

ADVANCES IN FRACTIONAL CALCULUS

Advances in Fractional Calculus

Theoretical Developments and Applications in Physics and Engineering

edited by

J. Sabatier

*Université de Bordeaux I
Talence, France*

O. P. Agrawal

*Southern Illinois University
Carbondale, IL, USA*

and

J. A. Tenreiro Machado

*Institute of Engineering of Porto
Portugal*



Springer

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN-13 978-1-4020-6041-0 (HB)
ISBN-13 978-1-4020-6042-7 (e-book)

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

www.springer.com

Printed on acid-free paper

The views and opinions expressed in all the papers of this book are the authors' personal one.

The copyright of the individual papers belong to the authors. Copies cannot be reproduced for commercial profit.

All Rights Reserved
© 2007 Springer

No part of this work may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without written permission from the Publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work.

**We dedicate this book to the honorable memory of our
colleague and friend Professor Peter W. Krempf**

Table of Contents

Preface	xi

1. Analytical and Numerical Techniques..... 1

Three Classes of FDEs Amenable to Approximation Using a Galerkin Technique	3
<i>S. J. Singh, A. Chatterjee</i>	

Enumeration of the Real Zeros of the Mittag-Leffler Function $E_\alpha(z)$, $1 < \alpha < 2$	15
<i>J. W. Hanneken, D. M. Vaught, B. N. Narahari Achar</i>	

The Caputo Fractional Derivative: Initialization Issues Relative to Fractional Differential Equations	27
<i>B. N. Narahari Achar, C. F. Lorenzo, T. T. Hartley</i>	

Comparison of Five Numerical Schemes for Fractional Differential Equations	43
<i>O. P. Agrawal, P. Kumar</i>	

Suboptimum H_2 Pseudo-rational Approximations to Fractional-order Linear Time Invariant Systems	61
<i>D. Xue, Y. Chen</i>	

Linear Differential Equations of Fractional Order	77
<i>B. Bonilla, M. Rivero, J. J. Trujillo</i>	

Riesz Potentials as Centred Derivatives	93
<i>M. D. Ortigueira</i>	

2. Classical Mechanics and Particle Physics..... 113

On Fractional Variational Principles	115
<i>D. Baleanu, S. I. Muslih</i>	

Fractional Kinetics in Pseudochaotic Systems and Its Applications	127
<i>G. M. Zaslavsky</i>	

Semi-integrals and Semi-derivatives in Particle Physics	139
<i>P. W. Kreml</i>	

Mesoscopic Fractional Kinetic Equations versus a Riemann–Liouville Integral Type	155
<i>R. R. Nigmatullin, J. J. Trujillo</i>	

3. Diffusive Systems..... 169

Enhanced Tracer Diffusion in Porous Media with an Impermeable Boundary	171
<i>N. Krepysheva, L. Di Pietro, M. C. Néel</i>	

Solute Spreading in Heterogeneous Aggregated Porous Media.....	185
<i>K. Logvinova, M. C. Néel</i>	

Fractional Advective-Dispersive Equation as a Model of Solute Transport in Porous Media.....	199
<i>F. San Jose Martinez, Y. A. Pachepsky, W. J. Rawls</i>	

Modelling and Identification of Diffusive Systems using Fractional Models.....	213
<i>A. Benchellal, T. Poinot, J. C. Trigeassou</i>	

4. Modeling..... 227

Identification of Fractional Models from Frequency Data	229
<i>D. Valério, J. Sá da Costa</i>	

Dynamic Response of the Fractional Relaxor–Oscillator to a Harmonic Driving Force.....	243
<i>B. N. Narahari Achar, J. W. Hanneken</i>	

A Direct Approximation of Fractional Cole–Cole Systems by Ordinary First-order Processes	257
<i>M. Haschka, V. Krebs</i>	

Fractional Multimodels of the Gastrocnemius Muscle for Tetanus Pattern	271
<i>L. Sommacal, P. Melchior, J. M. Cabelguen, A. Oustaloup, A. Ijspeert</i>	

Limited-Bandwidth Fractional Differentiator: Synthesis and Application in Vibration Isolation.....	287
<i>P. Serrier, X. Moreau, A. Oustaloup</i>	

5. Electrical Systems..... 303

A Fractional Calculus Perspective in the Evolutionary Design of Combinational Circuits	305
<i>C. Reis, J. A. Tenreiro Machado, J. B. Cunha</i>	

Electrical Skin Phenomena: A Fractional Calculus Analysis.....	323
<i>J. A. Tenreiro Machado, I. S. Jesus, A. Galhano, J. B. Cunha, J. K. Tar</i>	

Implementation of Fractional-order Operators on Field Programmable Gate Arrays.....	333
<i>C. X. Jiang, J. E. Carletta, T. T. Hartley</i>	

Complex Order-Distributions Using Conjugated order Differintegrals....	347
<i>J. L. Adams, T. T. Hartley, C. F. Lorenzo</i>	

6. Viscoelastic and Disordered Media..... 361

Fractional Derivative Consideration on Nonlinear Viscoelastic Statical and Dynamical Behavior under Large Pre-displacement	363
<i>H. Nasuno, N. Shimizu, M. Fukunaga</i>	

Quasi-Fractals: New Possibilities in Description of Disordered Media ...	377
<i>R. R. Nigmatullin, A. P. Alekhin</i>	

Fractional Damping: Stochastic Origin and Finite Approximations.....	389
<i>S. J. Singh, A. Chatterjee</i>	

Analytical Modelling and Experimental Identification of Viscoelastic Mechanical Systems.....	403
<i>G. Catania, S. Sorrentino</i>	

7. Control 417

LMI Characterization of Fractional Systems Stability 419
M. Moze, J. Sabatier, A. Oustaloup

Active Wave Control for Flexible Structures Using Fractional
 Calculus 435
M. Kuroda

Fractional-order Control of a Flexible Manipulator 449
V. Feliu, B. M. Vinagre, C. A. Monje

Tuning Rules for Fractional PIDs 463
D. Valério, J. Sá da Costa

Frequency Band-Limited Fractional Differentiator Prefilter in Path
 Tracking Design 477
P. Melchior, A. Poty, A. Oustaloup

Flatness Control of a Fractional Thermal System 493
P. Melchior, M. Cugnet, J. Sabatier, A. Poty, A. Oustaloup

Robustness Comparison of Smith Predictor-based Control
 and Fractional-Order Control 511
P. Lanusse, A. Oustaloup

Robust Design of an Anti-windup Compensated 3rd-Generation
 CRONE Controller 527
P. Lanusse, A. Oustaloup, J. Sabatier

Robustness of Fractional-order Boundary Control of Time Fractional
 Wave Equations with Delayed Boundary Measurement Using
 the Smith Predictor 543
J. Liang, W. Zhang, Y. Chen, I. Podlubny

Preface

Fractional Calculus is a field of applied mathematics that deals with derivatives and integrals of arbitrary orders (including complex orders), and their applications in science, engineering, mathematics, economics, and other fields. It is also known by several other names such as Generalized Integral and Differential Calculus and Calculus of Arbitrary Order. The name “Fractional Calculus” is holdover from the period when it meant calculus of ration order. The seeds of fractional derivatives were planted over 300 years ago. Since then many great mathematicians (pure and applied) of their times, such as N. H. Abel, M. Caputo, L. Euler, J. Fourier, A. K. Grünwald, J. Hadamard, G. H. Hardy, O. Heaviside, H. J. Holmgren, P. S. Laplace, G. W. Leibniz, A. V. Letnikov, J. Liouville, B. Riemann M. Riesz, and H. Weyl, have contributed to this field. However, most scientists and engineers remain unaware of Fractional Calculus; it is not being taught in schools and colleges; and others remain skeptical of this field. There are several reasons for that: several of the definitions proposed for fractional derivatives were inconsistent, meaning they worked in some cases but not in others. The mathematics involved appeared very different from that of integer order calculus. There were almost no practical applications of this field, and it was considered by many as an abstract area containing only mathematical manipulations of little or no use.

Nearly 30 years ago, the paradigm began to shift from pure mathematical formulations to applications in various fields. During the last decade Fractional Calculus has been applied to almost every field of science, engineering, and mathematics. Some of the areas where Fractional Calculus has made a profound impact include viscoelasticity and rheology, electrical engineering, electrochemistry, biology, biophysics and bioengineering, signal and image processing, mechanics, mechatronics, physics, and control theory. Although some of the mathematical issues remain unsolved, most of the difficulties have been overcome, and most of the documented key mathematical issues in the field have been resolved to a point where many of the mathematical tools for both the integer- and fractional-order calculus are the same. The books and monographs of Oldham and Spanier (1974), Oustaloup (1991, 1994, 1995), Miller and Ross (1993), Samko, Kilbas, and Marichev (1993), Kiryakova (1994), Carpinteri and Mainardi (1997), Podlubny (1999), and Hilfer (2000) have been helpful in introducing the field to engineering, science, economics and finance, pure and applied mathematics communities. The progress in this field continues. Three

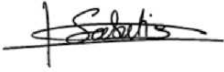
recent books in this field are by West, Grigolini, and Bologna (2003), Kilbas, Srivastava, and Trujillo (2005), and Magin (2006).

One of the major advantages of fractional calculus is that it can be considered as a super set of integer-order calculus. Thus, fractional calculus has the potential to accomplish what integer-order calculus cannot. We believe that many of the great future developments will come from the applications of fractional calculus to different fields. For this reason, we are promoting this field. We recently organized five symposia (the first symposium on Fractional Derivatives and Their Applications (FDTAs), ASME-DETC 2003, Chicago, Illinois, USA, September 2003; IFAC first workshop on Fractional Differentiations and its Applications (FDAs), Bordeaux, France, July 2004; Mini symposium on FDTAs, ENOC-2005, Eindhoven, the Netherlands, August 2005; the second symposium on FDTAs, ASME-DETC 2005, Long Beach, California, USA, September 2005; and IFAC second workshop on FDAs, Porto, Portugal, July 2006) and published several special issues which include *Signal Processing*, Vol. 83, No. 11, 2003 and Vol. 86, No. 10, 2006; *Nonlinear dynamics*, Vol. 29, No. 1–4, 2002 and Vol. 38, No. 1–4, 2004; and *Fractional Differentiations and its Applications*, Books on Demand, Germany, 2005. This book is an attempt to further advance the field of fractional derivatives and their applications.

In spite of the progress made in this field, many researchers continue to ask: “What are the applications of this field?” The answer can be found right here in this book. This book contains 37 papers on the applications of Fractional Calculus. These papers have been divided into seven categories based on their themes and applications, namely, analytical and numerical techniques, classical mechanics and particle physics, diffusive systems, viscoelastic and disordered media, electrical systems, modeling, and control. Applications, theories, and algorithms presented in these papers are contemporary, and they advance the state of knowledge in the field. We believe that researchers, new and old, would realize that we cannot remain within the boundaries of integral order calculus, that fractional calculus is indeed a viable mathematical tool that will accomplish far more than what integer calculus promises, and that fractional calculus is the calculus for the future.

Most of the papers in this book are expanded and improved versions of the papers presented at the Mini symposium on FDTAs, ENOC-2005, Eindhoven, The Netherlands, August 2005, and the second symposium on FDTAs, ASME-DETC 2005, Long Beach, California, USA, September 2005. We sincerely thank the ASME for allowing the authors to submit modified versions of their papers for this book. We also thank the authors for submitting their papers for this book and to Springer-Verlag for its

publication. We hope that readers will find this book useful and valuable in the advancement of their knowledge and their field.



Dr. J. Sabatier
LAPS – CRONE group
Bordeaux 1 University
France



Prof. O. P. Agrawal
College of Engineering
Southern Illinois
University
USA



Prof. J. A. Tenreiro Machado
Institute of Engineering of
Porto
Dept. of Electrotech.
Portugal

Part 1

Analytical and Numerical Techniques

THREE CLASSES OF FDEs AMENABLE TO APPROXIMATION USING A GALERKIN TECHNIQUE

Mechanical Engineering Department, Indian Institute of Science, Bangalore 560012, India

Abstract

We have recently presented elsewhere a Galerkin approximation scheme for fractional order derivatives, and used it to obtain accurate numerical solutions of second-order (mechanical) systems with fractional-order damping terms. Here, we demonstrate how that approximation can be used to find accurate numerical solutions of three different classes of fractional differential equations (FDEs), where for simplicity we assume that there is a single fractional-order derivative, with order between 0 and 1. In the first class of FDEs, the highest derivative has integer order greater than one. An example of a traveling point load on an infinite beam resting on an elastic, fractionally damped, foundation is studied. The second class contains FDEs where the highest derivative has order 1. Examples of the so-called generalized Basset's equation are studied. The third class contains FDEs where the highest derivative is the fractional-order derivative itself. Two specific examples are considered. In each example studied in the paper, the Galerkin-based numerical approximation is compared with analytical or semi-analytical solutions obtained by other means. In each case, the Galerkin approximation is found to be very good. We conclude that the Galerkin approximation can be used with confidence for a variety of FDEs, including possibly nonlinear ones for which analytical solutions may be difficult or impossible to obtain.

Keywords

1 Introduction

Consider the fractional differential equation

$${}^C_0 D_t^\alpha x(t) + \frac{d}{dt} \left[\int_0^t \frac{x(\tau)}{t-\tau} d\tau \right] = 0,$$

$$\frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} f(x) e^{-x^2} dx = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} f(x) e^{-x^2} dx$$

$$^{\alpha} x t \left[\frac{1}{\Gamma - \alpha} \int_0^t \frac{x \tau}{t - \tau} d\tau - \frac{1}{\Gamma - \alpha} \int_0^t \frac{x t - \tau}{\tau^{\alpha}} d\tau \right]$$

A musical score for a string quartet, featuring four staves with various musical notations including notes, rests, and dynamic markings like 'ff'. The notation is dense and complex, typical of a 20th-century composition.

Aa Ba c x t

1.

$$\alpha_{x,t} \approx \frac{1}{\Gamma_{\alpha} \Gamma_{-\alpha}} \mathbf{c}^T \mathbf{a},$$

[illegible]

The image shows a page of handwritten musical notation. At the top, the title "L'Espresso" is written in a stylized, cursive font. Below the title, there are ten staves of music. The notation includes various musical symbols such as notes, rests, and dynamic markings like "ff" (fortissimo). The handwriting is in black ink on a white background. The music appears to be a single melodic line, possibly for a piano or violin. The staves are connected by a single line on the left side. The overall style is that of a personal manuscript or a composer's sketch.

¹ A Maple-8 worksheet to compute the matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{c}$ is available on [5].

2 Traveling Load on an Infinite Beam

$$u_{xxxx} - \frac{m}{EI} u_{tt} - \frac{c}{EI} t^{1/2} u - \frac{k}{EI} u = -\frac{1}{EI} \delta(x - vt), \quad u \pm \infty, t \equiv 0.$$

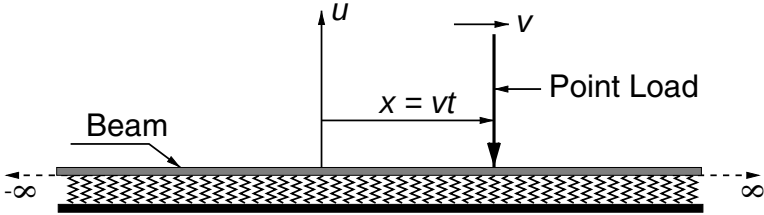


Fig. 1. Traveling point load on an infinite beam with a fractionally damped elastic foundation.

2.1 With Galerkin

$$u_{xxxx} - \frac{m}{EI} u_{tt} - \frac{c}{EI \Gamma(\frac{1}{2})} u - \frac{k}{EI} u = -\frac{1}{EI} \delta(x - vt)$$

$$\mathbf{A} \ddot{\mathbf{a}} + \mathbf{B} \mathbf{a} = \mathbf{c} u_t,$$

$$u_{\xi\xi\xi\xi} = \frac{m}{EI} \left(v^2 u_{\xi\xi} - v u_{\xi\tau} - u_{\tau\tau} - \frac{c}{\Gamma} \frac{c}{\Gamma} \mathbf{c}^T \mathbf{a} - k u \right) - \frac{\delta}{EI} \xi$$

$$\mathbf{A} \mathbf{a}_\tau - v \mathbf{a}_\xi = \mathbf{B} \mathbf{a} - \mathbf{c} u_\tau - v u_\xi.$$

$$u_{\xi\xi\xi\xi} = \frac{m}{EI} \left(v^2 u_{\xi\xi} - \frac{c}{\Gamma} \frac{c}{\Gamma} \mathbf{c}^T \mathbf{a} - k u \right) - \frac{\delta}{EI} \xi$$

$$-v \mathbf{A} \mathbf{a}_\xi = \mathbf{B} \mathbf{a} - v \mathbf{c} u_\xi.$$

2.2 Without Galerkin

$$\frac{1}{t} u(t, x) = \frac{1}{\Gamma} \int_0^t \frac{u(z, x)}{\sqrt{t-z}} dz.$$

$$\frac{1}{t} u(t, x) = \frac{1}{\Gamma} \int_0^t \frac{u(t-w, x)}{\sqrt{w}} dw.$$

$$\begin{aligned} & \xi = x - vt, \quad \tau = t, \quad u = -v u_\xi - u_\tau \\ & u(t-w, x) = -v u_\xi \xi + v w \quad \xi = x - vt \Rightarrow x - v(t-w) = \xi + v w \end{aligned}$$

$$\begin{aligned} \frac{1}{t} u(t, x) &= \frac{-v}{\Gamma} \int_0^\tau \frac{u_\xi \xi + v w}{\sqrt{w}} dw \\ &= \frac{-v}{\Gamma} \left(\int_0^\infty \frac{u_\xi \xi + v w}{\sqrt{w}} dw - \int_\tau^\infty \frac{u_\xi \xi + v w}{\sqrt{w}} dw \right). \end{aligned}$$

$$\frac{1}{t} u(t, x) = \frac{-v}{\Gamma} \int_0^\infty \frac{u_\xi \xi + v w}{\sqrt{w}} dw.$$

$$y = \xi + v w$$

$$\begin{aligned}
 {}^{1/2}u(t, x) &= \frac{-\sqrt{v}}{\Gamma} \int_{\xi}^{\infty} \frac{u'(y)}{\sqrt{y-\xi}} dy - \frac{-\sqrt{v}}{\Gamma} \int_{-\infty}^{\infty} \frac{y-\xi}{\sqrt{y-\xi}} \frac{u'(y)}{\sqrt{y-\xi}} dy, \\
 &\quad \text{without } \int_{-\infty}^{\infty} \frac{y-\xi}{\sqrt{y-\xi}} \frac{u'(y)}{\sqrt{y-\xi}} dy \\
 u_{\xi\xi\xi} &= \frac{mv^2}{EI} u_{\xi\xi} - \frac{c\sqrt{v}}{EI\Gamma} \int_{-\infty}^{\infty} \frac{y-\xi}{\sqrt{y-\xi}} \frac{u'(y)}{\sqrt{y-\xi}} dy - \frac{k}{EI} u - \frac{\delta}{EI} \xi.
 \end{aligned}$$

2.3 Solutions, with Galerkin and without

$$\begin{aligned}
 &\xi > 0 \\
 &\xi < 0 \\
 &u_{\xi\xi\xi} = \frac{mv^2}{EI} u_{\xi\xi} - \frac{c\sqrt{v}}{EI\Gamma} \int_{-\infty}^{\infty} \frac{y-\xi}{\sqrt{y-\xi}} \frac{u'(y)}{\sqrt{y-\xi}} dy - \frac{k}{EI} u - \frac{\delta}{EI} \xi
 \end{aligned}$$

$$U(\omega) = \frac{\sqrt{-i\omega}}{-EI\omega^4\sqrt{-i\omega} - mv^2\omega^2\sqrt{-i\omega} - ic\sqrt{v}\omega - k\sqrt{-i\omega}}$$

$$\begin{aligned}
 &\xi > 0 \\
 &\xi < 0 \\
 &u_{\xi\xi\xi} = \frac{mv^2}{EI} u_{\xi\xi} - \frac{c\sqrt{v}}{EI\Gamma} \int_{-\infty}^{\infty} \frac{y-\xi}{\sqrt{y-\xi}} \frac{u'(y)}{\sqrt{y-\xi}} dy - \frac{k}{EI} u - \frac{\delta}{EI} \xi
 \end{aligned}$$

2.4 Results

$$\begin{aligned}
 &m, EI, k, v, c, \delta \\
 &u_{\xi\xi\xi} = \frac{mv^2}{EI} u_{\xi\xi} - \frac{c\sqrt{v}}{EI\Gamma} \int_{-\infty}^{\infty} \frac{y-\xi}{\sqrt{y-\xi}} \frac{u'(y)}{\sqrt{y-\xi}} dy - \frac{k}{EI} u - \frac{\delta}{EI} \xi
 \end{aligned}$$

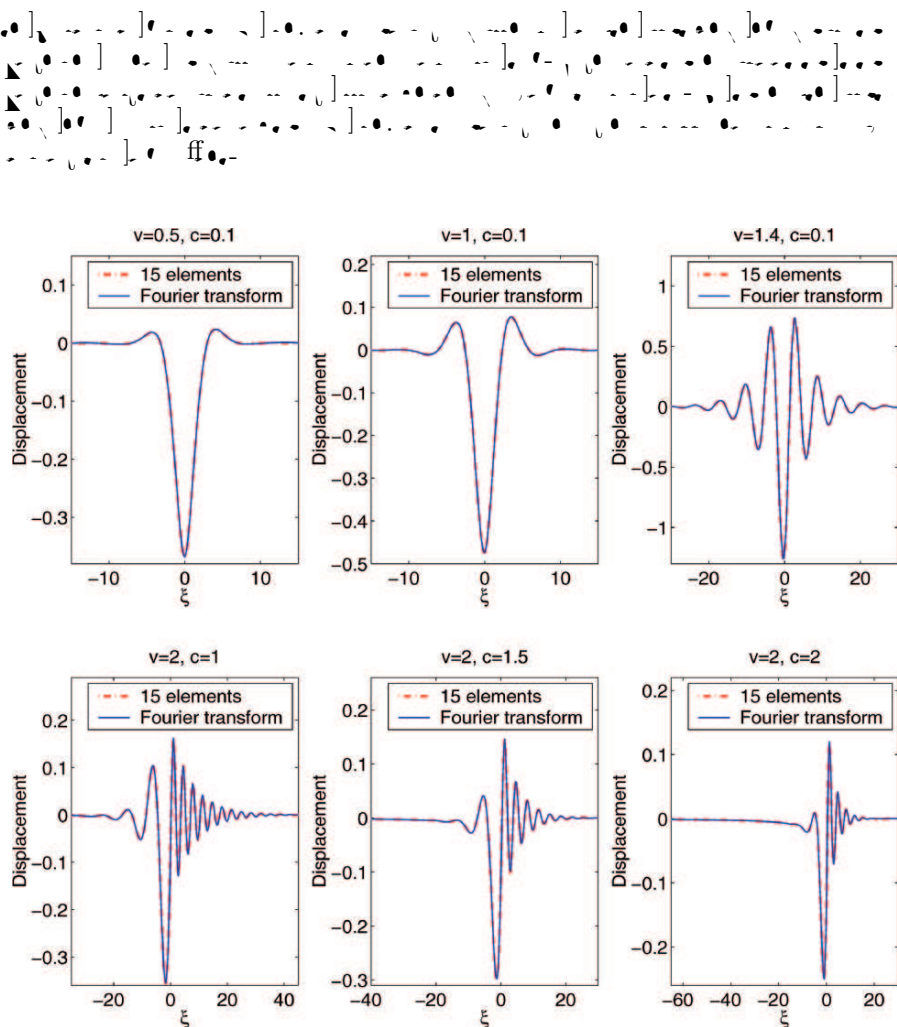
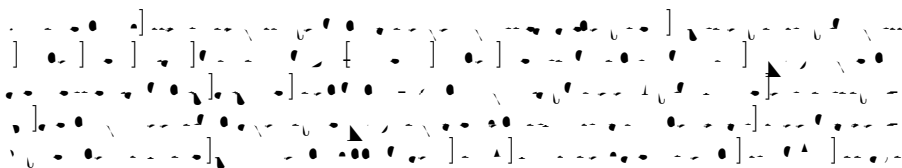


Fig. 2. Numerical results for a traveling point load on an infinite beam at steady state.

3 Off Spheres Falling Through Viscous Liquids



3.3 Results

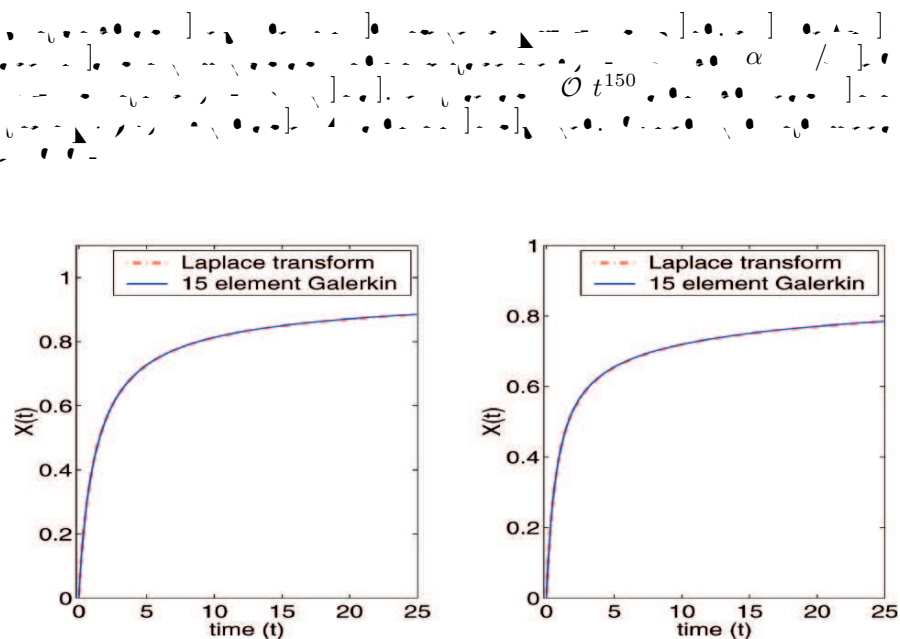


Fig. 3. Comparison between Laplace transform and 15-element Galerkin approximation solutions: Left: $\alpha = 1/2$ and sum in Eq. (14) upto $\mathcal{O}(t^{150})$ term. Right: $\alpha = 1/3$ and sum in Eq. (14) upto $\mathcal{O}(t^{150})$ term.

4 FDEs With Highest Derivative Fractional

$${}^{\alpha}x_t - x_t = f(t), \quad x(0) = x_0.$$

4.1 Adaptation of the Galerkin approximation

$${}^{\alpha}x_t - x_t = f(t), \quad x(0) = x_0.$$

$f t$ ff
 $h t$ $x t$ α
 $h t$ $g t$ $h t$ $x t$ $h t$
 $h t$ $g t$ $g t$ $g t$ g
 t $h t$ $g t$ $h t$ $x t$
 α $g t$ $x t$
 $h t$ $g t$, g

$$g^{\alpha} g^{\beta} t \left[\frac{1}{\Gamma - \alpha \Gamma - \alpha} \mathbf{c}^T \mathbf{a} \right]$$

$$g(t) = \frac{1}{\Gamma - \alpha} \int_0^t \frac{g(\tau)}{t - \tau} d\tau \quad \frac{1}{\Gamma - \alpha} \int_0^t \frac{h(\tau)}{t - \tau} d\tau \quad \alpha^{-1} h(t),$$

$$\alpha^{-1} h t \left[\frac{1}{\Gamma - \alpha \Gamma - \alpha} \mathbf{c}^T \mathbf{a} \right.$$

$$x(t) = \frac{1}{\Gamma(\alpha)\Gamma(-\alpha)} \mathbf{c}_{1-\alpha}^T \mathbf{a} \quad .$$

$$x(t) = e^{-\alpha y(t)}$$

4.2 Analytical solutions

$$X(s) = \frac{1}{s} - \frac{s^{-1/3}}{s^{4/3}} = \frac{1}{s} - \frac{s^{-1/3}}{s^{4/3}}.$$

4.3 Results

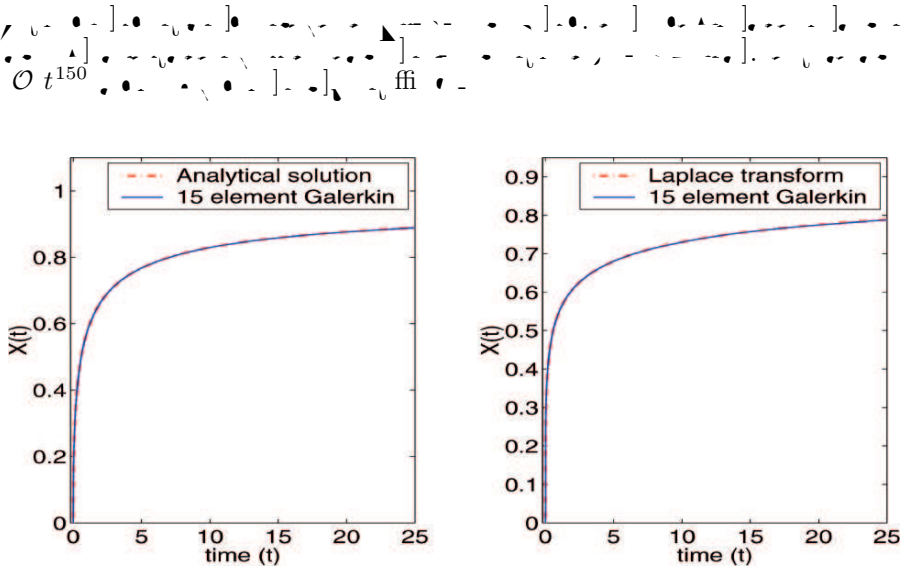


Fig. 4. Comparison between analytical and 15-element Galerkin approximation solutions. Left: $\alpha = 1/2$. Right: $\alpha = 1/3$. For $\alpha = 1/3$, the series is summed up to $\mathcal{O}(t^{150})$.

5 Discussion and Conclusions

References

1. Samko SG, Kilbas AA, Marichev OI (1993) *Fractional Integrals and Derivatives: Theory and Applications*. Gordon and Breach, Amsterdam.
2. Oldham KB (1974) *The Fractional Calculus*. Academic Press, New York.
3. Koh CG, Kelly JM (1990) Earthquake Eng. Struc. Dyn., 19:229–241.
4. Singh SJ, Chatterjee A (2005) *Nonlinear Dynamics* (in press).
5. http://www.geocities.com/dynamics_iisc/SystemMatrices.zip
6. Basset AB (1910) Quart. J. Math. 41:369–381.
7. Mainardi F, Pironi P, Tampieri F (1995) On a Generalization of Basset Problem via Fractional Calculus, in: Proceedings CANCAM 95.
8. Mainardi F (1996) Chaos, Solitons Fractals, 7(9):1461–1477.
9. Hairer E, Wanner G (1991) *Solving Ordinary Differential Equations II: Stiff and Differential Algebraic Problems*. Springer, Berlin.

ENUMERATION OF THE REAL ZEROS OF THE MITTAG-LEFFLER FUNCTION $E_\alpha(z)$, $1 < \alpha < 2$

John W. Hanneken¹, David M. Vaught², and B. N. Narahari Achar³

¹ University of Memphis, Physics Department, Memphis, TN 38152; Tel: 901.678.2417, Fax: 901.678.4733, E-mail: jhannekn@memphis.edu

² University of Memphis, Physics Department, Memphis, TN 38152; DHawkeye10@aol.com

³ University of Memphis, Physics Department, Memphis, TN 38152; Tel: 901.678.3122, Fax: 901.678.4733, E-mail: nachar@memphis.edu

Abstract

The Mittag-Leffler function $E_\alpha(z)$, which is a generalization of the exponential function, arises frequently in the solutions of physical problems described by differential and/or integral equations of fractional order. Consequently, the zeros of $E_\alpha(z)$ and their distribution are of fundamental importance and play a significant role in the dynamic solutions. The Mittag-Leffler function $E_\alpha(z)$ is known to have a finite number of real zeros in the range $1 < \alpha < 2$ which is applicable for many physical problems. What has not been known is the exact number of real zeros of $E_\alpha(z)$ for a given value of α in this range. An iteration formula is derived for calculating the number of real zeros of $E_\alpha(z)$ for any value of α in the range $1 < \alpha < 2$ and some specific results are tabulated.

Keywords

Mittag-Leffler functions, zeros, fractional calculus.

1 Introduction

The single parameter Mittag-Leffler function $E_\alpha(z)$ is defined over the entire complex plane by

$$E_\alpha(z) = \sum_{k=0}^{\infty} \frac{z^k}{\Gamma(\alpha k + 1)} \quad \alpha > 0, z \in \mathbb{C} \quad (1)$$

and is named after Mittag-Leffler who introduced it in 1903 [1,2]. The two parameter generalized Mittag-Leffler function, which was introduced later [3,4], is also defined over the entire complex plane, and is given by

$$E_{\alpha,\beta}(z) = \sum_{k=0}^{\infty} \frac{z^k}{\Gamma(\alpha k + \beta)} \quad \alpha, \beta > 0, z \in \mathbb{C} \quad (2)$$

It may be noted that when $\beta = 1$, $E_{\alpha,1}(z) = E_{\alpha}(z)$. Properties of the Mittag-Leffler functions have been summarized in several references [5–7]. Although others have considered complex α [8,9] and complex β [10], the present work is restricted to real α and β . The Mittag-Leffler functions are natural extensions of the exponential function and solutions of fractional-order differential equations are often expressed in terms of Mittag-Leffler functions in much the same way that solutions of many integer order differential equations may be expressed in terms of exponential functions. Consequently, the zeros of $E_{\alpha,1}(z)$, which play a significant role in the dynamic solutions, are of intrinsic interest.

Except for the special case of $\alpha = 1$, in general $E_{\alpha,1}(z)$ has an infinite number of zeros [11,12] and all complex zeros of $E_{\alpha}(z)$ appear as pairs of complex conjugates [13]. To facilitate the discussion of the zeros, the domain of α values can be conveniently divided into four ranges: $0 < \alpha < 1$, $\alpha = 1$, $1 < \alpha < 2$, and $\alpha \geq 2$ based on the nature of the zeros, but $E_{\alpha,1}(z)$ and its zeros exhibit similar properties within each range. For $0 < \alpha < 1$, $E_{\alpha}(z)$ has no real zeros [14] and thus must have an infinite number of complex zeros. For $\alpha = 1$, $E_{1,1}(z)$ can be written as $E_1(z) = \exp(z)$, which has no zeros real or complex. For $1 < \alpha < 2$, $E_{\alpha}(z)$ has a finite number of zeros on the negative real axis [5,8,9,11,14] and must in addition have an infinite number of complex zeros [11,15]. For $\alpha \geq 2$, $E_{\alpha}(z)$ has an infinite number of zeros that are real, negative, and simple and no complex zeros [8–10,16]. Note that regardless of the range of α , $E_{\alpha,1}(z)$ has no positive real zeros. Thus, for convenience, the variable x will be used to represent a positive real number so that $E_{\alpha,1}(-x)$ clearly has a negative real argument. Real zeros occur only in the ranges $1 < \alpha < 2$, and $\alpha \geq 2$. The range $1 < \alpha < 2$ is the range for which the least is known and yet is quite relevant for many physical problems [6,17]. The objective of this paper is to determine the exact number of real zeros for $E_{\alpha,1}(-x)$ for arbitrary α in the range $1 < \alpha < 2$. These results will be discussed later in connection with an asymptotic formula for the number of real zeros valid near $\alpha = 2$ [14]. The first requirement is a discussion of how to calculate $E_{\alpha,1}(-x)$ accurately.

2 Numerical Evaluation of $E_{\alpha,\beta}(-x)$

Numerical values of $E_{\alpha,\beta}(z)$ are easily calculated using the power series given in Eq. (2) when the argument z is not too large. However, for large arguments this method is impractical because of the extremely slow convergence of the series. Instead, use will be made of the representation of $E_{\alpha,\beta}(z)$ as a Laplace inversion integral [6]

$$E_{\alpha,\beta}(z) = \frac{1}{2\pi i} \int_{Br} e^s \frac{s^{\alpha-\beta}}{s^\alpha - z} ds \quad (3)$$

where Br denotes the Bromwich path. Using standard techniques in the theory of calculus of residues [18], $E_{\alpha,\beta}(\lambda z^\alpha)$ can be decomposed into two parts [14]. For the special case of a negative real argument, the result is given by:

$$E_{\alpha,\beta}(-x) = g_{\alpha,\beta}(-x) + f_{\alpha,\beta}(-x) \quad (4a)$$

$$g_{\alpha,\beta}(-x) = \frac{\frac{2}{\alpha} \exp\left[x^{1/\alpha} \cos\left(\frac{\pi}{\alpha}\right)\right] \cos\left[\frac{\pi(1-\beta)}{\alpha} + x^{1/\alpha} \sin\left(\frac{\pi}{\alpha}\right)\right]}{x^{(\beta-1)/\alpha}} \quad (4b)$$

$$f_{\alpha,\beta}(-x) = \frac{\frac{1}{\pi} \int_0^\infty \left\{ \frac{\exp(-x^{1/\alpha} r) r^{\alpha-\beta} [r^\alpha \sin(\pi\beta) + \sin[\pi(\beta-\alpha)]]}{r^{2\alpha} + 2r^\alpha \cos(\pi\alpha) + 1} \right\} dr}{x^{(\beta-1)/\alpha}} \quad (4c)$$

where $\alpha + 1 > \beta$ and for $\alpha < 1$, $g_{\alpha,\beta}(-x) = 0$. For the special case of $\beta = 1$ Eqs. (4a-c) reduce to

$$E_{\alpha,1}(-x) = g_{\alpha,1}(-x) + f_{\alpha,1}(-x) \quad (5a)$$

$$g_{\alpha,1}(-x) = \frac{2}{\alpha} \exp\left[x^{\frac{1}{\alpha}} \cos\left(\frac{\pi}{\alpha}\right)\right] \cos\left[\frac{\pi(1-\beta)}{\alpha} + x^{\frac{1}{\alpha}} \sin\left(\frac{\pi}{\alpha}\right)\right] \quad (5b)$$

$$f_{\alpha,1}(-x) = \frac{1}{\pi} \int_0^\infty \left\{ \frac{\exp(-x^{1/\alpha} r) r^{\alpha-1} \sin(\pi\alpha)}{r^{2\alpha} + 2r^\alpha \cos(\pi\alpha) + 1} \right\} dr \quad (5c)$$

Numerical values of the Mittag-Leffler function $E_{\alpha,1}(-x)$ were computed primarily from Eqs. (5a–c) using Mathematica[19] with the integration performed using the built-in function `NIntegrate`. The values computed using Eqs. (5a–c) were in agreement to better than 40 significant digits with the values calculated directly from Eq. (1) for small values of the argument. As an alternative to the numerical integration required in Eq. (5c), $f_{\alpha,1}(-x)$ can be written in an asymptotic infinite series as follows[14]

$$f_{\alpha,1}(-x) = \frac{1}{x\Gamma(1-\alpha)} - \frac{1}{x^2\Gamma(1-2\alpha)} + \cdots = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{x^n \Gamma(1-n\alpha)} \quad (6)$$

This series is particularly useful when both x and the gamma function are large and the series converges very quickly. The value of the gamma function approaches infinity as its argument approaches a negative integer. Thus, Eq. (6) is most useful for α close to 2 and x large.

3 Zeros of $E_{\alpha,1}(-x)$ of Multiplicity 2

Critical to the derivation of a formula for the number of real zeros is an understanding of the nature of the zeros and this is best done by examining the graphs of $E_{\alpha,1}(-x)$. For $1 < \alpha < 2$, $E_{\alpha,1}(0) = 1$ and for large x values $E_{\alpha,1}(-x)$ is negative and asymptotically approaches zero governed predominately by $f_{\alpha,1}(-x)$, Eq. (5c), with the exponentially decreasing oscillations of $g_{\alpha,1}(-x)$, Eq. (5b), superimposed. The fact that the curves of $E_{\alpha,1}(-x)$ are positive at $x = 0$ and ultimately become negative for large x implies that $E_{\alpha,1}(-x)$ can only cross the x -axis an odd number of times[5]. This is illustrated in the plot of $E_{\alpha,1}(-x)$ for $\alpha = 1.3$ shown in Fig. 1.

The curve exhibits only one zero at $x \cong 2.293$ and for larger x remains negative with the superimposed oscillation of $g_{\alpha,1}(-x)$ imperceptible on this scale. The rate of exponential decay of $g_{\alpha,1}(-x)$ is determined by the exponent $x^{1/\alpha} \cos(\pi/\alpha)$, the $\cos(\pi/\alpha)$ being negative in the range $1 < \alpha < 2$. As α increases this exponent decreases resulting in larger amplitude oscillations. This is illustrated in the graph of $E_{\alpha,1}(-x)$ for $\alpha = 1.5$ also shown in Fig. 1. The larger amplitude oscillations of $g_{\alpha,1}(-x)$ give rise to a relative maximum at $x \cong 17.472$ extending above the x -axis and yielding two more zeros at $x \cong 13.765$ and $x \cong 24.243$ in addition to the one at $x \cong 2.110$.

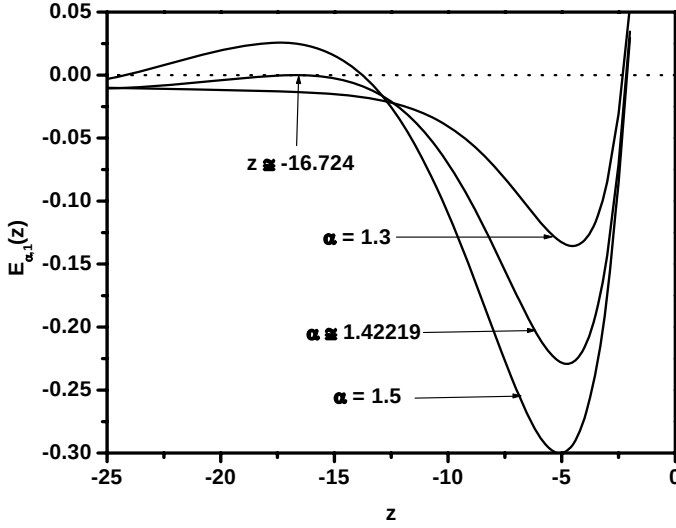


Fig. 1. Plots of $E_{\alpha,1}(-x)$ for various values of α .

Clearly, there is a value of α between $\alpha = 1.3$ and $\alpha = 1.5$ for which the curve of $E_{\alpha,1}(-x)$ is exactly tangential to the x-axis. This is illustrated in the graph of $E_{\alpha,1}(-x)$ for $\alpha \cong 1.422190690801$ also shown in Fig. 1. This curve has a zero at $x \cong 2.145$ and is tangential at $x \cong 16.724$ where it has a zero of multiplicity of 2 still yielding an odd total number of zeros. It may be noted that for $\alpha = 1.3$ the curve crosses the x-axis only once yielding one zero and for $\alpha = 1.5$ the curve crosses the x-axis 3 times yielding 3 zeros. Thus, the value of $\alpha \cong 1.422190690801$ separates the range of α values where $E_{\alpha,1}(-x)$ has only one zero from the range where $E_{\alpha,1}(-x)$ has three zeros. The next larger value of α where the curve is tangential to the x-axis is at $\alpha \cong 1.5718839229424$ where $E_{\alpha,1}(-x)$ has five zeros. The iteration formula for the number of real zeros described in the next section depends essentially on the existence of these values of α where the curve of $E_{\alpha,1}(-x)$ is tangential to the x-axis and for which one of the zeros has a multiplicity of 2. The first 5,641 of these α values where the curve of $E_{\alpha,1}(-x)$ is tangential to the x-axis have been numerically determined. A few selected values are given in Table 1. These values will be most useful in section 5 to establish ranges of reliability for the iteration results for $\alpha < 1.999$. In reading Table 1, for example, α_5 is the lowest value of α for which $E_{\alpha,1}(-x)$ has 5 zeros and $E_{\alpha,1}(-x)$ is tangential to the x-axis. Thus, $E_{\alpha,1}(-x)$ has 5 zeros for $\alpha_5 \leq \alpha < \alpha_7$, 7 zeros for $\alpha_7 \leq \alpha < \alpha_9$, 9 zeros for $\alpha_9 \leq \alpha < \alpha_{11}$, ..., 11281 zeros for $\alpha_{11281} \leq \alpha < \alpha_{11283}$.

Table 1. Values of α (truncated) at which $E_{\alpha,1}(-x)$ is tangent to the x-axis

n	α_n	n	α_n
3	1.422190690801	11217	1.998994787610
5	1.571883922942	11219	1.998994948054
7	1.649068237342	11221	1.998995108443
9	1.698516223760	11223	1.998995268780
11	1.733693032768	11225	1.998995429062
13	1.760338811725	11227	1.998995589290
15	1.781392651685	11229	1.998995749465
17	1.798543344750	11231	1.998995909586
19	1.812841949070	11233	1.998996069654
21	1.824982270661	11235	1.998996229667
23	1.835443517675	11237	1.998996389627
25	1.844568817828	11239	1.998996549534
27	1.852611186687	11241	1.998996709387
29	1.859761810886	11243	1.998996869186
31	1.866168176867	11245	1.998997028932
33	1.871946096560	11247	1.998997188625
35	1.877187921171	11249	1.998997348263
37	1.881968294552	11251	1.998997507849
39	1.886348272721	11253	1.998997667381
41	1.890378331112	11255	1.998997826860
43	1.894100597857	11257	1.998997986285
45	1.897550537931	11259	1.998998145657
47	1.900758240821	11261	1.998998304976
49	1.903749417395	11263	1.998998464241
51	1.906546180470	11265	1.998998623453
53	1.909167662339	11267	1.998998782612
55	1.911630507999	11269	1.998998941718
57	1.913949272538	11271	1.998999100770
59	1.916136743903	11273	1.998999259770
61	1.918204207029	11275	1.998999418716
63	1.920161661487	11277	1.998999577609
65	1.922018001994	11279	1.998999736450
67	1.923781169033	11281	1.998999895237
69	1.925458275243	11283	1.999000053971

4 Iteration Formula

Two conditions must be satisfied for $E_{\alpha,1}(-x)$ to be tangent to the x -axis, namely both the function and its derivative must be zero, or

$$E_{\alpha,1}(-x) = 0 \quad \text{and} \quad \frac{d}{dx} [E_{\alpha,1}(-x)] = 0 \quad (7a,b)$$

From Eq. (5a), $E_{\alpha,1}(-x) = g_{\alpha,1}(-x) + f_{\alpha,1}(-x)$ the condition, Eq. (7a), requires $g_{\alpha,1}(-x) = -f_{\alpha,1}(-x)$. Substituting Eq. (5b) for $g_{\alpha}(-x)$ and Eq. (6) for $f_{\alpha}(-x)$ yields:

$$\frac{2}{\alpha} \exp \left[x^{\frac{1}{\alpha}} \cos \left(\frac{\pi}{\alpha} \right) \right] \cos \left[x^{\frac{1}{\alpha}} \sin \left(\frac{\pi}{\alpha} \right) \right] = \sum_{i=1}^{\infty} \frac{(-1)^i}{x^i \Gamma(1 - i\alpha)} \quad (8)$$

It is easy to show that

$$\frac{d}{dx} [E_{\alpha,1}(-x)] = -\frac{1}{\alpha} E_{\alpha,\alpha}(-x) \quad (9)$$

and thus the second condition, Eq. (7b), requires $E_{\alpha,\alpha}(-x) = 0$. Since from Eq. (4a), $E_{\alpha,\alpha}(-x) = g_{\alpha,\alpha}(-x) + f_{\alpha,\alpha}(-x)$ it follows that $g_{\alpha,\alpha}(-x) = -f_{\alpha,\alpha}(-x)$ where $g_{\alpha,\alpha}(-x)$ is given by Eq. (4b) which for $\alpha = \beta$ becomes

$$g_{\alpha,\alpha}(-x) = \frac{-\frac{2}{\alpha} \exp \left[x^{\frac{1}{\alpha}} \cos \left(\frac{\pi}{\alpha} \right) \right] \cos \left[\frac{\pi}{\alpha} + x^{\frac{1}{\alpha}} \sin \left(\frac{\pi}{\alpha} \right) \right]}{x^{(\alpha-1)/\alpha}} \quad (10)$$

The asymptotic expansion of Eq. (4c) with $\alpha = \beta$ is given by[14]:

$$f_{\alpha,\alpha}(-x) = \frac{-x^{-2}}{\Gamma(-\alpha)} + \frac{x^{-3}}{\Gamma(-2\alpha)} + \frac{-x^{-4}}{\Gamma(-3\alpha)} + \dots \quad (11)$$

In the limit of α close to 2 (when x will also be large) even the first term of the expansion in Eq. (11) will be negligibly small and consequently $f_{\alpha,\alpha}(-x) \cong 0$. Thus, the condition that $E_{\alpha,\alpha}(-x) = 0$ is approximately satisfied when $g_{\alpha,\alpha}(-x)$ in Eq. (10) equals zero, or

$$\cos \left[\frac{\pi}{\alpha} + x^{\frac{1}{\alpha}} \sin \left(\frac{\pi}{\alpha} \right) \right] = 0 \quad (12)$$

Equation (12) is satisfied when the cosine argument is given by $\pi/2 + 2m\pi$, with $m = 0, 1, 2, 3, \dots$. Solving Eq. (12) for x yields

$$x = \left[\frac{\pi/2 - \pi/\alpha + 2m\pi}{\sin(\pi/\alpha)} \right]^\alpha \quad m = 0, 1, 2, 3, \dots \quad (13)$$

Note that although Eq. (12) is also satisfied at $3\pi/2 + 2m\pi$, $E_{\alpha,\alpha}(-x)$ cannot be zero when cosine is negative. Substituting Eq. (13) into Eq. (8) and solving for m yields:

$$m = \frac{\ln \left\{ \frac{-\alpha [\sin(\pi/\alpha)]^\alpha}{2^{\alpha+1} \pi^\alpha m^\alpha \cos(\pi/\alpha - \pi/2) \Gamma(1-\alpha)} \right\}}{2\pi \cot(\pi/\alpha)} + \frac{1}{2\alpha} - \frac{1}{4} \\ + \frac{-\alpha \ln \left(1 + \frac{1}{4m} - \frac{1}{2m\alpha} \right) + \ln \left(1 + \sum_{i=1}^{\infty} A_i \right)}{2\pi \cot(\pi/\alpha)} \quad (14)$$

where

$$A_i = \frac{(-1)^i \Gamma(-\alpha) [\sin(\pi/\alpha)]^{i\alpha}}{(i+1)(\pi/2 - \pi/\alpha + 2m\pi)^{i\alpha} \Gamma[-(i+1)\alpha]} \quad i = 1, 2, 3, \dots$$

The A_i 's come from keeping terms beyond $i = 1$ in the infinite series in Eq. (8). In Eq. (14), m cannot be solved explicitly, but can be determined iteratively by guessing a value of m and using this value of m in Eq. (14) to calculate a new guess for m and repeating the process until consecutive values of m differ by less than some predetermined value (10^{-15} in this case). In an attempt to satisfy both Eqs. (7a and 7b), the iteration process converges to a value of m such that x given by Eq. (13) is just beyond the largest zero for that α value. Note that m represents the number of relative maxima of $E_{\alpha,1}(-x)$ from $x=0$ to the x value of the largest zero. To determine the number of real zeros, consider the representation, Eq. (5a), of $E_{\alpha,1}(-x)$ as a sum of two functions $g(-x)$ and $f(-x)$. The function $f_{\alpha,1}(-x)$ is negative for all x and is a completely monotonic function which decreases toward zero with increasing x [14]. The function $g_{\alpha,1}(-x)$ exhibits oscillations with an amplitude which decays exponentially. Each full period oscillation of the $\cos[x^{1/\alpha} \sin(\pi/\alpha)]$ term in $g_{\alpha,1}(-x)$ results in two zeros for when cosine is positive and $g_{\alpha,1}(-x)$ is larger than $f_{\alpha,1}(-x)$ it gives rise to a relative maximum in $E_{\alpha,1}(-x)$ above the x -axis. This process continues as long as the magnitude of $g_{\alpha,1}(-x)$ is larger than that of

$f_{\alpha,1}(-x)$, but when $|g_{\alpha,1}(-x)| < |f_{\alpha,1}(-x)|$ their sum is less than zero and no more zeros are possible. Thus, the number of real zeros of $E_{\alpha,1}(-x)$, n , is then given by

$$n = 2 [m] + 1 \quad (15)$$

where $[m]$ is the greatest integer $\leq m$. The greatest integer function is required because the largest zero does not coincide with the end of one full period. In addition, the 1 must be included because the largest zero occurs in a period during which the magnitude of $g_{\alpha,1}(-x)$ has decayed to less than $f_{\alpha,1}(-x)$, resulting in only one zero during this interval. Equations (14) and (15) are the main results of this paper.

5 Accuracy of the Iteration Results

Using Eqs.(14) and (15), the number of real zeros of $E_{\alpha,1}(-x)$ can be calculated for arbitrary α in the range $1 < \alpha < 2$ with some restrictions based on the number of significant digits in α . These restrictions result because of the approximate solution of $E_{\alpha,\alpha}(-x) = 0$ used in the derivation. The approximation that $f_{\alpha,\alpha}(-x) \cong 0$ in Eq. (11) improves as α approaches 2 and consequently the results of using Eqs. (14) and (15) become more accurate. When the value of α deviated further from 2, the results from Eqs. (14) and (15) become less accurate. However, the total number of real zeros in this case can be easily enumerated by a brute-force technique described later. Equations (14) and (15) do not yield reliable results for $1 < \alpha < 1.42$ but in this range $E_{\alpha,1}(-x)$ has only one real zero. Equations (14) and (15) do give the correct number of real zeros of $E_{\alpha,1}(-x)$ for $\alpha \geq 1.42$ when at most 3 significant digits in α are specified. As α gets closer to 2, α can be specified to an increasing number of significant digits. However, an increased number of significant digits in α does not guarantee the correct number of zeros, as illustrated by the following example. For $\alpha = 1.9796275$, Eqs. (14) and (15) correctly predict that $E_{\alpha,1}(-x)$ will have 349 zeros. However, at $\alpha = 1.9796276$, Eqs. (14) and (15) incorrectly predict 349 zeros instead of the correct 351. At $\alpha = 1.9796276$, the approximations used in deriving Eqs. (14) and (15) are not accurate enough to discriminate between 349 zeros at $\alpha = 1.9796275$ and 351 zeros at $\alpha = 1.9796276$. Thus, if α is specified to 8 significant digits, α must be ≥ 1.9796277 to be guaranteed that Eqs. (14) and (15) will predict the correct number of real zeros. If α is specified to a certain number of significant digits, Table 2 gives the range of α that will guarantee that the results of Eqs. (14) and (15) yield the correct number of real zeros.

Table 2. Ranges of reliability for the results

Significant digits in α	Range of α for reliable results from Eqs. (14) and (15)
3	$1.42 \leq \alpha < 2$
4	$1.573 \leq \alpha < 2$
5	$1.7815 \leq \alpha < 2$
6	$1.86618 \leq \alpha < 2$
7	$1.951713 \leq \alpha < 2$
8	$1.9796277 \leq \alpha < 2$
9	$1.99571096 \leq \alpha < 2$
10	$1.997045583 \leq \alpha < 2$
11	$1.9986590973 \leq \alpha < 2$

6 Results and Conclusions

Table 3 gives the number of real zeros of $E_{\alpha,1}(-x)$ computed from Eqs. (14) and (15) for various values of α all of which have been verified by the brute force counting method. Table 4 extends Table 3 to values of α closer to 2 where the results of Eqs. (14) and (15) are most accurate. For values of α not listed in either table, Eqs.(14) and(15) correctly predict the number of real zeros of $E_{\alpha,1}(-x)$ for any arbitrary α provided the restrictions on the number of significant digits specified in α are observed (Table 2).

Table 3. Number of real zeros of $E_{\alpha,1}(-x)$

α	# of zeros	α	# of zeros	α	# of zeros
1.000	0	1.900	45	1.990	815
1.100	1	1.910	53	1.991	923
1.200	1	1.920	61	1.992	1,059
1.300	1	1.930	73	1.993	1,237
1.400	1	1.940	91	1.994	1,479
1.500	3	1.950	115	1.995	1,825
1.600	5	1.960	153	1.996	2,357
1.700	9	1.970	219	1.997	3,273
1.800	17	1.980	357	1.998	5,181
1.900	45	1.990	815	1.999	1,1281

Table 4. Number of real zeros of $E_{\alpha,1}(-x)$ for $\alpha > 1.999$

$2-\alpha$	Number of real zeros
10^{-4}	142,803
10^{-5}	1,723,335
10^{-6}	20,160,229
10^{-7}	230,691,031
10^{-8}	2,596,455,273
10^{-9}	28,849,564,429
10^{-10}	317,262,155,731
10^{-11}	3,459,601,473,763
10^{-12}	37,460,093,329,007
10^{-13}	403,193,222,273,617
10^{-14}	4,317,438,639,773,315
10^{-15}	46,025,834,494,632,015
10^{-16}	488,741,129,109,758,967
10^{-17}	5,171,958,979,244,453,601
10^{-18}	54,562,572,375,712,516,775
10^{-19}	574,033,197,647,837,786,487
10^{-20}	6,024,205,251,646,954,541,059

References

1. Mittag-Leffler GM (1903) Sur la nouvelle fonction $E_{\alpha}(X)$. Comptes Rendus de l'Academie des Sciences, Paris Series II, Vol. 137, pp. 554–558.
2. Mittag-Leffler GM (1903) Sopra la funzione $E_{\alpha}(X)$. Rendiconti Accademia Nazionale dei Lincei Series V, Vol. 13, pp. 3–5.
3. Humbert P (1953) Quelques resultants relatifs a la fonction de Mittag-Leffler. Comptes Rendus de l'Academie des Sciences, Paris, Vol. 236, pp. 1467–1468.
4. Agarwal RP (1953) A propos d'une note de M. Pierre Humbert Comptes Rendus de l'Academie des Sciences, Paris, Vol. 236, pp. 2031–2032.
5. Erdelyi A, Magnus W, Oberhettinger F, Tricomi FG (1955) *Higher Transcendental Functions*, Vol. 3. McGraw-Hill, New York, pp. 206–212.
6. Mainardi F, Gorenflo R (1996) The Mittag-Leffler function in the Riemann-Liouville Fractional Calculus, in: Kilbas AA (ed.), *Boundary Value Problems, Special Functions and Fractional Calculus*. Belarusian State University, Minsk, Belarus, pp. 215–225.
7. Podlubny I (1999) *Fractional Differential Equations, Mathematics in Science and Engineering*, Vol. 198. Academic Press, San Diego, pp. 16–37.
8. Wiman A (1905) Über den Fundamentalsatz in der Theorie der Funktionen $E_{\alpha}(X)$. Acta Math. 29:191–201.

9. Wiman A (1905) Über die Nullstellen der Funktionen $E_\alpha(X)$. *Acta Math.* 29:217–234.
10. Ostrovskii V, Peresyokkova IN (1997) Nonasymptotic results on distribution of zeros of the function $E_\rho(z, \mu)$. *Anal. Math.* 23(4):283–296.
11. Djrbashian MM (ed.) (1993) Harmonic analysis and boundary value problems in the complex domain, in: *Operator Theory Advances and Applications*, Vol. 65. Birkhauser Verlag, Basel, Switzerland.
12. Sedletskii AM (2000) On zeros of functions of Mittag-Leffler type, *Math. Notes* 68(5):602–613.
13. Gorenflo R, Luchko Yu, Rogozin S (1997) Mittag-Leffler type functions: notes on growth properties and distribution of zeros. Fachbereich Mathematik und Informatik, A04/97, Freie Universitaet, Berlin, pp. 1–23. Downloadable from <http://www.math.fu-berlin.de/publ/index.html>
14. Gorenflo R, Mainardi F (1996) Fractional oscillations and Mittag-Leffler functions. Fachbereich Mathematik und Informatik, A14/96, Freie Universitaet, Berlin, pp. 1–22. Downloadable from <http://www.math.fuberlin.de/publ/index.html>
15. Ostrovskii V, Peresyokkova IN (1997) Nonasymptotic results on distribution of zeros of the function $E_\rho(z, \mu)$. *Anal. Math.* 23(4):283–296.
16. Popov AYu (2002) The spectral values of a boundary value problem and the zeros of Mittag-Leffler functions. *Differential Equations* 38(5):642–653.
17. Gorenflo R, Mainardi F (1997) Fractional calculus: integral and differential equations of fractional order. in: Carpenteri A, Mainardi R (eds.), *Fractals and Fractional Calculus in Continuum Mechanics*. Springer, Wien, pp. 223–276.
18. McLachlan NM (1963) *Complex Variable Theory and Transform Calculus with Technical Applications*, Second Edition, Cambridge University Press, Cambridge.
19. Mathematica Software System, Version 4, Wolfram Research, Champaign, IL.

THE CAPUTO FRACTIONAL DERIVATIVE: INITIALIZATION ISSUES RELATIVE TO FRACTIONAL DIFFERENTIAL EQUATIONS

B. N. Narahari Achar¹, Carl F. Lorenzo², and Tom T. Hartley³

¹ University of Memphis, Memphis, TN 38152; Tel: (901)678-3122, Fax: (901)678-4733,
E-mail: nachar@memphis.edu

² NASA Glenn Research Center, Cleveland, OH 44135

³ University of Akron, Akron, OH 44325

Abstract

Recognizing the importance of proper initialization of a system, which is evolving in time according to a differential equation of fractional order, Lorenzo and Hartley developed the method of properly incorporating the effect of the past (history) by means of an initialization function for the Riemann–Liouville and the Grunwald formulations of fractional calculus. The present work addresses this issue for the Caputo fractional derivative and cautions that the commonly held belief that the Caputo formulation of fractional derivatives properly accounts for the initialization effects is not generally true when applied to the solution of fractional differential equations.

Keywords

Caputo fractional derivatives, initialization issues.

1 Introduction

Lorenzo and Hartley (LH) [1,2] have clearly established the importance of time-dependent initialization function in taking into account the history of a system which evolves according to a differential equation of fractional order. They have considered both the Riemann–Liouville (RL) and the Grunwald formulations of fractional calculus [3–6] in developing the initialization function [7]. This paper examines the Caputo fractional derivative [8,9] with the objective of determining the inferred initialization, that is, the history function associated with the Caputo fractional derivative from the perspective of the Lorenzo–Hartley scheme. It will be shown that the commonly held belief that the Caputo derivative properly

accounts for the initialization effects is not generally true when applied to the solution of fractional differential equations.

After a brief description of the LH terminal initialization procedure for the RL fractional derivative, the initialization function for the Caputo derivative that would yield the same result as the initialized LH derivative is given. In the final section, initialization limitations of the Caputo derivative when applied to solutions of fractional differential equations are discussed.

2 Initialization of the Riemann–Liouville Fractional Differintegral

Consider the following q th order fractional integrals of $f(t)$, the first integral starting at time $t = a$, and the second, starting at time $t = c > a$, respectively:

$${}_a d_t^{-q} f(t) = \frac{1}{\Gamma(q)} \int_a^t (t - \tau)^{q-1} f(\tau) d\tau, \quad t > a, \quad (1)$$

and

$${}_c d_t^{-q} f(t) = \frac{1}{\Gamma(q)} \int_c^t (t - \tau)^{q-1} f(\tau) d\tau, \quad t > c. \quad (2)$$

It is assumed that the function $f(t)$ is zero for all $t \leq a$ the time interval between $t = a$ and $t = c$ being considered to be the “history” of the fractional integral ${}_c d_t^{-q} f(t)$. Initialization consists in adding a function ψ to the integral starting at time $t = c$ so that the result of fractional integration starting at time $t = c$ is equal to that of the integral starting at time $t = a$ for all $t > c$, i.e.,

$${}_c d_t^{-q} f(t) + \psi = {}_a d_t^{-q} f(t), \quad t > c \quad (3)$$

Or in other words,

$$\psi = \frac{1}{\Gamma(q)} \int_a^c (t - \tau)^{q-1} f(\tau) d\tau, \quad t > c. \quad (4)$$

Of the two types of initializations described by LH [7], only the “terminal initialization”, in which case the integral can only be initialized prior to the start time $t = c$, will be considered here. Then the **generalized fractional integral**, for arbitrary, real, and nonnegative values of v is defined by

$${}_c D_t^{-\nu} f(t) \equiv {}_c d_t^{-\nu} f(t) + \psi(f, -\nu, a, c, t), \quad \nu \geq 0, \quad (5)$$

$t > c \geq a$, and $f(t) = 0$ for all $t \leq a$

where $\psi(f, -\nu, a, c, t) = \frac{1}{\Gamma(\nu)} \int_a^c (t - \tau)^{\nu-1} f(\tau) d\tau$, as defined in Eq. (4)

The **generalized fractional derivative**, for q and p real is defined by

$${}_c D_t^q f(t) \equiv {}_c D_t^m {}_c D_t^{-p} f(t), \quad t > c, \quad (6)$$

where, m is an integer such that $m - 1 < q < m$, and $q = m - p$. Furthermore, $q \geq 0$ and $t > c \geq 0$.

In terms of the conventional notation,

$${}_c D_t^q f(t) = \frac{d^m}{dt^m} {}_c d_t^{-p} f(t) + \frac{d^m}{dt^m} \psi(f, -p, a, c, t) + \psi(h, m, a, c, t) \quad (7)$$

where, $t > c$, and $h = {}_a D_t^{-p} f(t)$. It is of course clear that $f(t)$ may be considered to be a composite function, for example a function different than $f(t)$ may be used for the history period, i.e., $a < t < c$, while $f(t)$ remains the function to be fractionally differintegrated, i.e., $t > c$. It has also been shown [7] that for terminal initialization of the integer derivative,

$$\psi(h, m, a, c, t) = 0, \quad t \geq c, \quad (8)$$

and the definition in Eq. (4) is applied for $\psi(f, -p, a, c, t)$ in Eq. (7). The next section considers the extention to the Caputo fractional derivative.

3 Initialization of the Caputo Fractional Derivative

The Caputo fractional derivative was introduced [8,9] to alleviate some of the difficulties associated with the RL approach to fractional differential equations when applied to the solution of physical problems and is defined by [8]:

$${}_a^C d_t^\alpha f(t) = \frac{1}{\Gamma(m - \alpha)} \int_a^t (t - \tau)^{m-\alpha-1} f^{(m)}(\tau) d\tau \quad (m - 1 < \alpha < m) \quad (9)$$

As is well known, in the solution of fractional differential equations, the initial conditions are specified in terms of fractional derivatives in the RL approach, but, in terms of integer order derivatives with known physical interpretations in the Caputo approach [10]. In view of the popularity of the Caputo formulation in applications of physical interest, the key question to be asked is: when viewed from the LH general initialization perspective, what “history” is inferred [11,12] for the Caputo derivative?

4 Relation Between the Initialized LH and Caputo Fractional Derivatives

As has been noted, the generalized initialization as applied to RL fractional derivative, according to LH is given by Eq. (7) and will be used in the following examples, where for convenience, α is used in the place of q , i.e., $\alpha = m - p > 0$, m is a positive integer, and as before, for terminal initialization, $\psi(h, m, a, c, t) = 0$. Hereafter $t = c$ corresponds to $t = 0$. Three cases will be considered below.

4.1 Simple cases: $0 < \alpha < 1$ and $1 < \alpha < 2$

We first consider the case when $0 < \alpha < 1$, i.e., $m = 1$, then

$$\begin{aligned} {}_0 D_t^\alpha f(t) &= {}_0 D_t^1 \left\{ {}_0 D_t^{-(1-\alpha)} f(t) \right\} \\ &= \frac{d}{dt} \left\{ {}_0 D_t^{-(1-\alpha)} f(t) \right\} + \psi(h, 1, a, 0, t), \quad t > 0, \end{aligned} \quad (10)$$

Noting that the initialization for the integer order derivative is zero

$${}_0 D_t^\alpha f(t) = \frac{d}{dt} \left\{ {}_0 d_t^{-(1-\alpha)} f(t) + \psi(f, -(1-\alpha), a, 0, t) \right\} + 0, \quad t > 0 \quad (11)$$

Substituting explicitly for the quantities in curly brackets in Eq. (11) yields

$${}_0D_t^\alpha f(t) = \frac{d}{dt} \left\{ \frac{1}{\Gamma(1-\alpha)} \int_0^t (t-\tau)^{-\alpha} f(\tau) d\tau \right\} + \frac{d}{dt} \left\{ \frac{1}{\Gamma(1-\alpha)} \int_a^0 (t-\tau)^{-\alpha} f(\tau) d\tau \right\} \quad (12)$$

Recasting the convolution integral by interchanging the arguments and carrying out the differentiation of the integral using Leibnitz' rule, yields

$$\begin{aligned} {}_0D_t^\alpha f(t) &= \frac{1}{\Gamma(1-\alpha)} \int_0^t \tau^{-\alpha} f'(t-\tau) d\tau + \frac{t^{-\alpha}}{\Gamma(1-\alpha)} f(0) \\ &\quad + \frac{1}{\Gamma(1-\alpha)} \frac{d}{dt} \int_a^0 (t-\tau)^{-\alpha} f(\tau) d\tau, \quad t > 0. \end{aligned} \quad (13)$$

Rewriting the argument of the convolution integral as $(t-\tau)^{-\alpha} f'(\tau)$ and using the definition of the Caputo derivative, Eq. (9) with $m=1$ and $0 < \alpha < 1$, one can write the following expression relating the Caputo derivative to the initialized LH derivative for $0 < \alpha < 1$:

$${}_0D_t^\alpha f(t) = {}_0^C d_t^\alpha f(t) + \frac{t^{-\alpha} f(0)}{\Gamma(1-\alpha)} + \frac{d}{dt} \psi(f, -(1-\alpha), a, 0, t), \quad t > 0. \quad (14)$$

where the last integral in Eq. (13) is restated as an LH initialization.

For the case $1 < \alpha < 2$, $m=2$ and the initialized LH derivative given by

$${}_0D_t^\alpha f(t) = \frac{d}{dt} \left(\frac{d}{dt} \left\{ {}_0d_t^{-(2-\alpha)} f(t) + \psi(f, -(2-\alpha), a, 0, t) \right\} + 0 \right), \quad t > 0 \quad (15)$$

yields on substituting explicit expressions for the quantities in the curly brackets

$${}_0D_t^\alpha f(t) = \frac{d}{dt} \left(\frac{d}{dt} \left\{ \frac{1}{\Gamma(2-\alpha)} \int_0^t (t-\tau)^{1-\alpha} f(\tau) d\tau \right\} \right) + \frac{d^2}{dt^2} \psi(f, -(2-\alpha), a, 0, t) \quad (16)$$

Recasting the convolution integral in Eq. (16) by interchanging the arguments and carrying out the differentiation of the integral using Leibnitz' rule yields the expression relating the Caputo derivative to the initialized LH derivative for the case $1 < \alpha < 2$ as [11]:

$${}_0D_t^\alpha f(t) = {}^CD_t^\alpha f(t) + \frac{t^{1-\alpha} f'(0)}{\Gamma(2-\alpha)} + \frac{t^{-\alpha} f(0)}{\Gamma(1-\alpha)} + \frac{d^2}{dt^2} \psi(f, -(2-\alpha), a, 0, t), t > 0. \quad (17)$$

The expressions in Eq. (14) and Eq. (17) can be generalized as shown below.

4.2 General case $m-1 < \alpha < m$

Generalizing to the case when $m-1 < \alpha < m$ we get

$${}_0D_t^\alpha f(t) = \frac{d^m}{dt^m} {}_0d_t^{-(m-\alpha)} f(t) + \frac{d^m}{dt^m} \psi(f, -(m-\alpha), a, 0, t), t > 0, \quad (18)$$

$$\begin{aligned} {}_0D_t^\alpha f(t) &= \frac{1}{\Gamma(m-\alpha)} \int_0^t (t-\tau)^{m-\alpha-1} f^m(\tau) d\tau \\ &\quad + \sum_{k=0}^{m-1} \frac{t^{k-\alpha} f^k(0^+)}{\Gamma(k-\alpha+1)} + \frac{d^m}{dt^m} \psi(f, -(m-\alpha), a, 0, t), \end{aligned} \quad (19)$$

or,

$$\begin{aligned} {}_0D_t^\alpha f(t) &= {}^CD_t^\alpha f(t) + \sum_{k=0}^{m-1} \frac{t^{k-\alpha} f^k(0^+)}{\Gamma(k-\alpha+1)} \\ &\quad + \frac{d^m}{dt^m} \psi(f, -(m-\alpha), a, 0, t), \quad t > 0, \quad m-1 < \alpha < m. \end{aligned} \quad (20)$$

Equation (20) expresses the LH α order derivative ${}_0D_t^\alpha f(t)$ in terms of the α order Caputo derivative and additional terms. The additional terms consist of a polynomial in t with coefficients given by the values of the function $f(t)$ and its integer-order derivatives $f^{(k)}(t)$, all evaluated at $t = 0^+$, and the LH initialization for a fractional derivative under the assumption of terminal initialization. The polynomial contains a term ($k = 0$ term), which is singular at $t = 0$ for $\alpha > 0$. The details of the derivation can be found in ref. [11]. For the range $0 < \alpha < 1$, Eq. (20) simplifies to the Eq. (14), and for the range $1 < \alpha < 2$,

it reduces to Eq. (17). Expressions in Eq. (14), Eq. (17), and Eq. (20) will now be used to determine the history inferred by the use of the Caputo derivative.

5 Inferred History of the Caputo Derivative

It is important to determine the “history” inferred by use of the Caputo derivative of a function $f(t)$. This can be achieved by setting the Caputo derivative equal to the LH fractional derivative of the same order α , and for the same function $f(t)$, for $t > 0$.

5.1 Simple cases: $0 < \alpha < 1$ and $1 < \alpha < 2$

It follows from Eq. (14) that the two derivatives will be equal for $0 < \alpha < 1$ if

$$\frac{d}{dt} \psi(f, -(1-\alpha), a, 0, t) = -\frac{t^{-\alpha} f(0^+)}{\Gamma(1-\alpha)} \quad t > 0. \quad (21)$$

For clarity of presentation we will call the initialization function, yet to be determined, $f_1(t)$ for $a \leq t \leq 0$, to differentiate it from $f(0^+)$ on the right-hand side (RHS). For the terminal initialization considered in this note, it follows that “history” would be given by a function $f_1(\tau)$, satisfying the following equation:

$$\frac{d}{dt} \left\{ \frac{1}{\Gamma(1-\alpha)} \int_a^0 (t-\tau)^{-\alpha} f_1(\tau) d\tau \right\} = -\frac{t^{-\alpha} f(0^+)}{\Gamma(1-\alpha)}, \quad t > 0. \quad (22)$$

It is important to note that the left-hand side (LHS) of Eq. (22), which is the required initialization, is only related to the value of the function evaluated at $t=0$, on the RHS, and not to the function or its derivatives at any instant prior to $t=0$. Specifically, RHS of Eq. (22) is not a function of “ a ”. To determine the inferred history of the Caputo derivative we require a general representation for $f_1(\tau)$. We will consider continuous functions and assume that $f_1(\tau)$ may be represented by a polynomial in τ , that is $f_1(\tau) = \sum_{i=0}^{\infty} b_i \tau^i$. Specifically, the Maclaurin series, or

$$f_1(\tau) = \sum_{n=0}^{\infty} \frac{f_1^{(n)}(0^-)}{\Gamma(n+1)} \tau^n, \quad \tau \leq 0. \quad (23)$$

will be used as the desired representation. Substituting Eq. (23) into the integrand of Eq. (22) and interchanging the order of integration and summation yields

$$\frac{d}{dt} \left\{ \sum_{n=0}^{\infty} \frac{f_1^{(n)}(0^-)}{\Gamma(n+1)} \int_a^0 (t-\tau)^{-\alpha} \tau^n d\tau \right\} = -t^{-\alpha} f(0^+), \quad t > 0. \quad (24)$$

A general solution for the definite integral can be derived [11] and substituting this result, we obtain

$$\begin{aligned} \frac{d}{dt} \left\{ \sum_{n=0}^{\infty} \frac{f_1^{(n)}(0^-)}{n!} \left[\sum_{i=1}^{n+1} \frac{(t-a)^{-\alpha+i} a^{n+1-i}}{\prod_{j=1}^i (-\alpha+j)} \frac{n!}{(n-i+1)!} - \frac{n! t^{-\alpha+n+1}}{\prod_{j=1}^{n+1} (-\alpha+j)} \right] \right\} \\ = -t^{-\alpha} f(0^+), \quad t > 0 \end{aligned} \quad (25)$$

Differentiating with respect to t gives

$$\begin{aligned} \left\{ \sum_{n=0}^{\infty} f_1^{(n)}(0^-) \left[\sum_{i=1}^{n+1} \frac{(i-\alpha)(t-a)^{-\alpha+i-1} a^{n+1-i}}{\prod_{j=1}^i (-\alpha+j)} \frac{1}{(n-i+1)!} - \frac{(-\alpha+n+1)t^{-\alpha+n}}{\prod_{j=1}^{n+1} (-\alpha+j)} \right] \right\} \\ = -t^{-\alpha} f(0^+), \quad t > 0. \end{aligned} \quad (26)$$

It is clear that only the $n = 0$ case on the LHS can match the exponent of t on the RHS, (the summation of the higher power terms, $t^{n-\alpha}$, cannot sum to a $t^{-\alpha}$ term), and because all derivatives $f_1^{(n)}(0^-) = 0$, $n \geq 1$ we have

$$f_1(0^-) \left[\frac{1}{(t-a)^\alpha} - t^{-\alpha} \right] = -t^{-\alpha} f(0^+), \quad t > 0. \quad (27)$$

Because the starting point of the initialization “ a ” does not occur on the RHS of Eq. (26), we must have $a = -\infty$, to force the first term of Eq. (27) to zero. Therefore, for $0 < \alpha < 1$ we have

$$f_1(0^-) = f(0^+), \text{ and from Eq. (23), } f_1(\tau) = f(0^+), \quad -\infty = a < \tau < 0. \quad (28)$$

Therefore, the only history that can make the Caputo derivative the same as the LH derivative, and that is tacitly assumed when evaluating a Caputo fractional derivative, for $0 < \alpha < 1$ is the “constant” function of time, that is

$$f(t) = f_1(t) = \text{constant} = f(0^+), \quad \text{for } -\infty = a \leq t \leq 0. \quad (29)$$

The above arguments can be extended to the case when $1 < \alpha < 2$ as outlined below. It follows from Eq. (17) that the Caputo derivative and the LH fractional derivative of the same order α and the same function $f(t)$ would be equal to each other if

$$\frac{d^2}{dt^2} \left\{ \frac{1}{\Gamma(2-\alpha)} \int_a^0 (t-\tau)^{1-\alpha} f_1(\tau) d\tau \right\} = -\frac{t^{1-\alpha} f'(0^+)}{\Gamma(2-\alpha)} - \frac{t^{-\alpha} f(0^+)}{\Gamma(1-\alpha)}, \quad t > 0. \quad (30)$$

Substituting as before from the McLaurin expansion in Eq. (23), and interchanging the order of integration and summation yields

$$\frac{d^2}{dt^2} \left\{ \sum_{n=0}^{\infty} \frac{f_1^{(n)}(0^-)}{\Gamma(n+1)} \int_a^0 (t-\tau)^{-\alpha} \tau^n d\tau \right\} = -t^{1-\alpha} f'(0^+) - (1-\alpha)t^{-\alpha} f(0^+), \quad t > 0. \quad (31)$$

Substituting the result of integration and performing the differentiation operation yields

$$\begin{aligned}
& \sum_{n=0}^{\infty} f_1^{(n)}(0^-) \left[\sum_{i=1}^{n+1} \frac{(i-\alpha)(1-\alpha+i)(t-a)^{-\alpha+i-1} a^{n+1-i}}{\prod_{j=1}^i (1-\alpha+j)} \frac{n!}{(n-i+1)!} \right] \\
& - \sum_{n=0}^{\infty} f_1^{(n)}(0^-) \left[\sum_{i=1}^{n+1} \frac{n!(2-\alpha+n)(-\alpha+n+1)t^{-\alpha+n}}{\prod_{j=1}^{n+1} (1-\alpha+j)} \right] \\
& = -t^{1-\alpha} f'(0^+) - (1-\alpha)t^{-\alpha} f(0^+), \quad t > 0.
\end{aligned} \tag{32}$$

It is clear that only the $n=0$ and $n=1$ cases can allow exponents of t that will match those of the terms on the RHS. It is required therefore that $f_1^{(n)}(0^-) = 0$ for all $n \geq 2$. Because the starting point of the initialization a does not occur on the RHS and because $1 < \alpha < 2$ we must have $a = -\infty$. Thus we must have

$$\begin{aligned}
& -\frac{f_1(0^-)(2-\alpha)(1-\alpha)}{(2-\alpha)} t^{-\alpha} - \frac{f_1^{(1)}(0^-)(3-\alpha)(2-\alpha)}{(2-\alpha)(3-\alpha)} t^{1-\alpha} \\
& = -t^{1-\alpha} f^{(1)}(0^+) - (1-\alpha)t^{-\alpha} f(0^+), \quad t > 0
\end{aligned} \tag{33}$$

Therefore we must have $f_1(0^-) = f(0^+)$, $f_1^{(1)}(0^-) = f_1^{(1)}(0^+)$. It follows from Eq. (23) that

$$f(t) = f(0) + f'(0)t \quad -\infty \leq t \leq 0, \tag{34}$$

for the case of $1 < \alpha < 2$. Both the results of Eq. (29) and Eq. (34) can be obtained as a special case of the more general result derived below.

5.2 General case: $m-1 < \alpha < m$

In this case [11] setting the two derivatives in Eq. (20) to be equal yields

$$\frac{d^m}{dt^m} \psi(f, -(m-\alpha), a, 0, t) = - \sum_{k=0}^{m-1} \frac{t^{k-\alpha} f^{(k)}(0^+)}{\Gamma(k-\alpha+1)}, t > 0, m-1 < \alpha < m. \quad (35)$$

Again for clarity of presentation we set $f = f_1$ in the initialization function. Then under the assumption of terminal initialization, this becomes

$$\frac{d^m}{dt^m} \frac{1}{\Gamma(m-\alpha)} \int_a^0 (t-\tau)^{(m-\alpha-1)} f_1(\tau) d\tau = - \sum_{k=0}^{m-1} \frac{t^{k-\alpha} f^{(k)}(0^+)}{\Gamma(k-\alpha+1)}, \quad (36)$$

$$t > 0, \quad m-1 < \alpha < m.$$

Again representing $f_1(\tau)$ as a continuous function by Eq. (23), gives

$$\frac{1}{\Gamma(m-\alpha)} \frac{d^m}{dt^m} \int_a^0 (t-\tau)^{(m-\alpha-1)} \sum_{n=0}^{\infty} \frac{f_1^{(n)}(0^-)}{\Gamma(n+1)} \tau^n d\tau \quad (37)$$

$$= - \sum_{k=0}^{m-1} \frac{t^{k-\alpha} f^{(k)}(0^-)}{\Gamma(k-\alpha+1)}, t > 0, m-1 < \alpha < m.$$

Interchanging the order of integration and summation, and substituting the result for the definite integral, and noting that the maximum value of the exponent of t on the RHS is $(m-1-\alpha)$, and hence, terms on the LHS, (after the m th order differentiation,) with exponents greater than this must have zero coefficients, that is $f_1^{(n)}(0^-) = 0$, for $n > m-1$, yields

$$\frac{1}{\Gamma(m-\alpha)} \sum_{n=0}^{m-1} \frac{f_1^{(n)}(0^-) t^{n-\alpha} \prod_{j=1}^m (n-\alpha+j)}{\prod_{j=1}^{n+1} (m-1-\alpha+j)} = \sum_{k=0}^{m-1} \frac{t^{k-\alpha} f^{(k)}(0^+)}{\Gamma(k-\alpha+1)}, \quad (38)$$

$$t > 0, m-1 < \alpha < m.$$

In general, the equality will only hold when

$$f_1^{(i)}(0^-) = f^{(i)}(0^+), \quad i = 0, 1, \dots, m-1. \quad (39)$$

Placing these results into Eq. (23) we have, for $m-1 < \alpha < m$

$$f(t) = \sum_{n=0}^{m-1} \frac{f^{(n)}(0^+)}{\Gamma(n+1)} t^n, \quad -\infty = a \leq t \leq 0, \quad (40)$$

as the only allowable initialization that will make the Caputo derivative equal to the LH derivative. Thus in general, we find for $m-1 < \alpha < m$, that the Caputo derivative infers a history in the form of a polynomial in t back to $-\infty$ with maximum order of $m-1$. The coefficients of the polynomial are related to the values of the integer-order derivatives evaluated at $t=0^+$. It is also observed that derivatives of $f(t)$ with order higher than order $m-1$ will in general be discontinuous at $t=0$. Eq. (40) yields Eq. (29) and Eq. (34) as special cases.

5.3 Example

A simple example will illustrate the profound differences between the Caputo derivative and the LH initialization of the RL derivative. Consider the semi-derivative of $f(t) = (t+2)^2$ with a history period starting at $t=a=-2$ (inferring $f(t)=0$ for $t < a=-2$), and with the differentiation of interest starting at $t=0$.

The Caputo derivative is given by

$${}_0^C d_t^{1/2} (t+2)^2 = \frac{1}{\Gamma(\frac{1}{2})} \int_0^t (t-\tau)^{-1/2} \frac{\partial}{\partial \tau} (\tau+2)^2 d\tau = \frac{1}{\Gamma(\frac{1}{2})} \left[\frac{8}{3} t^{3/2} + 8t^{1/2} \right], t > 0 \quad (41)$$

which has removed the effect of the singularity at $t=0$. Because $\alpha=1/2$, we have by Eq. (29) the inferred initialization function (history for Caputo derivative) given as $f(t) = f(0^+) = 4$ for $-\infty \leq t \leq 0$.

We now consider the LH initialization of the RL derivative for terminal initialization of the function $f(t) = (t+2)^2$

$$\begin{aligned}
{}_0D_t^{1/2}(t+2)^2 &= D_t^1 \left({}_0D_t^{-1/2}(t+2)^2 \right) \\
&= \frac{d}{dt} \left({}_0d_t^{-1/2}(t+2)^2 + \psi(f, -1/2, -2, 0, t) \right) + 0, t > 0,
\end{aligned} \tag{42}$$

$${}_0D_t^{1/2}(t+2)^2 = \frac{d}{dt} \left(\int_0^t \frac{(t-\tau)^{\frac{1}{2}-1}}{\Gamma(\frac{1}{2})} (\tau+2)^2 d\tau + \int_{-2}^0 \frac{(t-\tau)^{\frac{1}{2}-1}}{\Gamma(\frac{1}{2})} (\tau+2)^2 d\tau \right), t > 0, \tag{43}$$

Integrating, collecting terms, and simplifying yields

$${}_0D_t^{1/2}(t+2)^2 = \frac{40(t+2)^{3/2}}{15\Gamma(\frac{1}{2})}, \quad t > 0 \tag{44}$$

The results of Eq. (41) and Eq. (44), are shown graphically in Figure 1. It is clear from the figure that the Caputo inferred history has discontinuous integer order derivatives at $t=0$, while the chosen LH initialization is a smooth continuation of the function being integrated. The difference in the behavior of the two derivatives is profound for t significantly larger than zero! For t much larger than zero the derivatives will have a common functional form, namely $\propto t^{3/2}$. Also shown in Figure 1 is the uninitialized LH semi-derivative of the function starting at $t=a=-2$, namely ${}_2D_t^{1/2}(t+2)^2$. It is of course a smooth (backward) continuation of ${}_0D_t^{1/2}(t+2)^2$. It is also noted that the LH semi-derivative using the Caputo inferred initialization ($f(t)=4, -\infty < t < 0$) is the same as the Caputo semi-derivative as expected.

5.4 Caputo derivative in application

Here we examine application of the inferred history of the Caputo derivative developed in earlier sections to fractional differential equations to gain further insight into the initialization issues associated with the Caputo derivative.

Case 1

Suppose we consider only fractional differential equations of the form

$${}_0^C d_t^\alpha f(t) = y(t), \tag{45}$$

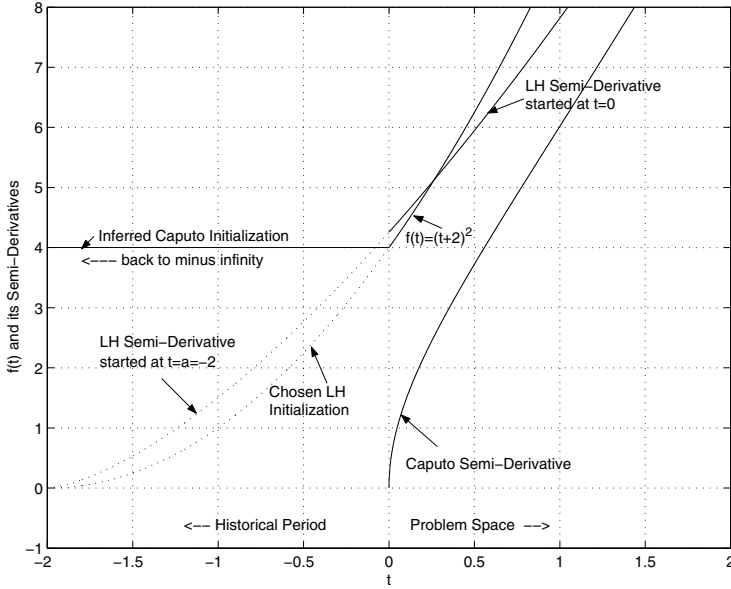


Fig. 1. Initialization for Caputo and LH semi-derivatives of $(t+2)^2$.

Use of the Caputo derivative and its inferred α – order history, Eq. (34), *may be acceptable* if i) it is found that the history acceptable to the physics defining the problem and ii) if it is acceptable to have discontinuity of the derivatives of $f(t)$ of order m and greater at $t=0$, where m is defined by $m-1 < \alpha < m$.

Case 2

Suppose we now consider the following fractional differential equation involving more than a single fractional derivative:

$${}_0^C D_t^{2\alpha} f(t) + {}_0^C D_t^\alpha f(t) + f(t) = y(t), \quad 1/2 < \alpha < 1. \quad (46)$$

Suppose we stipulate that both fractional derivative terms have the same history, i.e., $f(t)$, $t < 0$. Thus for this case, the orders of the terms lie between different integers, that is for the 2α -order derivative the order is $1 < 2\alpha < 2$ and for the α -order derivative it is $0 < \alpha < 1$. Thus based on the results of the

previous section, the allowable history for the α -order term is $f(t) = f(0)$ for $-\infty \leq t \leq 0$ from Eq. (29) and the allowable history for the 2α -order term is $f(t) = f(0) + f'(0)t$, $-\infty \leq t \leq 0$, from Eq (34). Clearly the only history that will satisfy both is $f(t) = f(0)$, $-\infty \leq t \leq 0$ with $f'(0) = 0$. But this forces the function being differentiated, that is, $f(t)$, $t > 0$ to have $f'(0) = 0$ which in general may not be the case.

If on the other hand, the individual terms in the differential equation of Eq. (46) each have separate and independent histories. This means that each differential term in the equation is disconnected from the other differential terms, and is acted on by its own individual input (history) in negative time. Then at time zero, all of the individual differential terms are connected together, with the requirement that all of the individual positions, velocities, etc. (including fractional derivatives), have the same values at time zero. Under this scenario the initialization of each term reverts the situation of Case 1. Thus the Caputo derivative can be used in the description of Eq. (46) if it is assumed that all derivative terms (elements) have separate and independent histories, are compatibly connected at time $t = 0$, and that each derivative term according to the magnitude of its order has the history specified by Eq. (40). That is, the inferred initialization of each term depends on the order of that term and the limitations of Case 1 apply.

6 Summary

Using the LH initialization of the RL derivative it is shown that the α -order Caputo derivative with $m-1 < \alpha < m$ infers a history in the form of a polynomial with maximum order of the polynomial being $m-1$ and given by

$$f(t) = \sum_{n=0}^{m-1} \frac{f^{(n)}(0^+)}{\Gamma(n+1)} t^n, \quad m-1 < \alpha < m, \quad a \leq t \leq 0.$$

implying that integer order derivatives of $f(t)$ with order higher than $m-1$ will in general be discontinuous at $t = 0$. While it has been known for quite sometime that the Caputo derivative is more restrictive than the RL derivative [13], it is now clear that the Caputo derivative can not represent *general* initializations required for most analysis, physics, and engineering problems.

References

1. Lorenzo CF, Hartley TT (2000) Initialized fractional calculus. *Int. J. Appl. Math.* 3(3):249–265.
2. Lorenzo CF, Hartley TT (2001) Initialization in Fractional Order Systems, in: *Proceedings of the European Control Conference*, Porto, Portugal pp. 1471–1476.
3. Oldham KB, Spanier J (1974) *The Fractional Calculus- Theory and Applications of Differentiation and Integration to Arbitrary Order*. Academic Press, New York.
4. Miller KS, Ross B (1993) *An Introduction to the Fractional Calculus and Fractional Differential Equations*. Wiley, New York.
5. Samko SG, Kilbas AA, Marichev OI (1993) *Fractional Integrals and Derivatives: Theory and Applications*. Gordon and Breach Science Publishers, Philadelphia, PA.
6. Podlubny L (1999) *Fractional Differential Equations*. Academic Press, San Diego, CA.
7. Lorenzo CF, Hartley TT (1998) Initialization, Conceptualization, and Application in the Generalized Fractional Calculus. NASA TM-1998.
8. Caputo M (1969) *Elasticita e Dissipazione*. Zanichelli, Bologna.
9. Caputo M (1967) Linear Model of dissipation whose Q is almost frequency independent-II. *Geophys. J. R. Astron. Soc.* 13:529–539.
10. Gorenflo R, Mainardi F (1997) Fractional calculus, integral and differential equations of fractional order in: *Fractals and Fractional Calculus in Continuum Mechanics*. Carpeneri, A, Mainardi, F (eds.), Springer, New York.
11. Achar BNN, Lorenzo CF, Hartley T (2003) Initialization and the Caputo Fractional Derivative. NASA TM-2003.
12. Achar BN, Narahari, Lorenzo CF, Hartley T (2005) Initialization Issues of the Caputo Fractional Derivative, *Proceedings of IDETC/CIE 2005*, Sept. 24–28, Long Beach CA. DETC2005. pp. 1–8.
13. Mainardi F, Gorenflo R (2000) On Mittag-Leffler -type functions in fractional evolution processes. *J. Comp. Appl. Math.* 118:283–299.

COMPARISON OF FIVE NUMERICAL SCHEMES FOR FRACTIONAL DIFFERENTIAL EQUATIONS

Om Prakash Agrawal¹ and Pankaj Kumar²

¹ Mechanical Engineering, Southern Illinois University, Carbondale, IL 62901;
E-mail: om@engr.siu.edu

² Mechanical Engineering, Southern Illinois University, Carbondale, IL 62901;
E-mail: pkumar@engr.siu.edu

Abstract

This paper presents a comparative study of the performance of five different numerical schemes for the solution of fractional differential equations. The schemes considered are a linear, a quadratic, a cubic, a state-space non-integer integrator, and a Direct discretization method. Results are presented for five different problems which include two linear 1-D, two nonlinear 1-D and one linear multidimensional. Both homogeneous and nonhomogeneous initial conditions (ICs) are considered. The stability, accuracy, and computational speeds for these algorithms are examined. Numerical simulations exhibit that the choice of a numerical scheme will depend on the problem considered and the performance criteria selected.

Keywords

Fractional differential equations, fractional derivatives, numerical schemes for fractional differential equations, Volterra integral equation, Grünwald–Letnikov approximation.

1 Introduction

Fractional derivatives (FDs) and fractional integrals (FIs) have received considerable interest in recent years. In many applications, FDs and FIs provide more accurate models of the systems than ordinary derivatives and integrals do. Many applications of FDs and FIs in the areas of solid mechanics and modeling of viscoelastic damping, electrochemical processes, dielectric polarization, colored noise, bioengineering, and various branches of science and engineering could be found, among others, in [1, 2, 3, 4, 5, 6, 7, 8]. Analysis and design of many of the systems require solution of fractional differential equations (FDEs) [3]. Several methods have recently been proposed

to solve these equations. These methods include Laplace and Fourier transforms [3, 9, 10], eigenvector expansion [11], method based on Laguerre integral formula [12], direct solution based on Grünwald–Letnikov approximation [3], truncated Taylor series expansion [13], diffusive representation method [14], approximate state-space representations [15, 16], and numerical methods [17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27]. Sabatier and Malti [28] grouped the FDEs appearing in the above applications into number of classes to develop a benchmark to evaluate performance of numerical algorithms. They also presented numerical techniques and results for each group of FDEs.

This brief review of applications of FDs and various analytical and numerical techniques to solve FDEs is by no means complete. Examples of many other applications could be found in [3, 4, 6, 7, 8] and the references there in, and many other analytical and numerical schemes to solve the FDEs are cited in [21, 21, 26, 27].

This paper presents a comparative study of the performance of five different numerical schemes to solve FDEs, namely, the linear [21, 22], the quadratic [25], the cubic [26], the direct method based on the Grünwald–Letnikov approximation [3], and the state-space approximation of the fractional integrator [15, 16]. For completeness, the associated algorithms/procedures are discussed briefly. Details of these algorithms could be found in the references cited above. The issues investigated include the numerical stability, accuracy, and computational times for these algorithms. Numerical results for five examples, two linear one dimensional, two nonlinear one-dimensional, and one-linear multidimensional are presented. Both homogeneous and nonhomogeneous boundary conditions are considered.

2 Notations and Definitions

We begin with the Riemann–Liouville definition of the *fractional integral* of order $\alpha > 0$, which is given as

$$I^\alpha y(t) = \frac{1}{\Gamma(\alpha)} \int_0^t (t - \tau)^{\alpha-1} y(\tau) d\tau, \quad (\alpha > 0), \quad (1)$$

where Γ is the gamma function. For integer $\alpha > 0$, Eq. 1 is known as the Cauchy integral formula. Here we take the lower limit of the integral as 0, however, a nonzero limit can also be taken. It can be verified that the integral operator I^α commutes, i.e.,

$$I^\alpha I^\beta y(t) = I^\beta I^\alpha y(t) = I^{\alpha+\beta} y(t) \quad \alpha, \beta > 0. \quad (2)$$

We will largely deal with Caputo fractional derivatives (CFDs). However, we will also come across the Riemann–Liouville fractional derivatives (RLFDs). These two derivatives are given as:

Caputo Fractional Derivative (CFD)

$$D_{\star}^{\alpha}y(t) = I^{n-\alpha}D^n y(t) = \frac{1}{\Gamma(n-\alpha)} \int_0^t (t-\tau)^{n-\alpha-1} \left(\frac{d}{d\tau}\right)^n y(\tau) d\tau, \quad (3)$$

Riemann–Liouville Fractional Derivative (RLFD)

$$D^{\alpha}y(t) = D^n I^{n-\alpha}y(t) = \frac{1}{\Gamma(n-\alpha)} \left(\frac{d}{dt}\right)^n \int_0^t (t-\tau)^{n-\alpha-1} y(\tau) d\tau. \quad (4)$$

where $\alpha > 0$, n is the smallest integer greater than or equal to α , and the operator D^n is the ordinary differential operator. These two derivatives are related by the formula

$$D^{\alpha}y(t) = D_{\star}^{\alpha}y(t) + \sum_{i=0}^{n-1} \frac{t^{i-\alpha}}{\Gamma(i-\alpha+1)} y^{(i)}(0^+). \quad (5)$$

Observe that for zero ICs, the two derivatives are the same. Thus, for this condition we may switch between the two derivatives as necessary.

3 Statement of the Problem

We consider the following FDEs and the ICs

$$D_{\star}^{\alpha}y(t) = f(t, y(t)), \quad (6)$$

$$y^{(i)}(0) = y_0^{(i)}, \quad i = 1, \dots, n-1. \quad (7)$$

Observe that here we consider the FDE in terms of Caputo derivatives. This allows us to account for physical initial conditions. Equations 6 and 7 are applicable for both scalar and vector y . In the discussions to follow, we use scalar y to derive an equation. However, when solving a problem in which y is a vector, we will use vector equivalent of the formulation without explicitly writing these equations. Note that [29] discusses the problem of finding the correct form of the initial conditions in a more general setting, not necessarily assuming that the entire history of the process can be observed. A similar treatment for the Caputo derivative is presented in [30].

Applying the operator I^{α} to Eq. 6, and using Eqs. 1, 2, 3, and 7, we obtain

$$y(t) = g(t) + \frac{1}{\Gamma(\alpha)} \int_0^t (t-\tau)^{\alpha-1} f(\tau, y(\tau)) d\tau, \quad (8)$$

where

$$g(t) = \sum_{i=0}^{n-1} y_0^{(i)} \frac{t^i}{i!}. \quad (9)$$

Equation 8 is a *Volterra integral equation*, and it plays a significant role in three of the schemes discussed below.

4 The Numerical Schemes

In this section, we briefly review the five numerical schemes stated above. The first three schemes essentially attempts to solve Eq. 8, the direct scheme approximates the fractional derivative terms, and the state-space scheme is based on the state-space approximation of a fractional integral operator.

4.1 The linear scheme

This scheme, presented by Diethelm, Ford, and Freed, is also called $P(EC)E$ and $P(EC)^ME$ schemes, where P , E , and C stands for predict, evaluate, and correct, and M represents the iteration number [21, 22]. The difference between the $P(EC)E$ and $P(EC)^ME$ schemes is that in the former scheme only one corrective step is taken whereas in the later scheme multiple corrective steps are taken.

Let T be the maximum simulation time. To explain the scheme, divide the time T into N equal parts, and let $h = T/N$ be the time interval of each part. The times at the grid points are given as $t_j = jh$, $j = 0, \dots, N$. For simplicity in the discussion to follow, we use the following notations: $y(t_j) = y(jh) = y_j$, $g(t_j) = g(jh) = g_j$, and $f(t_j, y(t_j)) = f(jh, y(jh)) = F_j$. Note that in many numerical analysis papers $y(t_j)$ and y_j represent the true and the numerically computed value of y at t_j . No such distinction is made here. Where such distinction is necessary, the true and the computed values are explicitly identified.

Now assume that the approximate numerical values for $y(t)$ have been determined at the grid points t_j , $j = 0, \dots, m$, $t_j < T$. Assuming that y and $f(t, y(t))$ vary linearly over each part and using Eq. 8, y_{m+1} is given as [21]

$$y_{m+1} = g_{m+1} + \frac{h^\alpha}{\Gamma(\alpha + 2)} \sum_{j=0}^{m+1} a_{j,m+1} F_j, \quad (10)$$

where

$$a_{j,m+1} = \begin{cases} m^{\alpha+1}(m-\alpha)(m+1)^\alpha, & \text{if } j = 0, \\ (m-j+2)^{\alpha+1} + (m-j)^{\alpha+1}2(m-j+1)^{\alpha+1}, & \text{if } 1 \leq j \leq m, \\ 1, & \text{if } j = m+1. \end{cases} \quad (11)$$

and $m = 0, \dots, N-1$. Note that y_{m+1} appears on both sides of Eq. 11, which for nonlinear $f(t, y(t))$ leads to a nonlinear equation. To solve this equation, [21] describes a $P(EC)E$ type scheme in which at the prediction step the integral in Eq. 8 is approximated using a product rectangular rule, and Eq. 10 is used to correct the values. In [22], this iteration is continued several times, except that after first iteration the value evaluated using Eq. 10 is used as the predicted value for the subsequent iteration. This is essentially equivalent to a fixed point iteration. The details of the algorithms can be found in [21, 22].

These authors also present a Richardson extrapolation-type scheme to further improve the accuracy of the results. Here we take a slightly different approach. For linear case, we solve Eq. 10 explicitly, and for nonlinear case we solve it using the Newton–Raphson scheme for which we take y_m as the starting guess for y_{m+1} .

Note that in this class of schemes y and $f(t, y(t))$ are approximated using linear functions, and therefore we call them the linear schemes.

4.2 The quadratic scheme

In this scheme, N is taken as an even number, and y and $f(t, y(t))$ are approximated over two adjacent parts using quadratic polynomials. Assume that y_j , $j = 1, \dots, 2m$ have already been computed. Using Eq. 8, the expressions for y_{2m+1} and y_{2m+2} are given as

$$\begin{aligned} y_{2m+1} = & g_{2m+1} + \frac{1}{\Gamma(\alpha)} \int_0^{2mh} ((2m+1)h - \tau)^{\alpha-1} f(\tau, y(\tau)) d\tau \\ & + \frac{1}{\Gamma(\alpha)} \int_{2mh}^{(2m+1)h} ((2m+1)h - \tau)^{\alpha-1} f(\tau, y(\tau)) d\tau \end{aligned} \quad (12)$$

and

$$\begin{aligned} y_{2m+2} = & g_{2m+2} + \frac{1}{\Gamma(\alpha)} \int_0^{2mh} ((2m+2)h - \tau)^{\alpha-1} f(\tau, y(\tau)) d\tau \\ & + \frac{1}{\Gamma(\alpha)} \int_{2mh}^{(2m+2)h} ((2m+2)h - \tau)^{\alpha-1} f(\tau, y(\tau)) d\tau \end{aligned} \quad (13)$$

Since y_j , $j = 0, \dots, 2m$ are known, the first integrals in both Eqs. 12 and 13 can be computed explicitly. To compute the second integral in Eq. 13, $f(t, y(t))$ is approximated over $[2mh, (2m+2)h]$ in terms of F_{2m} , F_{2m+1} , and F_{2m+2} , as

$$f(t, y(t)) = \sum_{j=0}^2 \phi_j(t) F_{2m+j} \quad (14)$$

where $\phi_j(t)$, $j = 0, 1$, and 2 are the quadratic interpolating polynomials (QIPs), which is 1 at node $2m+j$ and 0 at the two other nodes. Substituting Eq. 14 into Eq. 13, we obtain y_{2m+2} in terms of F_{2m+1} , and F_{2m+2} . Note that F_{2m} is not included here as it can be computed directly from y_{2m} . To compute the second integral in Eq. 12, $f(t, y(t))$ is approximated over $[2mh, (2m+1)h]$ in terms of F_{2m} , $F_{2m+1/2}$ and F_{2m+1} using QIPs similar to the one used in Eq. 14. Using Eq. 14, $F_{2m+1/2}$ is expressed in terms of F_{2m} , F_{2m+1} , and F_{2m+2} . This leads to y_{2m+1} in terms of F_{2m+1} , and F_{2m+2} . Thus, we obtain two equations in terms of two unknowns y_{2m+1} and y_{2m+2} , which are solved using the Newton–Raphson method. The details of the algorithm can be found in [25].

4.3 The cubic scheme

In this scheme N is taken as a multiple of 3, y , and $f(t, y(t))$ are approximated over three adjacent parts using cubic polynomials, and expressions are generated for y_{3m+1} , y_{3m+2} and y_{3m+3} in terms of F_{3m+1} , F_{3m+2} and F_{3m+3} . These expressions are solved using the Newton–Raphson method as before. For brevity, the details of the algorithm is omitted here, and the readers are referred to [26] where further details can be found.

4.4 The direct scheme

To explain this scheme, assume that y_j , $j = 0, \dots, m$ have already been computed, and we want to compute y_{m+1} . In the direct scheme, the CFD in Eq. 6 is first replaced with the RLFD using Eq. 5 and then the RLFD is approximated at t_{m+1} using a Grünwald–Letnikov definition [3]. This leads to

$$h^{-\alpha} \sum_{j=0}^{m+1} w_j^\alpha y_{(m+1-j)} = F_{m+1} - \sum_{i=0}^{n-1} \frac{((m+1)h)^{i-\alpha}}{\Gamma(i-\alpha+1)} y^{(i)}(0^+). \quad (15)$$

where the coefficients w_j^α satisfy the following recurrence relationship,

$$w_0^\alpha = 1, \quad w_j^\alpha = \left(1 - \frac{1+\alpha}{j}\right) w_{j-1}^\alpha, \quad j = 1, 2, \dots \quad (16)$$

Note that the CFD can be approximated directly using a slightly different scheme (see [24]), the approach considered here is believed to be computationally efficient. For nonlinear $f(t, y(t))$, Eq. 15 leads to a nonlinear equation in terms of y_{m+1} which is solved using the Newton–Raphson method as before. Note that if $1 < \alpha < 2$, then y_1 is computed as

$$y_1 = y_0 + y_0^{(1)} h \quad (17)$$

Similar modifications are made if α is greater than 2. The details of the algorithm could be found in [3].

4.5 The state-space non-integer integrator

For $\alpha > 0$, consider a fractional integral operator $1/s^\alpha$ (also known as the Laplace operator), where s is the Laplace parameter. The basic idea behind this scheme is to approximate $1/s^\alpha$ in terms of a set of integer-order integrators and phase lead filters. Thus, the integral operator is written as [16],

$$\frac{1}{s^\alpha} = \frac{G_\alpha}{s} + \sum_{i=1}^N \frac{c_i}{s + w_i} \quad (18)$$

where $N + 1$ is the number of states considered, and G_α , c_i and w_i are coefficients which depend on the frequency range of the application, the number of state variables considered, and the order of the derivatives. Several schemes have been presented that model a fractional integrator using this or a similar technique (see, e.g. [15, 16]). Here, we use the scheme described in [16].

5 Numerical Results

In this section, we present numerical results for 5 examples obtained using the five schemes, namely the linear, the quadratic, the cubic, the state-space non-integer integrator, and the direct discretization. For simplicity in the discussion to follow, these schemes will be called schemes S1, S2, S3, S4, and S5, respectively. The examples considered include two linear one-dimensional, two nonlinear one-dimensional, and one linear multidimensional. These examples have also been considered by other investigators. All problems were solved on a 2.80 GHz Pentium 4 desktop computer that had 1 GB of ram and Microsoft Windows XP, service pack 2 operating system. The algorithms were developed and solved using Matlab 7.0. For scheme S4, a fractional integrator block was developed in Simulink, and the examples were solved using the Simulink block diagrams. The problems were solved for several orders of the FDs ranging between 0 and 2 using different values of h .

We have generated a large volume of data/results. Because of space limit, not all of the data can be presented here. For scheme S4, the default integration scheme of the Simulink with default relative and absolute error was used, and the results were generated for $\alpha = 0.5$ and 1.5 only. Since Simulink controls the error internally, a maximum step size was specified, and it was allowed to compute the step size internally. In each case, except for Example 3, the error is computed as the difference between the numerical and the analytical solutions.

In the results below, we consider several combinations of α and h to maximize the spectrum of data presented. In the figures showing the numerical results, the symbols plus, multiplication, circle, triangle, square, and diamond will represent the results obtained analytically and using the five schemes S1 to S5, respectively. In each example, the second IC is for $1 < \alpha < 2$. To avoid repetition, we present first all the results, and the interpretation of the results.

5.1 Example 1

As the first example we consider the following linear FDE and the homogeneous initial conditions (ICs)

$$D_*^\alpha y(t) = 1 - y(t) \quad (19)$$

$$y(0) = 0, \quad \dot{y}(0) = 0 \quad (20)$$

The close form solution for this problem is given as

$$y(t) = E_{\alpha,1}(-t^\alpha) \quad (21)$$

where

$$E_{\alpha,\beta}(z) = \sum_{j=0}^{\infty} \frac{z^j}{\Gamma(\alpha j + \beta)} \quad (22)$$

is the generalized Mittag-Leffler function. The example is solved for $T = 6.4$ sec. Figure 1 shows the analytical and numerical results for $y(t)$ for $\alpha = 0.5$ and $h = 0.2$ (left) and for $\alpha = 1.25$ and $h = 0.0125$ (right). Table 1 compares the numerical errors for different schemes for $(\alpha = 0.5, h = 0.0125)$ and $(\alpha = 1.25, h = 0.2)$. The maximum errors for different schemes for $\alpha = 1.5$ and different values of h are given in Table 2. The CPU times for these schemes for $\alpha = 1.5$ and 0.75 and different values of h are given in Table 3.

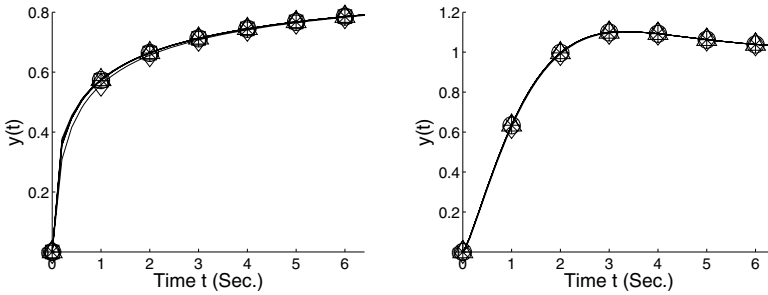


Fig. 1. Comparison of $y(t)$ obtained using different schemes for example 1. (Left: $\alpha = 0.5, h = 0.2$; Right: $\alpha = 1.25, h = 0.0125$.)

Table 1. Comparison of errors in $y(t)$ at different times for example 1

	$\alpha = 0.5, h = 0.0125$					$\alpha = 1.25, h = 0.2$				
t	S1	S2	S3	S4	S5	S1	S2	S3	S5	
0.8	5.88e-5	2.25e-5	2.83e-5	-8.85e-5	-1.11e-3	-5.76e-4	-1.77e-4	-2.87e-4	-1.44e-1	
1.6	2.74e-5	1.6e-5	1.33e-5	-2.09e-5	-5.81e-4	3.73e-4	-7.61e-5	-1.75e-4	-1.17e-1	
2.4	1.69e-5	6.59e-6	8.24e-6	-4.78e-6	-3.81e-4	8.26e-4	-2.12e-5	-7.16e-5	-7.24e-2	
3.2	1.19e-5	4.63e-6	5.78e-6	-4.01e-6	-2.77e-4	7.47e-4	1.30e-6	-1.29e-5	-3.33e-2	
4.0	8.91e-6	3.49e-6	4.35e-6	-2.64e-6	-2.14e-4	4.54e-4	7.36e-6	1.02e-5	-7.78e-3	
4.8	7.02e-6	2.75e-6	3.43e-6	-2.41e-6	-1.72e-4	1.79e-4	6.98e-6	1.49e-5	4.90e-3	
5.6	5.72e-6	2.25e-6	2.83e-6	-2.32e-6	-1.43e-4	7.77e-6	5.01e-6	1.21e-5	8.92e-3	
6.4	4.78e-6	1.88e-6	2.30e-6	-1.66e-6	-1.21e-4	-6.42e-5	3.22e-6	7.83e-6	8.44e-3	

Table 2. Comparison of maximum errors in $y(t)$ for example 1 for $\alpha = 1.5$

h	S1	S2	S3	S4	S5
0.2	1.56e-3	-1.87e-4	-1.98e-4	6.12e-6	-1.60e-1
0.1	3.94e-4	-2.71e-5	-3.57e-5	6.12e-6	-8.25e-2
0.05	-9.91e-5	-4.46e-6	-6.37e-6	6.12e-6	-4.19e-2
0.025	-2.53e-5	-7.69e-7	-1.13e-6	6.12e-6	-2.11e-2
0.0125	-6.44e-6	-1.35e-7	-2.00e-7	6.12e-6	-1.06e-2
0.00625	-1.63e-6	-2.38e-8	1.65e-6	6.12e-6	-5.30e-3
0.003125	-4.10e-7	-1.66e-8	1.77e-5	6.12e-6	-2.65e-3

Table 3. Comparison the CPU times in seconds for example 1

	$\alpha = 1.5$					$\alpha = 0.75$				
h	S1	S2	S3	S4	S5	S1	S2	S3	S5	
0.2	0.47	0.30	0.47	21.48	0.03	0.61	0.23	0.53	0.03	
0.1	1.30	0.75	1.08	21.59	0.02	1.66	0.70	1.13	0.05	
0.05	4.31	2.83	3.52	21.98	0.03	4.59	2.86	3.38	0.02	
0.025	15.02	11.20	11.64	22.27	0.06	15.34	10.64	11.45	0.06	
0.0125	56.02	43.02	44.44	22.17	0.14	56.08	42.39	43.42	0.14	
0.00625	218.39	172.06	170.47	22.25	0.33	212.45	168.61	167.41	0.28	
0.003125	856.41	695.05	668.06	20.53	0.78	836.39	674.67	659.41	0.75	

5.2 Example 2

As the second example we consider the following linear FDE and the inhomogeneous ICs

$$D_{\star}^{\alpha}y(t) = 0.1t - y(t) \quad (23)$$

$$y(0) = 1, \quad \dot{y}(0) = 0 \quad (24)$$

The analytical solution for this problem is given as

$$y(t) = 0.1t(1 - E_{\alpha,2}(-t^{\alpha})) + E_{\alpha,1}(-t^{\alpha})y(0). \quad (25)$$

Figure 2 compares the results for $y(t)$ for various schemes for $\alpha = 0.25$ and $h = 0.1$ (left) and for $\alpha = 1.5$ and $h = 0.00625$ (right). Table 4 presents errors for various schemes for $(\alpha = 0.25, h = 0.00625)$ and $(\alpha = 1.5, h = 0.1)$. The maximum errors in $y(t)$ for different schemes for $\alpha = 0.5$ and different h are given in Table 5. Table 6 compares the CPU times for different schemes $\alpha = 0.5$ and 1.75 and different h .

5.3 Example 3

As the third example we consider the following nonlinear FDE and the inhomogeneous ICs

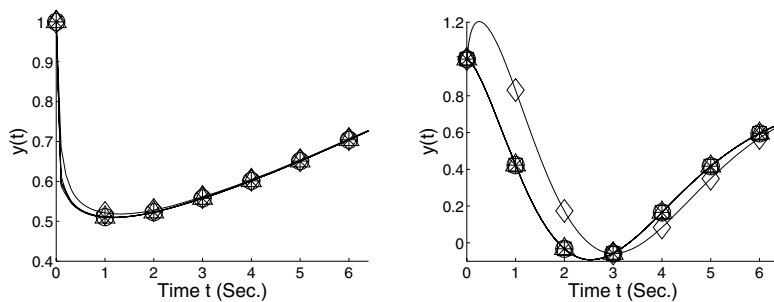


Fig. 2. Comparison of $y(t)$ obtained using different schemes for example 2. (Left: $\alpha = 0.25$, $h = 0.1$; Right: $\alpha = 1.5$, $h = 0.00625$.)

Table 4. Comparison of errors in $y(t)$ at different times for example 2

	$\alpha = 0.25$, $h = 0.00625$				$\alpha = 1.5$, $h = 0.1$				
t	S1	S2	S3	S5	S1	S2	S3	S4	S5
0.8	-5.01e-5	-2.76e-5	-3.29e-5	1.06e-3	3.53e-4	2.63e-5	3.57e-5	1.26e-6	1.61e-1
1.6	-2.46e-5	-1.36e-5	-1.62e-5	5.34e-4	1.77e-4	1.60e-5	2.84e-5	-2.53e-6	1.65e-1
2.4	-1.61e-5	-8.91e-6	-1.06e-5	3.58e-4	-1.96e-4	3.65e-6	1.19e-5	-5.67e-6	1.05e-1
3.2	-1.19e-5	-6.58e-6	-7.83e-6	2.70e-4	-4.07e-4	-3.73e-6	-1.06e-6	-4.60e-6	3.30e-2
4.0	-9.39e-6	-5.20e-6	-6.17e-6	2.17e-4	-3.58e-4	-5.52e-6	-6.89e-6	-2.75e-6	-1.96e-2
4.8	-7.73e-6	-4.28e-6	-4.91e-6	1.82e-4	-1.55e-4	-3.98e-6	-6.92e-6	-8.23e-7	-4.20e-2
5.6	-6.69e-6	-3.76e-6	-4.72e-6	1.57e-4	4.62e-5	-1.62e-6	-4.24e-6	1.86e-6	-3.94e-2
6.4	-9.55e-6	-7.02e-6	-7.25e-6	1.35e-4	1.50e-4	7.68e-8	-1.37e-6	3.47e-6	-2.41e-2

Table 5. Comparison of maximum errors in $y(t)$ for example 2 for $\alpha = 0.5$

h	S1	S2	S3	S4	S5
0.2	-2.15e-2	-1.40e-2	-1.10e-2	1.34e-4	9.26e-2
0.1	-1.19e-2	-7.77e-3	-6.15e-3	1.18e-4	8.54e-2
0.05	-6.37e-3	-4.19e-3	-3.33e-3	1.36e-4	7.95e-2
0.025	-3.35e-3	-2.21e-3	-1.76e-3	2.38e-5	7.51e-2
0.0125	-1.74e-3	-1.15e-3	-9.20e-4	1.89e-4	7.18e-2
0.00625	-8.90e-4	-5.92e-4	-4.74e-4	9.17e-5	6.95e-2
0.003125	-4.53e-4	-3.02e-4	-2.42e-4	1.93e-4	6.79e-2

Table 6. Comparison the CPU times in seconds for example 2

	Set1					Set2			
h	S1	S2	S3	S4	S5	S1	S2	S3	S5
0.2	0.70	0.27	0.48	22.48	0.03	0.55	0.27	0.52	0.03
0.1	1.64	0.86	1.16	22.31	0.00	1.63	0.77	1.17	0.00
0.05	5.25	2.94	3.63	22.95	0.03	5.06	2.91	3.78	0.03
0.025	18.13	11.88	12.27	22.58	0.05	17.63	11.64	13.05	0.02
0.0125	66.47	46.70	45.70	22.55	0.05	65.33	45.61	46.89	0.05
0.00625	260.98	183.64	174.31	22.67	0.11	250.75	181.56	178.63	0.11
0.003125	992.20	741.25	696.95	22.50	0.31	986.00	725.41	705.02	0.31

$$D_*^\alpha y(t) = 1 - y^2(t) \quad (26)$$

$$y(0) = \sqrt{2}, \quad \dot{y}(0) = -1 \quad (27)$$

Figure 3 compares the results for $y(t)$ for various schemes for $\alpha = 0.5$ and $h = 0.05$ (left) and for $\alpha = 1.75$ and $h = 0.003125$ (right). Since analytical solutions were not available, we considered the numerical results for $h = 0.00625$ obtained using the cubic method as the reference value and compute the error as the difference between the numerical solution and the reference solution. Table 7 presents errors for various schemes for $(\alpha = 0.5, h = 0.05)$ and $(\alpha = 1.75, h = 0.1)$. The maximum errors in $y(t)$ for different schemes for $\alpha = 1.5$ and different h are given in Table 8. Table 9 compares the CPU times for $\alpha = 1.5$ and different h .

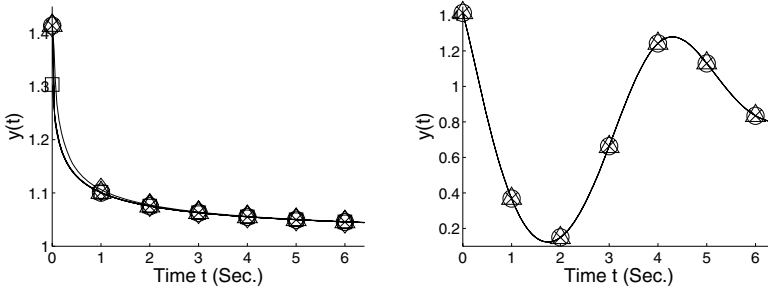


Fig. 3. Comparison of $y(t)$ obtained using different schemes for example 3. (Left: $\alpha = 0.5, h = 0.05$; Right: $\alpha = 1.75, h = 0.003125$.)

5.4 Example 4

As example 4 we consider the following nonlinear FDE and the homogeneous ICs

Table 7. Comparison of errors in $y(t)$ at different times for example 3

	$\alpha = 0.5, h = 0.05$					$\alpha = 1.75, h = 0.1$		
t	S1	S2	S3	S4	S5	S1	S2	S3
0.8	-4.17e-4	-1.57e-4	-2.33e-4	4.20e-4	7.09e-3	-2.19e-3	-1.52e-3	2.74e-2
1.6	-1.66e-4	-6.55e-5	-9.25e-5	2.16e-4	2.99e-3	1.39e-3	3.32e-3	-4.60e-2
2.4	-9.49e-5	-3.82e-5	-5.31e-5	1.37e-4	1.75e-3	3.88e-3	6.62e-3	-9.89e-2
3.2	-6.34e-5	-2.58e-5	-3.56e-5	9.67e-5	1.19e-3	3.78e-3	6.41e-3	-1.04e-1
4.0	-4.62e-5	-1.89e-5	-2.60e-5	7.31e-5	8.72e-4	4.72e-4	7.68e-4	-2.21e-2
4.8	-3.57e-5	-1.47e-5	-2.01e-5	5.77e-5	6.76e-4	-1.20e-3	-3.34e-3	5.02e-2
5.6	-2.85e-5	-1.17e-5	-1.60e-5	4.72e-5	5.45e-4	-6.88e-4	-2.07e-3	3.63e-2
6.4	-2.37e-5	-9.91e-6	-1.35e-5	3.93e-5	4.51e-4	-5.93e-5	5.74e-4	-5.52e-3

Table 8. Comparison of maximum errors in $y(t)$ for example 3 for $\alpha = 1.5$

h	S1	S2	S3	S4	S5
0.2	-5.5e-3	-4.6e-3	1.8e-1	-5.2e-3	-1.0e+0
0.1	-5.8e-3	-5.8e-3	9.1e-2	-5.9e-3	-5.2e+1
0.05	-6.1e-3	-6.1e-3	4.3e-2	-6.1e-3	-3.4e+2
0.025	-6.2e-3	-6.2e-3	1.9e-2	-6.2e-3	-9.7e+1
0.0125	-6.2e-3	-6.2e-3	6.2e-3	-6.2e-3	-7.4e+3
0.00625	-6.2e-3	-6.2e-3	0.0e+0	-6.2e-3	-2.0e+4

Table 9. Comparison the CPU times in seconds for example 3

	$\alpha = 1.5$					$\alpha = 0.25$			
h	S1	S2	S3	S4	S5	S1	S2	S3	S5
0.2	0.59	0.45	0.75	24.95	0.13	0.66	0.48	1.02	0.16
0.1	1.67	1.25	1.77	25.23	0.25	1.86	1.34	2.00	0.25
0.05	5.14	3.92	4.88	23.39	0.53	5.48	3.97	5.47	0.63
0.025	17.72	13.52	14.77	23.50	1.05	18.16	13.70	15.53	1.13
0.0125	65.81	49.61	51.06	23.22	2.09	66.98	49.59	52.13	2.14
0.00625	252.64	189.53	186.20	23.77	4.22	253.09	189.61	186.30	4.25
0.003125	989.27	740.64	713.03	23.55	8.53	989.09	742.73	712.84	8.63

$$D_{\star}^{\alpha}y(t) = \frac{40320}{\Gamma(9-\alpha)}t^{8-\alpha} - 3\frac{\Gamma(5+\alpha/2)}{\Gamma(5-\alpha/2)}t^{4-\alpha/2} + \frac{9}{4}\Gamma(\alpha+1) \\ + \left(\frac{3}{2}t^{\alpha/2} - t^4\right)^3 - [y(t)]^{3/2} \quad (28)$$

$$y(0) = 0, \quad \dot{y}(0) = 0. \quad (29)$$

The closed form solution for this example is given as

$$y(t) = t^8 3t^{4+\alpha/2} + \frac{9}{4}t^{\alpha}. \quad (30)$$

Figure 4 compares the results for $y(t)$ for various schemes for $\alpha = 0.75$ and $h = 0.025$ (left) and for $\alpha = 1.5$ and $h = 0.00625$ (right). Table 10 presents errors for various schemes for $(\alpha = 0.75, h = 0.00625)$ and $(\alpha = 1.5, h = 0.025)$. The maximum errors in $y(t)$ for different schemes for $\alpha = 0.5$ and different h are given in Table 11. Table 12 compares the CPU times for $\alpha = 0.5$ and 1.25 and different h .

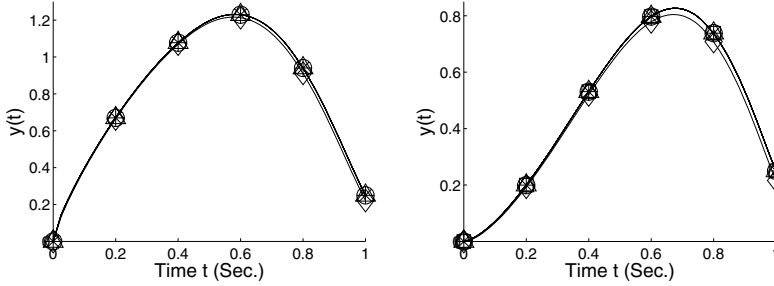


Fig. 4. Comparison of $y(t)$ obtained using different schemes for example 4. (Left: $\alpha = 0.75, h = 0.025$; Right: $\alpha = 1.5, h = 0.00625$.)

Table 10. Comparison of errors in $y(t)$ at different times for example 4

	$\alpha = 0.75, h = 0.00625$					$\alpha = 1.5, h = 0.025$				
t	S1	S2	S3	S5		S1	S2	S3	S4	S5
0	0	0	0	0		0	0	0	0	0
0.1	-5.60e-7	-3.81e-9	-1.08e-9	-1.99e-3		-5.06e-6	2.72e-7	-9.78e-8	-4.26e-6	-1.94e-2
0.2	-2.72e-6	-6.52e-9	-1.02e-9	-1.51e-3		-3.32e-5	4.65e-7	-1.78e-7	-5.79e-6	-2.81e-2
0.3	-6.56e-6	-8.43e-9	-0.97e-9	-1.48e-3		-9.83e-5	5.84e-7	-2.14e-7	-6.42e-6	-3.60e-2
0.4	-1.17e-5	-8.83e-9	-0.46e-9	-1.88e-3		-2.07e-4	5.84e-7	-1.26e-7	-5.91e-6	-4.54e-2
0.5	-1.74e-5	-6.41e-9	0.22e-9	-2.73e-3		-3.51e-4	3.97e-7	7.45e-8	-4.00e-6	-5.81e-2
0.6	-2.22e-5	0.67e-9	2.14e-9	-4.02e-3		-5.08e-4	-6.17e-8	5.30e-7	-7.84e-7	-7.49e-2
0.7	-2.35e-5	1.48e-8	5.09e-9	-5.61e-3		-6.25e-4	-8.90e-7	1.43e-6	3.68e-6	-9.57e-2
0.8	-1.75e-5	3.88e-8	7.44e-9	-7.18e-3		-6.18e-4	-2.20e-6	2.55e-6	8.91e-6	-1.18e-1
0.9	2.89e-6	7.55e-8	1.45e-8	-8.08e-3		-3.54e-4	-4.13e-6	4.30e-6	1.34e-5	-1.35e-1
1.0	4.97e-5	1.28e-7	2.45e-8	-6.98e-3		3.57e-4	-6.87e-6	7.06e-6	1.44e-5	-1.35e-1

5.5 Example 5

As the fifth example we consider the following linear FDE

Table 11. Comparison of maximum errors in $y(t)$ for example 4 for $\alpha = 0.5$

h	S1	S2	S3	S4	S5
0.2	2.30e-2	-5.30e-2	1.65e-2	-5.00e-5	-1.54e-1
0.1	8.39e-3	3.57e-3	9.22e-4	-9.95e-5	-7.84e-2
0.05	2.48e-3	3.94e-4	1.01e-4	-7.96e-5	-4.65e-2
0.025	6.79e-4	3.95e-5	7.30e-6	-1.67e-4	-3.56e-2
0.0125	1.79e-4	3.76e-6	3.95e-7	-1.61e-4	-2.65e-2
0.00625	4.65e-5	3.50e-7	2.53e-8	-2.32e-4	-1.93e-2
0.003125	1.19e-5	3.19e-8	-1.20e-8	-2.06e-4	-1.39e-2

Table 12. Comparison the CPU times in seconds for example 4

	$\alpha = 0.5$						$\alpha = 1.25$				
h	S1	S2	S3	S4	S5		S1	S2	S3	S5	
0.2	0.16	0.19	0.14	4.27	0.03		0.14	0.16	0.20	0.03	
0.1	0.31	0.23	0.02	4.64	0.09		0.31	0.23	0.06	0.06	
0.05	0.75	0.55	0.08	4.56	0.16		0.70	0.53	0.08	0.16	
0.025	1.91	1.41	0.14	4.59	0.31		1.77	1.38	0.16	0.31	
0.0125	5.42	4.08	0.41	4.56	0.61		5.33	4.14	0.56	0.63	
0.00625	17.72	13.63	1.28	4.59	1.25		17.63	13.52	1.50	1.23	
0.003125	62.47	47.22	4.38	4.53	2.47		62.63	48.83	5.28	2.48	

$$D_0^{0.2}X(t) = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & -1 & 0 & 0 & -\xi & 0 & 0 \end{pmatrix} X(t) + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} u(t) \quad (31)$$

where

$$X(t) = [x(t) \ D_0^{0.2}x(t) \ D_0^{0.4}x(t) \ D_0^{0.6}x(t) \ D_0^{0.8}x(t) \ D_0^1x(t) \ D_0^{1.2}x(t)]^T \quad (32)$$

and the homogeneous ICs

$$X(0) = [0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0]^T. \quad (33)$$

The output variable considered is given as

$$y(t) = [K \ 0 \ 0 \ K\xi \ 0 \ 0 \ 1 + K]X(t) \quad (34)$$

The closed form solution for this example is given as [28],

$$Y_{Anal}(t) = \frac{Kt^{0.2}}{\Gamma(1.2)} - \frac{5}{3}exp(tk_1)cos(tk_2 + \frac{\pi}{3n}) + \frac{5}{9}exp(tk_1)\sqrt{3}sin(tk_2 + \frac{\pi}{3n}) +$$

$$\frac{\sin(n\pi)}{\pi} \int_0^\infty \frac{x^\alpha e^{-tx} (2x^n a - 1)}{1 - 2x^n a + x^{2n} (4a^2 - 1) - 2x^{3n} a + x^{4n}} dx \quad (35)$$

where $n = 0.6$, $a = \cos(n\pi)$, $k_1 = \cos(\pi/(3n))$, $k_2 = \sin(\pi/(3n))$, and $\alpha = 0.2$. Figure 5 compares the results for $y(t)$ for $h = 0.025$. Table 13 presents errors for various schemes for $h = 0.0125$. The maximum errors in $y(t)$ for different schemes and different h are given in Table 14. This Table was generated by comparing the results at the intervals of 0.1 s only. Table 15 compares the CPU times for different schemes and different h .

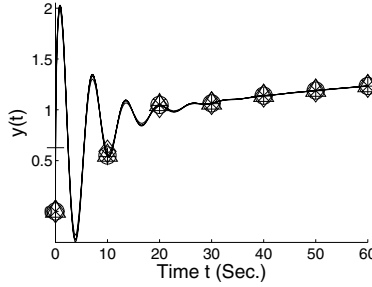


Fig. 5. Comparison of $y(t)$ for $h = 0.025$ for example 5.

Table 13. Comparison of errors in $y(t)$ at different times for example 5

t	S1	S2	S3	S5
10	1.09e-2	1.09e-2	1.09e-2	3.12e-2
20	5.52e-3	5.48e-3	5.44e-3	-1.30e-3
30	3.52e-3	3.54e-3	3.85e-3	4.87e-3
40	2.61e-3	2.60e-3	2.65e-3	2.40e-3
50	2.03e-3	2.04e-3	6.74e-4	2.01e-3
60	1.66e-3	1.66e-3	4.74e-3	1.65e-3

The numerical results presented above suggest the following: (1). The numerical errors for schemes S1, S2, and S3 decrease as the step size is reduced. This suggests that these schemes are numerical. In scheme S4, to meet the error tolerance requirement the Simulinks default integrator automatically adjusts the step size. Although, it is possible to force the Simulink to select a specific integrator and take a fix step size, it is generally not recommended. This is largely because, in most cases, it does not seem to improve the results, and often its efficiency deteriorates. Scheme S5 seems to work well in some cases. However, it is observed that S5 does not work well for $\alpha > 1$ and nonzero initial conditions. For example, Figure 2 shows that results obtained using S5

Table 14. Comparison of maximum errors in $y(t)$ for example 5 for different h

h	S1	S2	S3	S5
0.2	1.24e-1	7.40e-2	7.39e-2	3.28e-1
0.1	-9.59e-2	-8.48e-2	-8.04e-2	1.92e-1
0.05	-7.34e-2	-6.32e-2	-6.35e-2	1.13e-1
0.025	-6.26e-2	-5.87e-2	-5.95e-2	-8.09e-2
0.0125	-5.83e-2	-5.67e-2	-5.70e-2	-6.80e-2
0.00625	-5.65e-2	-5.58e-2	-5.59e-2	-6.15e-2

Table 15. Comparison the CPU times in seconds for example 5

h	S1	S2	S3	S5
0.2	29.97	19.58	22.63	0.28
0.1	99.73	73.97	76.17	0.42
0.05	384.89	287.53	287.23	1.31
0.025	1524.59	1140.45	1112.77	3.56
0.0125	6128.41	4555.50	4422.42	12.61
0.00625	24742.19	18340.91	17652.16	46.34

have significant error. For example 3, results obtained using S5 for $\alpha > 1$ were divergent. For this reason, these results are not included in Figure 3, but they are included in Table 8. (2). For the given step size, results from S5 almost always have large errors. The errors in the results obtained using the other 4 schemes are very close to each other. However, for a given step size, scheme S2 seems to give, for most part, more accurate results. The maximum error table suggests that for larger step sizes, S4 may give more accurate results. However, as the step size is reduced, the other schemes give better results. This is because in S4 scheme, the Simulinks integrator adaptively changes the step size. 3. In most case, scheme S5 seems to be the fastest. However, as reported above, it gives large errors and in some cases it fails to give accurate results. As a result, this scheme could be used if the results for the given class of problems have been verified and a very high accuracy is not desired. The CPU time for scheme S4 seems to be stable. This is because the integrator automatically adjusts the step size. The CPU times for the other three schemes seems to be comparable. For larger step sizes, they are faster than S4, for smaller step sizes the opposite is true. For larger step sizes, for most part S2 takes less CPU time than S1 and S3 do.

6 Conclusions

Five numerical schemes to solve linear and nonlinear FDEs subjected to homogeneous and nonhomogeneous ICs are discussed. These schemes were used to

solve five different problems, two linear 1-D, two nonlinear 1-D, and one linear multidimensional. Performance studies included stability, accuracy, and computational speed. Results suggest that the choice of an algorithm will depend on the problem considered and the performance criteria selected.

References

1. Mainardi F (1997) Fractional calculus: some basic problem in continuum and statistical mechanics, in: *Fractals and Fractional Calculus in Continuum Mechanics*. Carpinteri, A, Mainardi, F (eds.), Springer, Wein, New York, pp. 291–348.
2. Rossikhin YA, Shitikova MV (1997) Applications of fractional calculus to dynamic problems of linear and nonlinear hereditary mechanics of solids. *Appl. Mech. Rev.* 50:15–67.
3. Podlubny I (1999) *Fractional Differential Equations*. Academic Press, New York.
4. Hilfer R (2000) *Applications of Fractional Calculus in Physics*. World Scientific, New Jersey.
5. West BJ, Bologna M, Grigolini P (2003) *Physics of Fractal Operators*. Springer, New York.
6. Magin RL (2004) Fractional calculus in bioengineering. *Crit. Rev. Biomed. Eng.* 32(1):1–104.
7. Magin RL (2004) Fractional calculus in bioengineering – Part 2. *Crit. Rev. Biomed. Eng.* 32(2):105–193.
8. Magin RL (2004) Fractional calculus in bioengineering – Part 3. *Crit. Rev. Biomed. Eng.* 32(3/4):194–377.
9. Miller KS, Ross B (1993) *An Introduction to the Fractional Calculus and Fractional Differential Equations*. Wiley, New York.
10. Gaul L, Klein P, Kempfle S (1989) Impulse response function of an oscillator with fractional derivative in damping description. *Mech. Res. Commun.* 16(5):4447–4472.
11. Suarez LE, Shokooh A (1997) An eigenvector expansion method for the solution of motion containing fractional derivatives. *ASME. J. Appl. Mech.* 64:629–635.
12. Yuan L, Agrawal OP (2002) A numerical scheme for dynamic systems containing fractional derivatives. *Transactions of the ASME, J. Vib. Acoust.* 124:321–324.
13. Machado JAT (2001) Discrete-time fractional-order controllers. *FCAA J.* 4:47–66.
14. Heleschewitz D, Matignon D (1998) Diffusive Realizations of Fractional integrodifferential Operators: Structural Analysis Under Approximation, in: *Proceedings IFAC Conference System, Structure and Control*, Nantes, France, 2:243–248.
15. Aoun M, Malti R, Levron F, Oustaloup A (2003) Numerical simulation of fractional systems, in: *Proceedings of DETC2003, 2003 ASME Design Engineering Technical Conferences*, September 2–6, Chicago, Illinois.
16. Pointot T, Trigeassou J (2003) Modeling and simulation of fractional systems using a non integer integrator, in: *Proceedings of DETC2003, 2003 ASME Design Engineering Technical Conferences*, September 2–6, Chicago, Illinois.
17. Padovan J (1987) Computational algorithms and finite element formulation involving fractional operators. *Comput. Mech.* 2:271–287.
18. Gorenflo R (1997) Fractional calculus: some numerical methods in: Carpinteri A, Mainardi, F (eds.), *Fractals and Fractional Calculus in Continuum Mechanics*. Springer, Wein, New York, pp. 277–290.

19. Ruge P, Wagner N (1999) Time-domain solutions for vibration systems with feeding memory. European Conference of Computational Mechanics, Munchen, Germany, August 31 September 3.
20. Diethelm K, Ford NJ (2002) Analysis of fractional differential equations. *J. Math. Anal. Appl.* 265:229–248.
21. Diethelm K, Ford NJ, Freed AD (2002) A predictor-corrector approach for the numerical solution of fractional differential equations. *Nonlinear Dynamics* 29(1-4): 3–22.
22. Diethelm K (2003) Efficient solution of multi-term fractional differential equation using P(EC)^mE methods. *Computing* 71:305–319.
23. Diethelm K, Ford NJ (2004) Multi-order fractional differential equations and their numerical solution. *Appl. Math. Comput.* 154(3):621–640.
24. Diethelm K, Ford NJ, Freed AD, Luchko Y (2005) Algorithms for the fractional calculus: a selection of numerical methods. *Comput. Methods Appl. Mechan. Eng.* 194:743–773.
25. Agrawal OP (2004) Block-by-Block Method for Numerical Solution of Fractional Differential Equations, in: *Proceedings of IFAC2004, First IFAC Workshop on Fractional Differentiation and Its Applications*. Bordeaux, France, July 19–21.
26. Kumar P, Agrawal OP (2005) A Cubic Scheme for Numerical Solution of Fractional Differential Equations, in: *Proceedings of the Fifth EUROMECH Nonlinear Dynamics Conference*, Eindhoven University of Technology, Eindhoven, The Netherland, August 7–12.
27. Kumar P, Agrawal OP (2005) Numerical Scheme for the Solution of Fractional Differential Equations, in: *Proceedings of the 2005 ASME Design Engineering Technical Conferences and Computer and Information Engineering Conference*, Long Beach, California, September 24–28.
28. Sabatier J, Malti R (2004) Simulation of Fractional Systems: A Benchmark, in: *Proceedings of IFAC2004, First IFAC Workshop on Fractional Differentiation and Its Applications*. Bordeaux, France, July 19–21.
29. Lorenzo CF, Hartley TT (2000) Initialized fractional calculus. *Int. J. Appl. Math.* 3:249–265.
30. Achar BN, Lorenzo CF, Hartley TT (2005) Initialization issue of the Caputo fractional derivative, in: *Proceedings of the 2005 ASME Design Engineering Technical Conferences*, Long Beach, California, September 24–28.

SUBOPTIMUM H_2 PSEUDO-RATIONAL APPROXIMATIONS TO FRACTIONAL-ORDER LINEAR TIME INVARIANT SYSTEMS

Dingyü Xue¹ and YangQuan Chen²

¹ Institute of Artificial Intelligence and Robotics, Faculty of Information Science and Engineering, Northeastern University, Shenyang 110004, PR China,
E-mail: xuedingyu@ise.neu.edu.cn

² Center for Self-Organizing and Intelligent Systems (CSOIS), Department of Electrical and Computer Engineering, Utah State University, 4120 Old Main Hill, Logan, UT 84322-4120;
E-mail: yqchen@ece.usu.edu

Abstract

In this paper, we propose a procedure to achieve pseudo-rational approximation to arbitrary fractional-order linear time invariant (FO-LTI) systems with suboptimum H_2 -norm. The proposed pseudo-rational approximation is actually a rational model with a time delay. Through illustrations, we show that the pseudo-rational approximation is simple and effective. It is also demonstrated that this suboptimum approximation method is effective in designing integer-order controllers for FO-LTI systems in general non-commensurate form. Useful MATLAB codes are also included in the appendix.

Keywords

Fractional-order systems, model reduction, optimal model reduction, time delay systems, H_2 -norm approximation.

1 Introduction

Fractional order calculus, a 300-years-old topic [1, 2, 3, 4], has been gaining increasing attention in research communities. Applying fractional-order calculus to dynamic systems control, however, is just a recent focus of interest [5, 6, 7, 8, 9]. We should point out references [10, 11, 12, 13] for pioneering works and [14, 15, 16] for more recent developments. In most cases, our objective is to apply fractional-order control to enhance the system control performance. For example, as in the CRONE, where CRONE is a French abbreviation for “*Commande robuste d’ordre non-entier*” (which means non-integer order robust control), [17, 7, 8], *fractal robustness* is pursued. The

desired frequency template leads to fractional transmittance [18, 19] on which the CRONE controller synthesis is based. In CRONE controllers, the major ingredient is the fractional-order derivative s^r , where r is a real number and s is the Laplace transform symbol of differentiation. Another example is the $PI^\lambda D^\mu$ controller [6, 20], an extension of PID controller. In general form, the transfer function of $PI^\lambda D^\mu$ is given by $K_p + T_i s^{-\lambda} + T_d s^\mu$, where λ and μ are positive real numbers; K_p is the proportional gain, T_i the integration constant and T_d the differentiation constant. Clearly, taking $\lambda = 1$ and $\mu = 1$, we obtain a classical PID controller. If $T_i = 0$ we obtain a PD^μ controller, etc. All these types of controllers are particular cases of the $PI^\lambda D^\mu$ controller. It can be expected that the $PI^\lambda D^\mu$ controller may enhance the systems control performance due to more tuning knobs introduced.

Actually, in theory, $PI^\lambda D^\mu$ itself is an infinite dimensional linear filter due to the fractional order in the differentiator or integrator. It should be pointed out that a band-limit implementation of FOC is important in practice, i.e., the finite dimensional approximation of the FOC should be done in a proper range of frequencies of practical interest [21, 19]. Moreover, the fractional order can be a complex number as discussed in [21]. In this paper, we focus on the case where the fractional order is a real number.

For a single term s^r with r a real number, there are many approximation schemes proposed. In general, we have analog realizations [22, 23, 24, 25] and digital realizations. The key step in digital implementation of an FOC is the numerical evaluation or discretization of the fractional-order differentiator s^r . In general, there are two discretization methods: *direct discretization* and *indirect discretization*. In *indirect discretization* methods [21], two steps are required, i.e., frequency-domain fitting in continuous time domain first and then discretizing the fit s -transfer function. Other frequency-domain fitting methods can also be used but without guaranteeing the stable minimum-phase discretization. Existing *direct discretization* methods include the application of the direct power series expansion (PSE) of the Euler operator [26, 27, 28, 29], continuous fractional expansion (CFE) of the Tustin operator [27, 28, 29, 30, 31], and numerical integration-based method [26, 30, 32]. However, as pointed out in [33, 34, 35], the Tustin operator-based discretization scheme exhibits large errors in high-frequency range. A new mixed scheme of Euler and Tustin operators is proposed in [30] which yields the so-called Al-Alaoui operator [33]. These discretization methods for s^r are in IIR form. Recently, there are some reported methods to directly obtain the digital fractional-order differentiators in FIR (finite impulse response) form [36, 37]. However, using an FIR filter to approximate s^r may be less efficient due to very high order of the FIR filter. So, discretizing fractional differentiators in IIR forms is preferred [38, 30, 32, 31].

In this paper, we consider the general fractional-order LTI systems (FO-LTI) with noncommensurate fractional orders as follows:

$$G(s) = \frac{b_m s^{\gamma_m} + b_{m-1} s^{\gamma_{m-1}} + \cdots + b_1 s^{\gamma_1} + b_0}{a_n s^{\eta_n} + a_{n-1} s^{\eta_{n-1}} + \cdots + a_1 s^{\eta_1} + a_0}. \quad (1)$$

Using the aforementioned approximation schemes for a single s^γ and then for the general FO-LTI system (1) could be very tedious, leading to a very high-order model. In this paper, we propose to use a numerical algorithm to achieve a good approximation of the overall transfer function (1) using finite integer-order rational transfer function with a possible time delay term and illustrate how to use the approximated integer-order model for integer-order controller design. In Examples 1 and 2, approximation to a fractional-order transfer function is given and the fitting results are illustrated. In example 3, a fractional-order plant is approximated using the algorithm proposed in the paper, by a FOPD (first-order plus delay) model, and using an existing PID tuning formula, an integer order PID can be designed with a very good performance.

2 True Rational Approximations to Fractional Integrators and Differentiators: Oustaloup's Method

For comparison purpose, here we present Oustaloup's algorithm [18, 19, 39]. Assuming that the frequency range to fit is selected as (ω_b, ω_h) , the transfer function of a continuous filter can be constructed to approximate the pure fractional derivative term s^γ such that

$$G_{f,\gamma}(s) = K \prod_{k=-N}^N \frac{s + \omega'_k}{s + \omega_k} \quad (2)$$

where the zeros, poles, and the gain can be evaluated from

$$\omega'_k = \omega_b \left(\frac{\omega_h}{\omega_b} \right)^{\frac{k+N+\frac{1}{2}(1-\gamma)}{2N+1}}, \quad \omega_k = \omega_b \left(\frac{\omega_h}{\omega_b} \right)^{\frac{k+N+\frac{1}{2}(1+\gamma)}{2N+1}}, \quad K = \omega_h^\gamma. \quad (3)$$

where $k = -N, \dots, N$.

An implementation in MATLAB is given in Appendix 1. Substituting γ_i and η_i in (1) with $G_{f,\gamma_i}(s)$ and $G_{f,\eta_i}(s)$ respectively, the original fractional-order model $G(s)$ can be approximated by a rational function $\hat{G}(s)$. It should be noted that the order of the resulted $\hat{G}(s)$ is usually very high. Thus, there is a need to approximate the original model by reduced order ones using the optimal-reduction techniques.

3 A Numerical Algorithm for Suboptimal Pseudo-Rational Approximations

In this section, we are interested in finding an approximate integer-order model with a low order, possibly with a time delay in the following form:

$$G_{r/m,\tau}(s) = \frac{\beta_1 s^r + \dots + \beta_r s + \beta_{r+1}}{s^m + \alpha_1 s^{m-1} + \dots + \alpha_{m-1} s + \alpha_m} e^{-\tau s}. \quad (4)$$

An objective function for minimizing the H_2 -norm of the reduction error signal $e(t)$ can be defined as

$$\mathbf{J} = \min_{\theta} \left\| \widehat{G}(s) - G_{r/m,\tau}(s) \right\|_2 \quad (5)$$

where θ is the set of parameters to be optimized such that

$$\theta = [\beta_1, \dots, \beta_r, \alpha_1, \dots, \alpha_m, \tau]. \quad (6)$$

For an easy evaluation of the criterion $\mathbf{J}$, the delayed term in the reduced order model $G_{r/m,\tau}(s)$ can be further approximated by a rational function $\widehat{G}_{r/m}(s)$ using the Padé approximation technique [40]. Thus, the revised criterion can then be defined by

$$\mathbf{J} = \min_{\theta} \left\| \widehat{G}(s) - \widehat{G}_{r/m}(s) \right\|_2. \quad (7)$$

and the H_2 norm computation can be evaluated recursively using the algorithm in [41].

Suppose that for a stable transfer function type $E(s) = \widehat{G}(s) - \widehat{G}_{r/m}(s) = B(s)/A(s)$, the polynomials $A_k(s)$ and $B_k(s)$ can be defined such that,

$$A_k(s) = a_0^k + a_1^k s + \dots + a_k^k s^k, \quad B_k(s) = b_0^k + b_1^k s + \dots + b_{k-1}^k s^{k-1} \quad (8)$$

The values of a_i^{k-1} and b_i^{k-1} can be evaluated from

$$a_i^{k-1} = \begin{cases} a_{i+1}^k, & \text{i even} \\ a_{i+1}^k - \alpha_k a_{i+2}^k, & \text{i odd} \end{cases} \quad i = 0, \dots, k-1 \quad (9)$$

and

$$b_i^{k-1} = \begin{cases} b_{i+1}^k, & \text{i even} \\ b_{i+1}^k - \beta_k a_{i+2}^k, & \text{i odd} \end{cases} \quad i = 1, \dots, k-1 \quad (10)$$

where, $\alpha_k = a_0^k/a_1^k$, and $\beta_k = b_1^k/a_1^k$.

The H_2 -norm of the approximate reduction error signal $\hat{e}(t)$ can be evaluated from

$$J = \sum_{k=1}^n \frac{\beta_k^2}{2\alpha_k} = \sum_{k=1}^n \frac{(b_1^k)^2}{2a_0^k a_1^k} \quad (11)$$

The sub-optimal H_2 -norm reduced order model for the original high order fractional order model can be obtained using the following procedure [40]:

1. Select an initial reduced model $\widehat{G}_{r/m}^0(s)$.
2. Evaluate an error $\left\| \widehat{G}(s) - \widehat{G}_{r/m}^0(s) \right\|_2$ from (11).

3. Use an optimization algorithm (for instance, Powell's algorithm [42]) to iterate one step for a better estimated model $\widehat{G}_{r/m}^1(s)$.
4. Set $\widehat{G}_{r/m}^0(s) \leftarrow \widehat{G}_{r/m}^1(s)$, go to step 2 until an optimal reduced model $\widehat{G}_{r/m}^*(s)$ is obtained.
5. Extract the delay from $\widehat{G}_{r/m}^*(s)$, if any.

We call the above procedure suboptimal since the Oustaloup's method is used for each single term s^γ in (1), and also, Padé approximation is used for pure delay terms.

4 Illustrative Examples

Examples are given in the section to demonstrate the optimal-model reduction procedures with full MATLAB implementations. Also the integer-order PID controller design procedure is explored for fractional-order plants, based on the model reduction algorithm in the paper.

Example 1: Non-commensurate FO-LTI system

Consider the non-commensurate FO-LTI system

$$G(s) = \frac{5}{s^{2.3} + 1.3s^{0.9} + 1.25}.$$

Using the following MATLAB scripts,

```
w1=1e-3; w2=1e3; N=2;
g1=ousta_fod(0.3,N,w1,w2); g2=ousta_fod(0.9,N,w1,w2);
s=tf('s'); G=5/(s^2*g1+1.3*g2+1.25);
```

with the Oustaloup's filter, the high-order approximation to the original fractional-order model can be approximated by

$$G(s) = \frac{5s^{10} + 6677s^9 + 2.191 \times 10^6 s^8 + 1.505 \times 10^8 s^7 + 2.936 \times 10^9 s^6 + 1.257 \times 10^{10} s^5 + 1.541 \times 10^{10} s^4 + 4.144 \times 10^9 s^3 + 3.168 \times 10^8 s^2 + 5.065 \times 10^6 s + 1.991 \times 10^4}{7.943s^{12} + 8791s^{11} + 1.731 \times 10^6 s^{10} + 8.766 \times 10^7 s^9 + 1.046 \times 10^9 s^8 + 3.82 \times 10^9 s^7 + 6.099 \times 10^9 s^6 + 7.743 \times 10^9 s^5 + 5.197 \times 10^9 s^4 + 1.15 \times 10^9 s^3 + 8.144 \times 10^7 s^2 + 1.278 \times 10^6 s + 4987}.$$

The following statements can then be used to find the optimum reduced order approximations to the original fractional order model.

```
G1=opt_app(G,1,2,0); G2=opt_app(G,2,3,0);
G3=opt_app(G,3,4,0); G4=opt_app(G,4,5,0);
step(G,G1,G2,G3,G4)
```

where the four reduced order models can be obtained

$$\begin{aligned}
 G_1(s) &= \frac{-2.045s + 7.654}{s^2 + 1.159s + 1.917} \\
 G_2(s) &= \frac{-0.5414s^2 + 4.061s + 2.945}{s^3 + 0.9677s^2 + 1.989s + 0.7378} \\
 G_3(s) &= \frac{-0.2592s^3 + 3.365s^2 + 4.9s + 0.3911}{s^4 + 1.264s^3 + 2.25s^2 + 1.379s + 0.09797} \\
 G_4(s) &= \frac{1.303s^4 + 1.902s^3 + 11.15s^2 + 4.71s + 0.1898}{s^5 + 2.496s^4 + 3.485s^3 + 4.192s^2 + 1.255s + 0.04755}
 \end{aligned}$$

The step responses for the above four reduced-order models can be obtained as compared in Fig. 1. It can be seen that the 1/2th order model gives a poor approximation to the original system, while the other low-order approximations using the method and codes of this paper are effective.

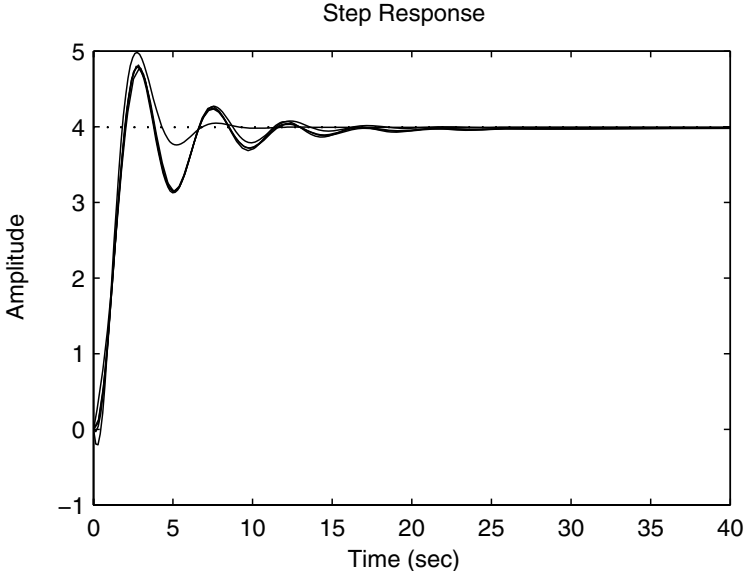


Fig. 1. Step responses comparisons of rational approximations.

Example 2: Non-commensurate FO-LTI system

Consider the following non-commensurate FO-LTI system:

$$G(s) = \frac{5s^{0.6} + 2}{s^{3.3} + 3.1s^{2.6} + 2.89s^{1.9} + 2.5s^{1.4} + 1.2}.$$

Using the following MATLAB scripts,

```

N=2; w1=1e-3; w2=1e3;
g1=ousta_fod(0.3,N,w1,w2); g2=ousta_fod(0.6,N,w1,w2);

```

```

g3=ousta_fod(0.9,N,w1,w2); g4=ousta_fod(0.4,N,w1,w2);
s=tf('s');
G=(5*g2+2)/(s^3*g1+3.1*s^2*g2+2.89*s*g3+2.5*s*g4+1.2);

```

an extremely high-order model can be obtained with the Oustaloup's filter, such that

$$\begin{aligned}
 G(s) = & \frac{317.5s^{25} + 8.05 \times 10^5 s^{24} + 7.916 \times 10^8 s^{23} + 3.867 \times 10^{11} s^{22} \\
 & + 1.001 \times 10^{14} s^{21} + 1.385 \times 10^{16} s^{20} + 1.061 \times 10^{18} s^{19} + 4.664 \times 10^{19} s^{18} \\
 & + 1.197 \times 10^{21} s^{17} + 1.778 \times 10^{22} s^{16} + 1.5 \times 10^{23} s^{15} + 7.242 \times 10^{23} s^{14} \\
 & + 2.052 \times 10^{24} s^{13} + 3.462 \times 10^{24} s^{12} + 3.459 \times 10^{24} s^{11} + 2.009 \times 10^{24} s^{10} \\
 & + 6.724 \times 10^{23} s^9 + 1.329 \times 10^{23} s^8 + 1.579 \times 10^{22} s^7 + 1.12 \times 10^{21} s^6 \\
 & + 4.592 \times 10^{19} s^5 + 1.037 \times 10^{18} s^4 + 1.314 \times 10^{16} s^3 + 9.315 \times 10^{13} s^2 \\
 & + 3.456 \times 10^{11} s + 5.223 \times 10^8}{7.943s^{28} + 2.245 \times 10^4 s^{27} + 2.512 \times 10^7 s^{26} + 1.427 \times 10^{10} s^{25} \\
 & + 4.392 \times 10^{12} s^{24} + 7.384 \times 10^{14} s^{23} + 6.896 \times 10^{16} s^{22} + 3.736 \times 10^{18} s^{21} \\
 & + 1.208 \times 10^{20} s^{20} + 2.343 \times 10^{21} s^{19} + 2.716 \times 10^{22} s^{18} + 1.896 \times 10^{23} s^{17} \\
 & + 8.211 \times 10^{23} s^{16} + 2.268 \times 10^{24} s^{15} + 4.076 \times 10^{24} s^{14} + 4.834 \times 10^{24} s^{13} \\
 & + 3.845 \times 10^{24} s^{12} + 2.134 \times 10^{24} s^{11} + 8.772 \times 10^{23} s^{10} + 2.574 \times 10^{23} s^9 \\
 & + 5.057 \times 10^{22} s^8 + 6.342 \times 10^{21} s^7 + 4.868 \times 10^{20} s^6 + 2.16 \times 10^{19} s^5 \\
 & + 5.176 \times 10^{17} s^4 + 6.863 \times 10^{15} s^3 + 5.055 \times 10^{13} s^2 \\
 & + 1.938 \times 10^{11} s + 3.014 \times 10^8}
 \end{aligned}$$

and the order of rational approximation to the original order model is the 28th, for $N = 2$. For larger values of N , the order of rational approximation may be even much higher. For instance, the order of the approximation may reach the 38th and 48th respectively for the selections $N = 3$ and $N = 4$, with extremely large coefficients. Thus the model reduction algorithm should be used with the following MATLAB statements

```

G2=opt_app(G,2,3,0); G3=opt_app(G,3,4,0);
G4=opt_app(G,4,5,0); step(G,G2,G3,G4,60)

```

the step responses can be compared in Fig. 2 and it can be seen that the third-order approximation is satisfactory and the fourth-order fitting gives a better approximation. The obtained optimum approximated results are listed in the following:

$$\begin{aligned}
 G_2(s) &= \frac{0.41056s^2 + 0.75579s + 0.037971}{s^3 + 0.24604s^2 + 0.22176s + 0.021915} \\
 G_3(s) &= \frac{-4.4627s^3 + 5.6139s^2 + 4.3354s + 0.15330}{s^4 + 7.4462s^3 + 1.7171s^2 + 1.5083s + 0.088476} \\
 G_4(s) &= \frac{1.7768s^4 + 2.2291s^3 + 10.911s^2 + 1.2169s + 0.010249}{s^5 + 11.347s^4 + 4.8219s^3 + 2.8448s^2 + 0.59199s + 0.0059152}
 \end{aligned}$$

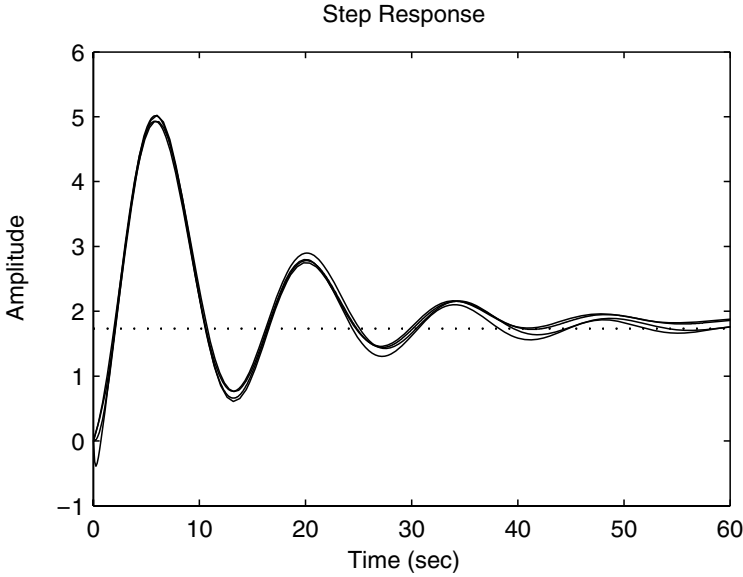


Fig. 2. Step responses comparisons.

Example 3: Sub-optimum pseudo-rational model reduction for integer-order PID controller design

Let us consider the following FO-LTI plant model:

$$G(s) = \frac{1}{s^{2.3} + 3.2s^{1.4} + 2.4s^{0.9} + 1}.$$

Let us first approximate it with Oustaloup's method and then fit it with a fixed model structure known as FOLPD (first-order lag plus deadtime) model,

where $G_r(s) = \frac{K}{Ts + 1} e^{-Ls}$. The following MATLAB scripts

```
N=2; w1=1e-3; w2=1e3;
g1=ousta_fod(0.3,N,w1,w2);
g2=ousta_fod(0.4,N,w1,w2);
g3=ousta_fod(0.9,N,w1,w2);
s=tf('s'); G=1/(s^2*g1+3.2*s*g2+2.4*g3+1);
G2=opt_app(G,0,1,1); step(G,G2)
```

can perform this task and the obtained optimal FOLPD model is given as follows:

$$G_r(s) = \frac{0.9951}{3.5014s + 1} e^{-1.634s}.$$

The comparison of the open-loop step response is shown in Fig. 3. It can be observed that the approximation is fairly effective.

Designing a suitable feedback controller for the original FO-LTI system G can be a formidable task. Now, let us consider designing an integer-order PID

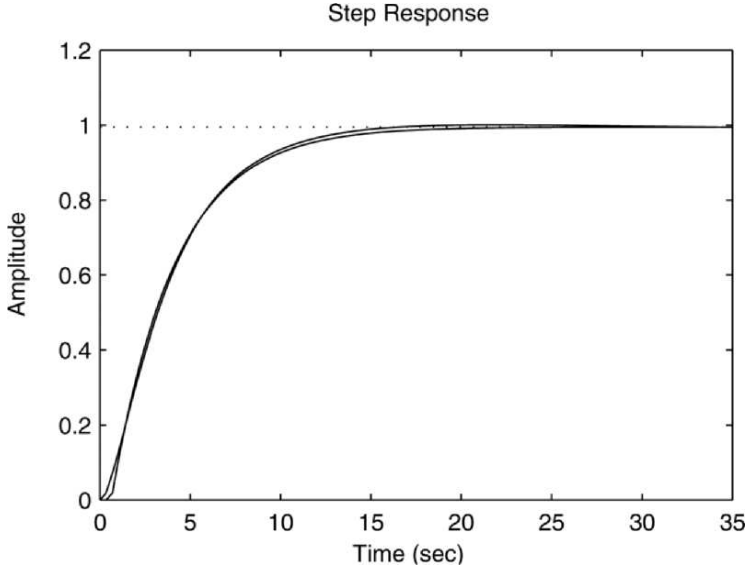


Fig. 3. Step response comparison of the optimum FOLPD and the original model.

controller for the optimally reduced model $G_r(s)$ and let us see if the designed controller still works for the original system.

The integer order PID controller to be designed is in the following form:

$$G_c(s) = K_p \left(1 + \frac{1}{T_i s} + \frac{T_d s}{T_d / N s + 1} \right). \quad (12)$$

The optimum ITAE criterion-based PID tuning formula [43] can be used

$$K_p = \frac{(0.7303 + 0.5307T/L)(T + 0.5L)}{K(T + L)}, \quad (13)$$

$$T_i = T + 0.5L, \quad T_d = \frac{0.5LT}{T + 0.5L}. \quad (14)$$

Based on this tuning algorithm, a PID controller can be designed for $G_r(s)$ as follows:

$$\begin{aligned} L &= 0.63; \quad T = 3.5014; \quad K = 0.9951; \quad N = 10; \quad T_i = T + 0.5L; \\ K_p &= (0.7303 + 0.5307T/L) * T_i / (K * (T + L)); \\ T_d &= (0.5 * L * T) / (T + 0.5L); \quad [K_p, T_i, T_d] \\ G_c &= K_p * (1 + 1/T_i/s + T_d*s/(T_d/N*s + 1)) \end{aligned}$$

The parameters of the PID controller are then $K_p = 3.4160$, $T_i = 3.8164$, $T_d = 0.2890$, and the PID controller can be written as

$$G_c(s) = \frac{1.086s^2 + 3.442s + 0.8951}{0.0289s^2 + s}$$

Finally, the step response of the original FO-LTI with the above-designed PID controller is shown in Fig. 4. A satisfactory performance can be clearly observed. Therefore, we believe, the method presented in this paper can be used for integer-order controller design for general FO-LTI systems.

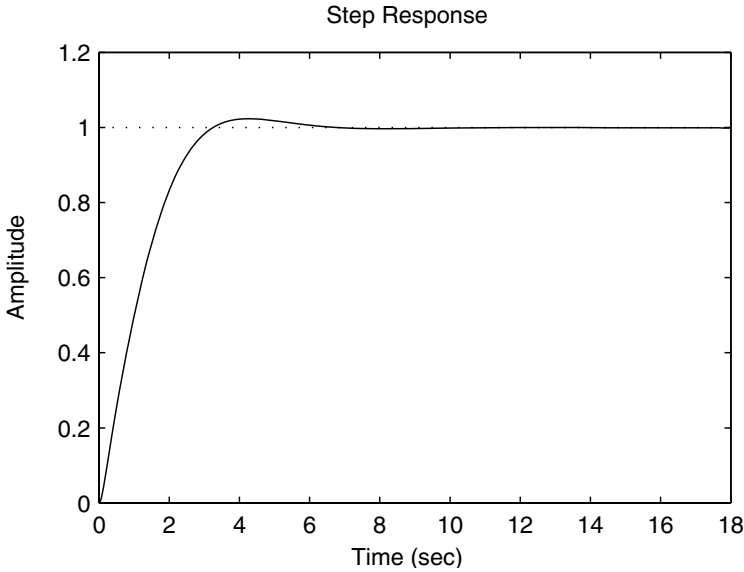


Fig. 4. Step response of fractional-order plant model under the PID controller.

5 Concluding Remarks

In this paper, we presented a procedure to achieve pseudo-rational approximation to arbitrary FO-LTI systems with suboptimum H_2 -norm. Relevant MATLAB codes useful for practical applications are also given in the appendix. Through illustrations, we show that the pseudo-rational approximation is simple and effective. It is also demonstrated that this suboptimum approximation method is effective in designing integer order controllers for FO-LTI systems in general form.

Finally, we would like to remark that the so-called pseudo-rational approximation is essentially by cascading irrational transfer function (a time delay) and a rational transfer function. Since a delay element is also infinite dimensional, it makes sense to approximate a general fractional-order LTI system involving time delay. Although it might not fully make physical sense, the pseudo-rational approximation proposed in this paper will find its practical applications in designing an integer-order controller for fractional-order systems, as illustrated in Example 3.

Acknowledgment

We acknowledge that this paper is a modified version of a paper published in the Proceedings of IDETC/CIE 2005 (Paper# DETC2005-84743). We would like to thank the ASME for granting us permission in written form to publish a modified version of IDETC/CIE 2005 (Paper# DETC2005-84743) as a chapter in the book entitled *Advances in Fractional Calculus: Theoretical Developments and Applications in Physics and in Engineering* edited by Professors Machado, Sabatier, and Agrawal (Springer).

Appendix 1 MATLAB functions for optimum fractional model reduction

- `ousta_fod.m` Outstaloup's rational approximation to fractional differentiator, with the syntax $G = \text{ousta\_fod}(r, N, \omega_L, \omega_H)$

```
function G=ousta_fod(r,N,w_L,w_H)
mu=w_H/w_L; k=-N:N; w_kp=(mu).^((k+N+0.5-0.5*r)/(2*N+1))*w_L;
w_k=(mu).^((k+N+0.5+0.5*r)/(2*N+1))*w_L;
K=(mu).^(-r/2)*prod(w_k./w_kp); G=tf(zpk(-w_kp',-w_k',K));
```

- `opt_app.m` Optimal model reduction function, and the pseudo-rational transfer function model G_r , i.e., the transfer function with a possible delay term, can be obtained. $G_r = \text{opt\_app}(G, r, d, \text{key}, G_0)$, where `key` indicates whether a time delay is required in the reduced order model. G_0 is the initial reduced order model, optional.

```
function G_r=opt_app(G,nn,nd,key,G0)
GS=tf(G); num=GS.num{1}; den=GS.den{1}; Td=totaldelay(GS);
GS.ioDelay=0; GS.InputDelay=0; GS.OutputDelay=0;
if nargin<5,
    n0=[1,1];
    for i=1:nd-2, n0=conv(n0,[1,1]); end
    G0=tf(n0,conv([1,1],n0));
end
beta=G0.num{1}(nd+1:nn:nd+1); alph=G0.den{1}; Tau=1.5*Td;
x=[beta(1:nn),alph(2:nd+1)]; if abs(Tau)<1e-5, Tau=0.5; end
if key==1, x=[x,Tau]; end
dc=dcgain(GS); y=opt_fun(x,GS,key,nn,nd,dc);
x=fminsearch('opt_fun',x,[],GS,key,nn,nd,dc);
alph=[1,x(nn+1:nn+nd)]; beta=x(1:nn+1); if key==0, Td=0; end
beta(nn+1)=alph(end)*dc;
if key==1, Tau=x(end)+Td; else, Tau=0; end
G_r=tf(beta,alph,'ioDelay',Tau);
```

- `opt_fun.m` internal function used by `opt_app`,

```
function y=opt_fun(x,G,key,nn,nd,dc)
ff0=1e10; alph=[1,x(nn+1:nn+nd)];
beta=x(1:nn+1); beta(end)=alph(end)*dc; g=tf(beta,alph);
if key==1,
    tau=x(end); if tau<=0, tau=eps; end
    [nP,dP]=pade(tau,3); gP=tf(nP,dP);
else, gP=1; end
G_e=G-g*gP;
G_e.num{1}=[0,G_e.num{1}(1:end-1)];
[y,ierr]=geth2(G_e);
if ierr==1, y=10*ff0; else, ff0=y; end
```

- `get2h.m` internal function to evaluate the H_2 norm of a rational transfer function model.

```
function [v,ierr]=geth2(G)
G=tf(G); num=G.num{1}; den=G.den{1}; ierr=0; n=length(den);
if abs(num(1))>eps
    disp('System not strictly proper'); ierr=1; return
else, a1=den; b1=num(2:end); end
for k=1:n-1
    if (a1(k+1)<=eps), ierr=1; v=0; return
    else,
        aa=a1(k)/a1(k+1); bb=b1(k)/a1(k+1); v=v+bb*bb/aa; k1=k+2;
        for i=k1:2:n-1
            a1(i)=a1(i)-aa*a1(i+1); b1(i)=b1(i)-bb*a1(i+1);
        end, end, end
v=sqrt(0.5*v);
```

References

1. Oldham KB, Spanier J (1974) *The Fractional Calculus*. Academic Press, New York.
2. Samko SG, Kilbas AA, Marichev OI (1987). *Fractional integrals and derivatives and some of their applications*. Nauka i tehnika, Minsk, Russia.
3. Miller KS, Ross B (1993) *An Introduction to the Fractional Calculus and Fractional Differential Equations*. Wiley, New York.

4. Podlubny I (1994) Fractional-order systems and fractional-order controllers. Tech. Rep. UEF-03-94, Slovak Academy of Sciences. Institute of Experimental Physics, Department of Control Engineering. Faculty of Mining, University of Technology. Kosice, Slovakia, November.
5. Lurie BJ (1994) 'Three-parameter tunable tilt-integral-derivative (TID) controller'. US Patent US5371670.
6. Podlubny I (1999) Fractional-order systems and $PI^\lambda D^\mu$ -controllers. IEEE Trans. Auto. Control, 44(1):208–214.
7. Oustaloup A, Mathieu B, Lanusse P (1995) The CRONE control of resonant plants: application to a flexible transmission. Eur. J. Contr. 1(2).
8. Oustaloup A, Moreau X, Nouillant M (1996) The CRONE suspension. Contr. Eng. Pract. 4(8):1101–1108.
9. Raynaud H, Zergainoh A (2000) State-space representation for fractional order controllers. Automatica, 36:1017–1021.
10. Manabe S (1960) The non-integer integral and its application to control systems. JIEE (Japanese Institute of Electrical Engineers) J. 80(860):589–597.
11. Manabe S (1961) The non-integer integral and its application to control systems. ETJ Jpn. 6(3–4):83–87.
12. Oustaloup A (1981) Fractional-order sinusoidal oscillators: optimization and their use in highly linear FM modulators. IEEE Trans. Circ. Syst. 28(10):1007–1009.
13. Axtell M, Bise EM (1990) Fractional calculus applications in control systems. In: Proceedings of the IEEE 1990 National Aerospace and Electronics Conference, pp. 563–566.
14. Vinagre BM, Chen Y (2002) Lecture notes on fractional calculus applications in automatic control and robotics. In the 41st IEEE CDC2002 Tutorial Workshop # 2, B. M. Vinagre and Y. Chen, eds., pp. 1–310. [Online] http://mechatronics.ece.usu.edu/foc/cdc02_tw2_1n.pdf.
15. JATMG (2002) Special issue on fractional calculus and applications. Nonlinear Dynamics, 29(March):1–385.
16. Ortigueira MD, JATMG (eds) (2003) Special issue on fractional signal processing and applications. Signal Processing, 83(11) November: 2285–2480.
17. Oustaloup A (1995) *La Dérivation non Entière*. HERMES, Paris.
18. Oustaloup A, Sabatier J, Lanusse P (1999) From fractal robustness to CRONE control. Fract. Calc. Appl. Anal. 2(1):1–30.
19. Oustaloup A, Mathieu B (1999) *La Commande CRONE: du Scalaire au Multi-variable*. HERMES, Paris.
20. Petrás I (1999) The fractional-order controllers: methods for their synthesis and application. J. Electr. Eng. 50(9–10):284–288.
21. Oustaloup A, Levron F, Nanot F, Mathieu B (2000) Frequency band complex non integer differentiator: characterization and synthesis. IEEE Trans. Circ. Syst. I: *Fundamental Theory and Applications*, 47(1) January: 25–40.
22. Ichise M, Nagayanagi Y, Kojima T (1971) An analog simulation of non-integer order transfer functions for analysis of electrode processes. J. electroanal. chem. 33:253–265.

23. Oldham KB (1973) Semiintegral electroanalysis: analog implementation. *Anal. Chem.* 45(1):39–47.
24. Sugi M, Hirano Y, Miura YF, Saito K (1999) Simulation of fractal immittance by analog circuits: an approach to optimized circuits. *IEICE Trans. Fundam.* (8):1627–1635.
25. Podlubny I, Petráš I, Vinagre BM, O’leary P, Dorcak L (2002) Analogue Realizations of Fractional-Order Controllers. *Nonlinear Dynamics*, 29(1-4) July: 281–296.
26. Machado JAT (1997) Analysis and design of fractional-order digital control systems. *J. Syst. Anal.-Model.-Simul.* 27:107–122.
27. Vinagre BM, Petráš I, Merchan P, Dorcak L (2001) Two digital realisation of fractional controllers: application to temperature control of a solid. In: *Proceedings of the European Control Conference (ECC2001)*, pp. 1764–1767.
28. Vinagre BM, Podlubny I, Hernandez A, Feliu V (2000) On realization of fractional-order controllers. In: *Proceedings of the Conference Internationale Francophone d’Automatique*.
29. Vinagre BM, Podlubny I, Hernandez A, Feliu V (2000) Some approximations of fractional-order operators used in control theory and applications. *Fract. Calc. Appl. Anal.* 3(3):231–248.
30. Chen Y, Moore KL (2002) Discretization schemes for fractional-order differentiators and integrators. *IEEE Trans. Circ. Syst. I: Fundamental Theory and Applications*, 49(3): March. 363–367.
31. Vinagre BM, Chen YQ, Petras I (2003) Two direct Tustin discretization methods for fractional-order differentiator/integrator. *J. Franklin Inst.* 340(5):349–362.
32. Chen YQ, Vinagre BM (2003) A new IIR-type digital fractional order differentiator. *Signal Processing*, 83(11) November: 2359–2365.
33. Al-Alaoui MA (1993) Novel digital integrator and differentiator. *Electron. Lett.* 29(4):376–378.
34. Al-Alaoui MA (1995) A class of second-order integrators and low-pass differentiators. *IEEE Trans. Cir. Syst. I: Fundamental Theory and Applications*, 42(4):220–223.
35. Al-Alaoui MA (1997) Filling the gap between the bilinear and the back-ward difference transforms: an interactive design approach. *Int. J. Electr. Eng. Educ.* 34(4):331–337.
36. Tseng C, Pei S, Hsia S (2000) Computation of fractional derivatives using Fourier transform and digital FIR differentiator. *Signal Processing*, 80:151–159.
37. Tseng C (2001) Design of fractional order digital FIR differentiator. *IEEE Signal Proc. Lett.* 8(3):77–79.
38. Chen Y, Vinagre BM, Podlubny I (2003) A new discretization method for fractional order differentiators via continued fraction expansion. In *Proceedings of The First Symposium on Fractional Derivatives and Their Applications at The 19th Biennial Conference on Mechanical Vibration and Noise, the ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference (ASME DETC2003)*, pp. 1–8, DETC2003/VIB.
39. Oustaloup A, Melchoir P, Lanusse P, Cois C, Dancla F (2000) The CRONE toolbox for Matlab. In: *Proceedings of the 11th IEEE International Symposium on Computer Aided Control System Design - CACSD*.
40. Xue D, Atherton DP (1994) A suboptimal reduction algorithm for linear systems with a time delay. *Int. J. Control*, 60(2):181–196.

41. Åström KJ (1970) *Introduction to Stochastic Control Theory*. Academic Press, London.
42. Press WH, Flannery BP, Teukolsky SA (1986) *Numerical Recipes, the Art of Scientific Computing*. Cambridge University Press, Cambridge.
43. Wang FS, Juang WS, Chan CT (1995) Optimal tuning of PID controllers for single and cascade control loops. Chem. Eng. Comm. 132:15–34.

LINEAR DIFFERENTIAL EQUATIONS OF FRACTIONAL ORDER

Blanca Bonilla¹, Margarita Rivero², and Juan J. Trujillo¹

¹ Departamento de Análisis Matemático, Universidad de la Laguna.

38271 La Laguna-Tenerife, Spain; E-mail: BBonilla@ull.es; JTrujillo@ull.es

² Departamento de Matemática Fundamental, Universidad de la Laguna.

38271 La Laguna-Tenerife, Spain; E-mail: MRivero@ull.es

Abstract

This manuscript presents the basic general theory for sequential linear fractional differential equations, involving the well known Riemann–Liouville fractional operators, D_{a+}^{α} ($a \in \mathbb{R}$, $0 < \alpha \leq 1$)

$$\mathbf{L}_{n\alpha}(y) = \left[\mathcal{D}_{a+}^{n\alpha} + \sum_{k=0}^{n-1} a_k(x) \mathcal{D}_{a+}^{k\alpha} \right] (y) = y^{(n\alpha)} + \sum_{k=0}^{n-1} a_k(x) y^{(k\alpha)} = f(x). \quad (1)$$

where $\{a_k(x)\}_{k=0}^{n-1}$ are continuous real functions defined in $[a, b] \subset \mathbb{R}$ and

$$\begin{aligned} \mathcal{D}_{a+}^{\alpha} &= D_{a+}^{\alpha} \\ \mathcal{D}_{a+}^{k\alpha} &= \mathcal{D}_{a+}^{\alpha} \mathcal{D}_{a+}^{(k-1)\alpha}. \end{aligned} \quad (2)$$

We also consider the case where $f(x)$ is a continuous real function in $(a, b] \subset \mathbb{R}$ and $f(a) = o(x^{\alpha-1})$.

We then introduce the Mittag-Leffler-type function $e_{\alpha}^{\lambda x}$, which we will call α -exponential. This function is the product of a Mittag-Leffler function and a power function. This function allows us to directly obtain the general solution to homogeneous and non-homogeneous linear fractional differential equations with constant coefficients. This method is a variation of the usual one for the ordinary case.

Keywords

Fractional differential equations, Caputo, Riemann–Liouville, linear.

1 Introduction

Questions as to what we mean by, and where we could apply, the fractional calculus operators have fascinated us all ever since 1695 when the so-called fractional calculus was conceptualised in connection with the infinitesimal

calculus, see [15] and [13]. A rigorous and encyclopedic study of fractional operators can be found in [17].

It is known that the classical calculus provides a powerful tool for explaining and modelling many important dynamic processes in most applied sciences. But experiments and reality teach us that there are many complex systems in nature and society with anomalous dynamics, such as charge transport in amorphous semiconductors, the spread of contaminants in underground water, relaxation in viscoelastic materials like polymers, the diffusion of pollution in the atmosphere, and many more.

In most of the above-mentioned cases, this kind of anomalous process has a complex macroscopic behaviour, the dynamics of which cannot be characterised by classical derivative models. Nevertheless, a heuristic solution to the corresponding models of some of those processes can be frequently obtained using tools from statistical physics. For such an explanation, one must use some generalised concepts from classical physics such as fractional Brownian motion, the continuous time random walk (CTRW) method involving Lévy stable distributions (instead of Gaussian distributions), the generalised central limit theorem (instead of the classical central limit theorem), and non-Markovian distributions which means non-local distributions (instead of the classical Markovian ones). From this approach it is also important to note that the anomalous behaviour of many complex processes includes multi-scaling in the time and space variables.

The above-mentioned tools have been used extensively during last 30 years. But the connection between these statistical models and certain fractional differential equations involving the fractional integral and derivative operators (Riemann–Liouville, Caputo, Liouville or Weyl, Riesz, etc.; see [17]) has only been formally established during the last 15 years; (see, for instance, [10], [9] [19], [14]).

We could ask ourselves, what are the useful properties of these fractional calculus operators, which help in the modelling of so many anomalous processes? From the point of view of the authors and from known experimental results, most of the processes associated with complex systems have non-local dynamics involving long memory in time, and the fractional integral and fractional-derivative operators do have some of those characteristics. Perhaps this is one of the reasons why these fractional calculus operators lose the above-mentioned useful properties of the ordinary derivative D .

This manuscript is organised as follows. Sections 2 and 3 presents some fractional operators and their main properties and introduce some types of Mittag-Leffler functions. In section 4 we develop a general theory for sequential linear fractional differential equations, while in section 5 we introduce a new direct method for solving the homogeneous and non-homogeneous case with constant coefficients, using the α -exponential function and certain fractional Green functions, including some illustrative examples.

We must point out that the Laplace or Fourier transform can only be used for the case where the fractional derivative involved in the fractional differential equation is D_{0+}^α , and never when we want to use the more general fractional derivative D_{a+}^α ($a < 0$), as must be done when the initial conditions of the corresponding model are given in the origin. On the other hand, it is clear that those mentioned integral transforms are not utile when the problem involve a distributional delta function as a initial condition.

2 Fractional operators

We will consider here the so-called sequential Riemann–Liouville and Caputo derivatives. See [17] and [1].

Let $\alpha \in \mathbb{R}$ ($\alpha > 0$), $m - 1 < \alpha \leq m$, $m \in \mathbb{N}$, $[a, b] \subset \mathbb{R}$ and f be a measurable function, that is $f \in L_1(a, b)$. Then the Riemann–Liouville integral operator of order α is defined by

$$(I_{a+}^\alpha f)(x) = \frac{1}{\Gamma(\alpha)} \int_a^x (x-t)^{\alpha-1} f(t) dt \quad (x > a), \quad (3)$$

and the corresponding Riemann–Liouville fractional derivative by

$$(D_{a+}^\alpha f)(x) = [D^m I_{a+}^{m-\alpha} f](x), \quad (4)$$

where $D = \frac{d}{dx}$ is the ordinary derivative.

Let us remember that, in general, when $\alpha, \beta \in \mathbb{R}^+$, the operators $D_{a+}^{\alpha+\beta}$ and $D_{a+}^\alpha D_{a+}^\beta$ are different. Also, as usual, we will use $AC([a, b])$ to refer to the set of absolutely continuous functions in $[a, b]$, and $AC^n([a, b])$ ($n \in \mathbb{N}$), for the set of functions f , such that there exist $(D^n)(f) = f^{(n)}$ in $[a, b]$ and $f^{(n)} \in AC$.

Property 1. Let $n - 1 \leq \alpha < n$, $m - 1 \leq \beta < m$. If $f \in L_1(a, b)$ with $f_{m-\beta} \in AC^{m+1}([a, b])$ and $f_{n-(\alpha+\beta)} \in AC^{n-1}([a, b])$ if $\alpha + \beta < n$ (or $f_{\{\alpha+\beta\}} \in AC^{\alpha+\beta}([a, b])$ if $\alpha + \beta > n$), where $f_{n-\alpha} = (I_{a+}^{n-\alpha} f)(x)$. Then we have the following index rule

$$(D_{a+}^\alpha D_{a+}^\beta f)(x) = (D_{a+}^{\alpha+\beta} f)(x) - \sum_{j=1}^m (D_{a+}^{\beta-j} f)(a+) \frac{(x-a)^{-j-\alpha}}{\Gamma(1-j-\alpha)}, \quad (5)$$

almost everywhere in $[a, b]$.

The following Property holds from the rule for the parametric derivation under the integral sign (see [14]).

Property 2. Let $0 < \eta \leq 1$, $(D_{a+}^\eta K) \in L_1(a, b)$ with a suitable f (for example, $f \in C([a, b])$). Then we have

$$D_{a+}^{\eta} \int_a^x K(x-t)f(t)dt = \int_a^x [D_{a+}^{\eta} K(x-a)](t)f(x-t+a)dt + f(x) \lim_{x \rightarrow a+} \left[I_{a+}^{1-\eta} K(t-a) \right](x). \quad (6)$$

As expected, a fractional differential equation of order α_n is an equation such as

$$F(x, y(x), (D^{\alpha_1}y)(x), (D^{\alpha_2}y)(x), \dots, (D^{\alpha_n}y)(x)) = g(x), \quad (7)$$

with $\alpha_1 < \alpha_2 < \dots < \alpha_n$, $F(x, y_1, \dots, y_n)$ and $g(x)$ known real functions, D^{α_k} ($k = 1, 2, \dots, n$) fractional differential operators and where $y(x)$ is the unknown function.

In 1993 Miller–Ross [11] introduced the sequential fractional derivative $\mathcal{D}^{\alpha}$ in the following way

$$\begin{aligned} \mathcal{D}^{\alpha} &= D^{\alpha}, \quad (0 < \alpha \leq 1) \\ \mathcal{D}^{k\alpha} &= \mathcal{D}^{\alpha} \mathcal{D}^{(k-1)\alpha}, \quad (k = 2, 3, \dots), \end{aligned} \quad (8)$$

where D^{α} is a fractional derivative.

A sequential fractional differential equation of order $n\alpha$ has the following relationship

$$F(x, y(x), (\mathcal{D}^{\alpha}y)(x), (\mathcal{D}^{2\alpha}y)(x), \dots, (\mathcal{D}^{n\alpha}y)(x)) = g(x). \quad (9)$$

Let $D^{\alpha} = D_{a+}^{\alpha}$ be the Riemann–Liouville fractional derivative. Then, taking into account Property 1, we can obtain the relation between $\mathcal{D}_{a+}^{n\alpha}$ and $D_{a+}^{n\alpha}$. When $n = 2$ such relation is given by

$$(\mathcal{D}_{a+}^{2\alpha}y)(x) = D_{a+}^{2\alpha} \left[y(x) - (I_{a+}^{1-\alpha}y)(a+) \frac{(x-a)^{\alpha-1}}{\Gamma(\alpha)} \right]. \quad (10)$$

On the other hand, if $\alpha = \frac{n}{p}$ ($n, p \in \mathbb{N}$) and $y(x)$ is a continuous real function defined in $[a, b]$, that is $y \in C([a, b])$, we can deduce from Property 1 the important property:

$$(D^n y)(t) = (\mathcal{D}_{a+}^{p\alpha} y)(t), \quad (t > a). \quad (11)$$

In this paper we study the linear sequential fractional differential equations of order $n\alpha$ which can be written in the following normalised form

$$\mathbf{L}_{n\alpha}(y) = \left[\mathcal{D}_{a+}^{n\alpha} + \sum_{k=0}^{n-1} a_k(x) \mathcal{D}_{a+}^{k\alpha} \right] (y) = y^{(n\alpha)} + \sum_{k=0}^{n-1} a_k(x) y^{(k\alpha)} = f(x), \quad (12)$$

where $\{a_k(x)\}_{k=0}^{n-1}$ are continuous real functions defined in an interval $[a, b] \subset \mathbb{R}$ and $f(x) \in C([a, b])$ or $f(x) \in C((a, b])$.

The existence and uniqueness of solutions to the Cauchy-type problem for fractional differential Eq. (12) was established in [4], [6], and [7] for different kinds of functional spaces. We present below two of the theorems which will be used in this paper.

Theorem 1. Let $x_0 \in (a, b) \subset \mathbb{R}$ and $\{y_0^k\}_{k=0}^{n-1} \in \mathbb{R}^n$. Let $f(x)$ and $\{a_k(x)\}_{k=0}^{n-1}$ be continuous real functions in $[a, b]$. Then there exists a unique continuous function $y(x)$ defined in $(a, b]$, which is a solution to the Cauchy-type problem

$$[\mathbf{L}_{n\alpha}(y)](x) = f(x) \quad (13)$$

$$(\mathcal{D}_{a+}^{k\alpha} y)(x_0) = y^{(k\alpha)}(x_0) = y_0^k \quad (k = 0, 1, \dots, n-1), \quad (14)$$

Moreover, this solution $y(x)$ satisfies

$$\lim_{x \rightarrow a+} (x - a)^{1-\alpha} y(x) < \infty, \quad (15)$$

and

$$(I_{a+}^{1-\alpha} y)(x) < \infty. \quad (16)$$

We denote with $\mathcal{C}_\gamma([a, b])$ ($\gamma \in \mathbb{R}$) the Banach space

$$\mathcal{C}_\gamma([a, b]) = \{g(x) \in C([a, b]) : \|g\|_{\mathcal{C}_\gamma} = \|(x - a)^\gamma g(x)\|_C < \infty\}. \quad (17)$$

In particular $\mathcal{C}_0([a, b]) = C([a, b])$.

Theorem 2. Let $\{a_k(x)\}_{k=0}^{n-1}$ be continuous functions in $[a, b]$, $f \in \mathcal{C}_{1-\alpha}([a, b])$ and $\{b_k\}_{k=0}^{n-1} \in \mathbb{R}^n$. Then there exists a unique continuous function $y(x)$ defined in $(a, b]$ which is a solution to the linear sequential fractional differential equation of order $n\alpha$

$$[\mathbf{L}_{n\alpha}(y)](x) = f(x), \quad (18)$$

and such that

$$\lim_{x \rightarrow a+} (x - a)^{1-\alpha} (\mathcal{D}_{a+}^{k\alpha} y)(x) = b_k \quad (19)$$

or such that

$$(I_{a+}^{1-\alpha} \mathcal{D}_{a+}^{k\alpha} y)(a+) = b_k. \quad (20)$$

For the particular case $f(x) = 0$ we have the following

Corollary 1. Let $x_0 \in (a, b]$, (or $x_0 = a$). Let $\{a_k(x)\}_{k=0}^{n-1}$ be continuous real functions defined in $(a, b]$ and such that $(x - a)^{1-\alpha} a_k(x)|_{x=a} < \infty$, $\forall k = 1, 2, \dots, n$. The homogeneous linear sequential fractional differential equation

$$[\mathbf{L}_{n\alpha}(y)](x) = 0 \quad (21)$$

has $y(x) = 0$ as the unique solution in $(a, b]$, satisfying the initial conditions

$$y^{(j\alpha)}(x_0) = 0 \quad (o \quad [(x - a)^{1-\alpha} y^{(k\alpha)}(x)]_{x=a+} = 0) \quad (k = 0, 1, \dots, n-1).$$

3 α -Exponential functions

In this section we introduce two special functions of the Mittag-Leffler type, which will be used in the next sections. See, for instance, [16], [12], [2], [5], and [3].

Definition 1. Let $\lambda, \nu \in \mathbb{C}$, $\alpha \in \mathbb{R}^+$ and $a \in \mathbb{R}$. We will call α -exponential function $e_\alpha^{\lambda(x-a)}$ the Mittag-Leffler-type function

$$e_\alpha^{\lambda(x-a)} = (x-a)^{\alpha-1} \sum_{k=0}^{\infty} \frac{\lambda^k (x-a)^{k\alpha}}{\Gamma[(k+1)\alpha]} \quad (x > a). \quad (22)$$

This function satisfies the following properties:

Proposition 1. Under the restrictions of definition 1, it is easy to prove the following properties

$$i) \quad D_{a+}^\alpha e_\alpha^{\lambda(x-a)} = \lambda e_\alpha^{\lambda(x-a)}. \quad (23)$$

$$ii) \quad e_\alpha^{\lambda(x-a)} = (x-a)^{\alpha-1} E_{\alpha,\alpha}(\lambda(x-a)^\alpha). \quad (24)$$

where $E_{\beta,\eta}(x-a)$ is the Mittag-Leffler function

$$E_{\beta,\eta}(x-a) = \sum_{k=0}^{\infty} \frac{(x-a)^k}{\Gamma(\beta k + \eta)} \quad (\eta, \beta \in \mathbb{R}^+).$$

$$iii) \quad \mathcal{L}(e_\alpha^{\lambda x}) = \frac{1}{s^\alpha - \lambda} \quad (|s|^\alpha < |\lambda|), \quad (25)$$

where $\mathcal{L}$ denotes the Laplace transform.

Definition 2. Let $\alpha \in \mathbb{R}^+$, $l \in \mathbb{N}_0$, $a \in \mathbb{R}$ and $\lambda = b + ic \in \mathbb{C}$. We will call $\mathcal{E}_{\alpha,l}^{\lambda x}$ the Mittag-Leffler-type function

$$\mathcal{E}_{\alpha,l}^{\lambda(x-a)} = (x-a)^{\alpha-1} \sum_{k=0}^{\infty} \frac{(l+k)!}{\Gamma[(k+l+1)\alpha]} \frac{(\lambda(x-a)^\alpha)^k}{k!} \quad (x > a). \quad (26)$$

Proposition 2. Under the restrictions of definition 2, it is easy to prove the following properties

$$i) \quad \frac{\partial^l}{\partial \lambda^l} e_\alpha^{\lambda(x-a)} = (x-a)^{l\alpha} \mathcal{E}_{\alpha,l}^{\lambda(x-a)}. \quad (27)$$

$$ii) \quad \mathcal{E}_{\alpha,l}^{\lambda x} = l! x^{\alpha-1} E_{\alpha,(l+1)\alpha}(\lambda x^\alpha). \quad (28)$$

$$iii) \quad \mathcal{L}\{x^{\alpha l} \mathcal{E}_{\alpha,l}^{\lambda x}\} = \frac{l!}{(s^\alpha - \lambda)^{l+1}} \quad (|s|^\alpha < |\lambda|). \quad (29)$$

4 General Theory for Linear Fractional Differential Equations

In this section we study the solutions to a homogeneous linear sequential fractional-differential equation

$$\mathbf{L}_{n\alpha}(y) = \left[\mathcal{D}_{a+}^{n\alpha} + \sum_{k=0}^{n-1} a_k(x) \mathcal{D}_{a+}^{k\alpha} \right] (y) = y^{(n\alpha)} + \sum_{k=0}^{n-1} a_k(x) y^{(k\alpha)} = 0, \quad (30)$$

where $\{a_k(x)\}_{k=0}^{n-1}$ are continuous real functions in $[a, b]$ and $[\mathcal{D}_{a+}^{n\alpha}](y) = y^{(n\alpha)}$ is the sequential Riemann–Liouville fractional derivative.

Definition 3. *As usual, a fundamental set of solutions to equation (30) in some interval $V \subset [a, b]$ is a set of n functions linearly independent in V , which are solutions to (30).*

Definition 4. *The α -Wronskian of the n functions $\{u_k(x)\}_1^n$, which admit iterated fractional derivatives up to order $(n-1)\alpha$ in some interval $V \subset (a, b]$, refers to the following determinant*

$$|\mathbf{W}_\alpha(u_1, \dots, u_n)(x)| = \begin{vmatrix} u_1(x) & u_2(x) & \cdot & \cdot & u_n(x) \\ u_1^{(\alpha)}(x) & u_2^{(\alpha)}(x) & \cdot & \cdot & u_n^{(\alpha)}(x) \\ u_1^{(2\alpha)}(x) & u_2^{(2\alpha)}(x) & \cdot & \cdot & u_n^{(2\alpha)}(x) \\ \dots & \cdot & \cdot & \cdot & \cdot \\ \dots & \cdot & \cdot & \cdot & \cdot \\ u_1^{((n-1)\alpha)}(x) & u_2^{((n-1)\alpha)}(x) & \dots & \cdot & u_n^{((n-1)\alpha)}(x) \end{vmatrix}. \quad (31)$$

To simplify the notation, this will be represented by $|\mathbf{W}_\alpha(x)| = |\mathbf{W}_\alpha(u_1, \dots, u_n)(x)|$. We will use $\mathbf{W}_\alpha(x)$ for the corresponding Wronskian matrix.

Theorem 3. *Let $\{u_k(x)\}_{k=1}^n$ be a family of functions with sequential fractional derivatives up to order $(n-1)\alpha$ in $(a, b]$ and such that, if $j = 1, 2, \dots, n$ and $k = 0, 1, \dots, n-1$*

$$\lim_{x \rightarrow a+} [(x-a)^{1-\alpha} u_j^{(k\alpha)}(x)] < \infty. \quad (32)$$

If the functions $\{(x-a)^{1-\alpha} u_j(x)\}_{j=1}^n$ are linearly dependent in $[a, b]$, it follows that for all $x \in [a, b]$

$$(x-a)^{n-n\alpha} |\mathbf{W}_\alpha(x)| = 0. \quad (33)$$

We can complete the above result, as in the ordinary case, with the following theorem

Theorem 4. *Let $\{u_k(x)\}_{k=1}^n$ be a solution family of functions to Eq. (30) in $(a, b]$ which satisfies*

$$\lim_{x \rightarrow a+} [(x-a)^{1-\alpha} u_j(x)] < \infty \quad (j = 1, 2, \dots, n).$$

Then the functions

$$\{(x-a)^{1-\alpha} u_j(x)\}_{j=1}^n$$

are linearly dependent in $[a, b]$ if, and only if, there exists an $x_0 \in [a, b]$ such that

$$[(x-a)^{n-n\alpha} |\mathbf{W}_\alpha(x)|]_{x=x_0} = 0 \quad (34)$$

From the above theorem we can always find, in a way similar to the ordinary case, a fundamental set of solutions for Eq. (30) in some interval $V \subset [a, b]$.

Usually, the general solution to a non-homogeneous linear sequential fractional differential equation

$$\mathbf{L}_{\mathbf{n}\alpha}(y) = f(x). \quad (35)$$

will be given as in the following proposition:

Proposition 3. *If $y_p(x)$ is a particular solution to (35) and $y_h(x)$ is a general solution to the corresponding homogeneous equation*

$$\mathbf{L}_{\mathbf{n}\alpha}(y) = 0, \quad (36)$$

that is,

$$y_h(x) = \sum_{k=1}^n c_k u_k(x), \quad (37)$$

with $\{c_k\}_{k=1}^n$ arbitrary real constants and $\{u_k(x)\}_{k=1}^n$ a fundamental set of (36), then a general solution to the non-homogeneous Eq. (35) is

$$y_g(x) = y_h(x) + y_p(x), \quad (38)$$

A general theory, similar to the above, can be established for the Caputo fractional derivative $D^\alpha \equiv {}^C D_{a+}^\alpha$, which was introduced by Caputo in 1969, see, for instance, [1].

$$({}^C D_{a+}^\alpha f)(x) = (I_{a+}^{n-\alpha} D^n f)(x) \quad (x > a \text{ and } n = -[-\alpha]). \quad (39)$$

Also it is usual to consider the following, more general, definition for the Caputo fractional derivative

$$({}^C D_{a+}^\alpha f)(x) = D_{a+}^\alpha \left[f(x) - \sum_{j=0}^{n-1} f^{(j)}(a+) \frac{(x-a)^j}{j!} \right], \quad (40)$$

which shows the close connection between the Caputo and the Riemann-Liouville derivatives.

5 Linear Sequential Fractional Differential Equations with Constant Coefficients

In this section we present a direct method for obtaining the explicit general solution to a linear sequential fractional differential equation with constant coefficients, such as

$$\mathbf{L}_{n\alpha}(y) = \left[\mathcal{D}_{a+}^{n\alpha} + \sum_{k=0}^{n-1} a_k \mathcal{D}_{a+}^{k\alpha} \right] (y) = f(x), \quad (41)$$

where a and $\{a_k\}_{k=0}^{n-1}$ are real constants and $\mathcal{D}_{a+}^{k\alpha}$ is the Riemann–Liouville sequential fractional derivative.

Several approaches have been developed for obtaining explicit solutions to some of these types of equations. The Laplace method was discussed by some authors, see, for instance, [11], [1], and [14], but this approach is applicable only if $a = 0$. With the restriction $a = 0$, it is not possible to consider Cauchy-type problems for Eq. (41) with conditions at $x = 0$. On the other hand, the direct method is very convenient for studying and solving boundary-value problems associated with Eq. (41) which cannot be solved by the Laplace method. At the end, we will introduce a fractional Green function to obtain an explicit particular solution to the non-homogeneous Eq. (41).

Let us consider now the corresponding homogeneous Eq. to (41)

$$\mathbf{L}_{n\alpha}(y) = \left[\mathcal{D}_{a+}^{n\alpha} + \sum_{k=0}^{n-1} a_k \mathcal{D}_{a+}^{k\alpha} \right] (y) = 0. \quad (42)$$

As in the ordinary case, if we try to find solutions to (42) of the type $y(x) = e_{\alpha}^{\lambda(x-a)}$, it follows that

$$\mathbf{L}_{n\alpha} \left(e_{\alpha}^{\lambda(x-a)} \right) = P_n(\lambda) e_{\alpha}^{\lambda(x-a)} \quad (43)$$

where

$$P_n(\lambda) = \lambda^n + \sum_{k=1}^{n-1} a_k \lambda^k, \quad (44)$$

is referred to as the characteristic polynomial associated with Eq. (42).

In the following it will be assumed that $\lambda \in \mathbb{C}$.

By the use of the properties of the α -exponential function, we obtain the following result

Lemma 1. *If λ is a root of characteristic polynomial (44), then*

$$\frac{\partial}{\partial \lambda} \mathbf{L}_{n\alpha} \left(e_{\alpha}^{\lambda(x-a)} \right) = \mathbf{L}_{n\alpha} \left(\frac{\partial}{\partial \lambda} e_{\alpha}^{\lambda(x-a)} \right) \quad (45)$$

and

$$\frac{\partial^l}{\partial \lambda^l} e_{\alpha}^{\lambda(x-a)} = (x-a)^{l\alpha} \mathcal{E}_{\alpha,l}^{\lambda(x-a)}. \quad (46)$$

So we can connect the solution of the characteristic polynomial (44) with solutions of (42) as in the usual case

Theorem 5. *Let $\{\lambda_j\}_{j=1}^k$ be all different real roots of the characteristic polynomial (44), whose orders of multiplicity are $\{\mu_j\}_{j=1}^k$, respectively. Let $\{r_j, \bar{r}_j\}_{j=1}^p$ ($r_j = b_j + ic_j$) be all distinct pairs of complex conjugate solutions of multiplicity $\{\sigma_j\}_{j=1}^p$, respectively, of (44). Then the union set of the sets*

$$\bigcup_{m=1}^k \left\{ (x-a)^{l\alpha} \mathcal{E}_{\alpha,l}^{\lambda_m(x-a)} \right\}_{l=1}^{\mu_m-1}, \quad (47)$$

$$\bigcup_{m=1}^p \left\{ \sum_{j=0}^{\infty} (-1)^j \frac{c_m^{2j}}{(2j)!} (x-a)^{(2j+l)\alpha} \mathcal{E}_{\alpha,l+2j}^{b_m(x-a)} \right\}_{l=1}^{\sigma_m-1} \quad (48)$$

and

$$\bigcup_{m=1}^p \left\{ \sum_{j=0}^{\infty} (-1)^j \frac{c_m^{2j+1}}{(2j+1)!} (x-a)^{(2j+l+1)\alpha} \mathcal{E}_{\alpha,l+2j+1}^{b_m(x-a)} \right\}_{l=1}^{\sigma_m-1}, \quad (49)$$

determines a fundamental system of solutions to fractional differential equation (42).

Note that only for the case where $a = 0$ can operational methods such as the Laplace transform be applied to solve the problem of constant coefficients.

Example 1. Let us consider the equation

$$\mathcal{D}_{a+}^{2\alpha} y + \lambda^2 y = 0. \quad (50)$$

Its characteristic equation is $P_2(x) = x^2 + \lambda^2 = (x - \lambda i)(x + \lambda i)$ and so the fundamental set of solutions to (50) is

$$\{\cos_\alpha[\lambda(x-a)], \sin_\alpha[\lambda(x-a)]\},$$

where

$$\cos_\alpha[\lambda(x-a)] = \sum_{j=0}^{\infty} (-1)^j \lambda^{(2j+1)} \frac{(x-a)^{(j+1)2\alpha-1}}{\Gamma[(j+1)2\alpha]} \quad (51)$$

and

$$\sin_\alpha[\lambda(x-a)] = \sum_{j=0}^{\infty} (-1)^j \lambda^{2j} \frac{(x-a)^{(2j+1)\alpha-1}}{\Gamma[(2j+1)\alpha]}. \quad (52)$$

These new functions $\sin_\alpha(x)$ and $\cos_\alpha(x)$ are a generalisation of the usual $\cos(x)$ and $\sin(x)$.

Since now we know how to obtain the general solution to homogeneous Eq. (42), then, in accordance with Proposition 4, to obtain the explicit general solution to (41) we only need to get a particular solution to (41).

First of all we will obtain the general solution to the simpler equation

$$y^{(\alpha)} - \lambda y = f(x) \quad (x > a) \quad (53)$$

where $y^{(\alpha)} = D_{a+}^{\alpha} y$.

Proposition 4. *Let $f \in L_1(a, b) \cap C([a, b])$. Then Eq. (53) admits*

$$y_g = ce_{\alpha}^{\lambda(x-a)} + y_p, \quad (54)$$

as a general solution in which

$$y_p = e_{\alpha}^{\lambda x} *^a f(x), \quad (55)$$

is a particular solution to (53), with $*^a$ being the following convolution

$$g(x) *^a f(x) = \int_a^x g(x-t)f(t)dt. \quad (56)$$

In addition, $y_p(a+) = 0$, if $f(x) \in C([a, b])$ and $(I_{a+}^{1-\alpha} y_p)(a+) = 0$, if $f(x) \in \mathcal{C}_{1-\alpha}([a, b])$.

Proof. It is sufficient to verify that $y_p(x)$ is a solution to (53). For this, if we apply Property 5 and we keep in mind (23) and that

$$\lim_{x \rightarrow a+} \left(I_{a+}^{1-\alpha} e_{\alpha}^{\lambda(t-a)} \right) (x) = 1,$$

then

$$\begin{aligned} (D_{a+}^{\alpha} y_p)(x) &= D_{a+}^{\alpha} \int_a^x e_{\alpha}^{\lambda(x-t)} f(t) dt \\ &= \int_a^x \{D_{a+}^{\alpha} e_{\alpha}^{\lambda(x-a)}\}(t) f(x-t+a) dt + f(x) \lim_{t \rightarrow a+} \{I_{a+}^{1-\alpha} e_{\alpha}^{\lambda(x-a)}\}(t) \\ &= \lambda \int_a^x e_{\alpha}^{\lambda(x-\xi)} f(\xi) d\xi + f(x) = \lambda y_p + f(x), \end{aligned}$$

which concludes the proof.

Theorem 6. *A particular solution to Eq. (41) is given by*

$$y_p = G_{\alpha}(x) *^a f(x) \quad (57)$$

where $G_{\alpha}(x)$ is

$$G_{\alpha}(x) = \prod_{j=1}^k *^a \left(\prod_{l=1}^{\sigma_j} e_{\alpha}^{\lambda_j(x-a)} \right) \quad (58)$$

where $\{\lambda_j\}_{j=1}^k$ are the k distinct complex roots of the characteristic polynomial (44) with multiplicity $\{\sigma_j\}_{j=1}^k$, respectively.

In addition, $y_p(a+) = 0$ if $f(x) \in C([a, b])$ and $(I_{a+}^{1-\alpha} y_p)(a+) = 0$ if $f(x) \in \mathcal{C}_{1-\alpha}([a, b])$. Moreover $(I_{a+}^{1-\alpha} G_{\alpha})(a+) = 0$.

Proof. It is sufficient to successively apply the result of Proposition 5 while keeping in mind the weak singularity presented by the function $e_\alpha^{\lambda x}$.

Remark 1. Since function $G_\alpha(x - \xi)$ plays the role of Green's function associated with non-homogeneous Eq. (41), analogous to the usual case, this function will be called Riemann–Liouville fractional Green's function.

Remark 2. Analogous results can be obtained if we consider the Caputo fractional derivative (39) or (40) instead of the Riemann–Liouville fractional derivative, by using the Mittag-Leffler function

$$E_\alpha(\lambda(x - a)) = \sum_{k=0}^{\infty} \frac{\lambda^k (x - a)^{k\alpha}}{\Gamma(\alpha k + 1)} \quad (\alpha > 0) \quad (59)$$

instead of the α -exponential function $e_\alpha^{\lambda(x-a)}$.

Example 2. Let us consider the equation

$${}^C\mathcal{D}_{a+}^{2\alpha} y + \lambda^2 y = 0. \quad (60)$$

Its corresponding characteristic polynomial is $P_2(x) = x^2 + \lambda^2$ and so the fundamental set of solutions to (60) is

$$\{\cos_\alpha^*[\lambda(x - a)], \sin_\alpha^*[\lambda(x - a)]\} \quad (61)$$

where

$$\cos_\alpha^*[\lambda(x - a)] = \operatorname{Re}\{E_\alpha(\lambda(x - a)^\alpha)\} \quad (62)$$

and

$$\sin_\alpha^*[\lambda(x - a)] = \operatorname{Im}\{E_\alpha(\lambda(x - a)^\alpha)\}. \quad (63)$$

We point out here that the $\sin_\alpha^*(x)$ and $\cos_\alpha^*(x)$ functions are a new generalisation of the usual $\cos(x)$ and $\sin(x)$ functions, which, like the $\sin_\alpha(x)$ and $\cos_\alpha(x)$ functions, could play a fundamental role, for instance, in the development of a fractional Fourier theory, or of Weierstrass-type fractal functions, which are solutions to elementary fractional differential equations.

In addition, the results previously presented may be applied to Riemann–Liouville **non-sequential** linear fractional differential equations. It is easy to prove the following:

Corollary 2. Let $f \in \mathcal{C}_{1-\alpha}([a, b])$ and $a_0, a_1 \in R$. Then equation

$$D_{a+}^{2\alpha} y + a_1 D_{a+}^\alpha y + a_0 y = f(x) \quad (0 < \alpha \leq 1) \quad (64)$$

has the general solution

$$y(x) = C_1 z_1(x) + C_2 z_2(x) + z_p(x) - \frac{C}{\Gamma(\alpha)} (x - a)^{\alpha-1}, \quad (65)$$

where z_i ($i = 1, 2$) is a fundamental system of solutions to the homogeneous sequential fractional differential equation

$$\mathcal{D}_{a+}^{2\alpha} z + a_1 \mathcal{D}_{a+}^{\alpha} z + a_0 z = 0, \quad (66)$$

and

$$z_p(x) = z_1(x) *^a z_2(x) *^a [f(x) + a_0 C(x-a)^{\alpha-1}] \quad (67)$$

is a particular solution to the non-homogeneous equation

$$\mathcal{D}_{a+}^{2\alpha} z + a_1 \mathcal{D}_{a+}^{\alpha} z + a_0 z = f(x) + a_0 C(x-a)^{\alpha-1} \quad (68)$$

where C , C_1 and C_2 are real constants such that $C_1 + C_2 = C$ if the roots of the characteristic Eq. of (66) are different, or $C_1 = C$, if they are not.

Example 3. Let $0 < \alpha \leq 1$ and $f \in \mathcal{C}_{1-\alpha}([a, b])$. A general solution to equation

$$D_{a+}^{2\alpha} y - 2D_{a+}^{\alpha} y + y = f(x) \quad (x > a), \quad (69)$$

is

$$y_g(x) = C e_{\alpha}^{(x-a)} + C_2 \mathcal{E}_{\alpha,1}^{(x-a)} + u(x) - \frac{C}{\Gamma(\alpha)} (x-a)^{\alpha-1} \quad (70)$$

C_2 and C being two arbitrary real constants, and

$$u(x) = e_{\alpha}^{(x-a)} *^a \mathcal{E}_{\alpha,1}^{(x-a)} *^a \left[f(x) + \frac{C}{\Gamma(\alpha)} (x-a)^{\alpha-1} \right], \quad (71)$$

Example 4. The ordinary differential equation

$$x'(t) - a^2 x(t) = 0, \quad (72)$$

according to the relation given in (11), may be transformed into the sequential linear fractional differential equation

$$(\mathcal{D}_{0+}^{2\alpha} x)(t) - a^2 x(t) = 0 \quad (\alpha = 1/2), \quad (73)$$

whose general solution is

$$x(t) = C_1 e_{\alpha}^{at} + C_2 e_{\alpha}^{-at}. \quad (74)$$

Any solution to (72) is included in the family of solutions to (74) because $x(0) < \infty$ and so $C_2 = -C_1$. Then

$$x(t) = C_1 \sum_{j=1}^{\infty} \frac{[1 - (-1)^j] a^j t^{j\alpha+\alpha-1}}{\Gamma[(j+1)\alpha]} \quad (75)$$

which is the well-known general solution to (72).

However, $x(t) = e_{\alpha}^{at}$ is a solution to (73) but it is not a solution to (72).

Acknowledgment

This work was supported, in part, by DGUI of G.A.CC (PI2003/133), by MEC (MTM2004-00327) and by ULL. This paper is a new version of a paper published in proceedings of IDETC/CIE 2005, September 24–28, Long Beach, California, USA. The authors want to express explicitly their gratitude to the ASME for its kind disposition to permit them publish a revised version of the paper as a chapter of this book.

References

1. Carpinteri A, Mainardi F (eds.) (1997) *Fractals and Fractional Calculus in Continuum Mechanics*, CIAM Courses and Lectures 378. Springer, New York.
2. Erdélyi A, Magnus W, Oberhettinger F, Tricomi FG (1953) *Higher Transcendental Functions*, Vol. I. McGraw-Hill, New York.
3. Gorenflo R, Kilbas AA, Rogosin SV (1998) On the generalized Mittag-Leffler Type functions, *Int. Trans. Spec. Funct.* 7:215–224.
4. Hayek N, Trujillo JJ, Rivero M, Bonilla B, Moreno JC (1999) An extension of Picard-Lindelöf theorem to fractional differential equations, *Appl. Anal.* 70(3–4):347–361.
5. Humbert P, Agarwal RP (1953) Sur la fonction de Mittag-Leffler et quelques-unes de ses généralisations, *Bull. Sci. Math. Ser. 2* 77:180–185.
6. Kilbas AA, Bonilla B, Trujillo JJ (2000) Fractional integrals and derivatives, and differential equations of fractional order in weighted spaces of continuous functions, *Dokl. Math.* 2(62):222–226.
7. Kilbas AA, Bonilla B, Trujillo JJ (2000) Existence and uniqueness theorems for nonlinear fractional differential equations, *Demonstratio Math.* 3(33):583–602.
8. Kilbas AA, Pierantozzi T, Vázquez L, Trujillo JJ (2004) On solution of fractional evolution equation, *J. Phys. A: Math Gen.* 37:1–13.
9. Kilbas AA, Srivastava HM, Trujillo JJ (2003) Fractional differential equations: an emergent field in applied and mathematical sciences, in: *Factorization, Singular Operators and Related Problems* (S. Samko, A. Lebre and A.F. dos Santos (eds.), Kluwer Academic London), pp. 151–174.
10. Metzler R, Klafter J (2000) The random walk's guide to anomalous diffusion: A fractional dynamic approach, *Phys. Rep.* 1(339)1–77.
11. Miller KS, Ross B (1993) *An Introduction to the Fractional Calculus and Fractional Differential Equations*. Wiley, New York.
12. Mittag-Leffler MG (1903) Sur la nouvelle fonction, *Compt. Rend. Acad. Sci. Paris*, 137:554–558.
13. Oldham KB, Spanier J (1974) *The Fractional Calculus*. Academic Press, New York.
14. Podlubny I (1999) Fractional differential equations, in: *Mathematics in Science and Engineering* 198. Academic Press, London.
15. Ross B (1977) The development of fractional calculus: 1695–1900, *Hist. Math.* 4:75–89.
16. Trujillo JJ, Rivero M, Bonilla B (1999) On a Riemann–Liouville generalized Taylor's formula, *J. Math Anal. Appl.* (1)(231):255–265.

17. Samko SG, Kilbas AA, Marichev OI (1993) *Fractional Integrals and Derivatives. Theory and Applications*. Gordon and Breach Science, Switzerland.
18. Schneider WR, Wyss W (1989) Fractional diffusion and wave equations, J. Math. Phys. 30:134–144.
19. Sokolov IM, Klafter J, Blumen A (2002) Fractional Kinetics, Physic. Today 11(55):48–54.

RIESZ POTENTIALS AS CENTRED DERIVATIVES

Manuel Duarte Ortigueira

UNINOVA and Department of Electrical Engineering of Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa¹

Campus da FCT da UNL, Quinta da Torre

2825 – 114 Monte da Caparica, Portugal;

Tel: +351 21 2948520, Fax: +351 21 2957786, E-mail: mdo@dee.fct.unl.pt

Abstract

Generalised fractional centred differences and derivatives are studied in this chapter. These generalise to real orders the existing ones valid for even and odd positive integer orders. For each one, suitable integral formulations are presented. The limit computation inside the integrals leads to generalisations of the Cauchy derivative. Their computations using a special path lead to the well known Riesz potentials. A study for coherence is done by applying the definitions to functions with Fourier transform. The existence of inverse Riesz potentials is also studied.

Keywords

Fractional centred difference, fractional centred derivative, Grünwald–Letnikov derivative, generalised Cauchy derivative.

1 Introduction

In previous works [1, 2, 3], we proposed a new approach to coherent fractional derivatives using as starting point the Diaz and Osler integral formulation for the fractional differences [4]. The framework we proposed was based on the following steps:

¹ Also with INESC-ID, R. Alves Redol, 9, 2º, Lisbon.

1. Use as starting point the Grünwald-Letnikov forward and backward derivatives
2. With integral formulations for the fractional differences and using the asymptotic properties of the Gamma function obtain the generalised Cauchy derivative
3. The computation of the integral defining the generalised Cauchy derivative is done with the Hankel path to obtain regularised fractional derivatives
4. The application of these regularised derivatives to functions with Laplace transform, leads to the Liouville fractional derivative [3]

Here we present a similar procedure for centred fractional derivatives [5, 6] We proceed according to the following steps:

1. Introduce the general framework for the centred differences, considering two cases that we will call type 1 and type 2. These are generalisations of the usual centred differences for even and odd positive orders respectively.
2. For those differences, integral representations will be proposed.
3. These differences lead to centred derivatives that are very similar to the usual Grünwald-Letnikov derivatives.
4. From the integral representations we obtain generalisations of the Cauchy derivative formula by using the properties of the Gamma function.
5. If the integration is performed over a two straight lines path that “closes” at infinite those integrals lead to the Riesz potentials.

A very important feature consequence of the theory we are going to present lies in the summation formulae for computing the Riesz potentials that are suitable for their numerical computation.

To test the coherence of the proposed definitions we apply them to the complex exponential. The results show that they are suitable for functions with Fourier transform. The formulation agrees also with the Okikiolu studies [7].

We must refer that we will not address the existence problem. We are mainly interested in obtaining a generalisation of a well-known formulation.

The paper outline is as follows. In section 2 we present the type 1 and type 2 centred differences and their integral representations. Centred derivative definitions similar to Grünwald-Letnikov ones are presented in section 3 and their integral representations obtained generalising the Riesz potentials. In section 4 we apply the definitions to the complex exponential to test the coherence of the definitions. At last we present some conclusions.

2 Centred Differences and Derivatives

2.1 Integer order centred differences

Let $f(t)$ be a complex variable function and $h \in \mathbb{C}$ and introduce Δ_c as finite “centred” difference defined by:

$$\Delta_c f(t) = f(t+h/2) - f(t-h/2) \quad (1)$$

By repeated application of this difference, we have:

$$\Delta_e^N f(t) = \sum_{k=-N/2}^{N/2} (-1)^{N/2-k} \frac{N!}{(N/2+k)! (N/2-k)!} f(t - kh) \quad (2)$$

when N is even, and

$$\Delta_o^N f(t) = \sum_{k=-N/2}^{N/2} (-1)^{N/2-k} \frac{N!}{(N/2+k)! (N/2-k)!} f(t - kh) \quad (3)$$

if N is odd. The symbol $\sum_{k=-N/2}^{N/2}$ used in the above formula means that the summation is done over half-integer values.

Using the gamma function, we can rewrite the above formulae in the format stated as follows:

Definition 1. Let N be a positive even integer. We define a centred difference by:

$$\Delta_e^N f(t) = (-1)^{N/2} \sum_{k=-N/2}^{N/2} (-1)^k \frac{\Gamma(N+1)}{\Gamma(N/2+k+1) \Gamma(N/2-k+1)} f(t - kh) \quad (4)$$

Definition 2. Let N be a positive odd integer. We define a centred difference by:

$$\Delta_o^N f(t) = (-1)^{(N+1)/2} \sum_{k=-(N-1)/2}^{(N+1)/2} (-1)^k \frac{\Gamma(N+1)}{\Gamma((N+1)/2-k+1) \Gamma((N-1)/2+k+1)} f(t - kh + h/2) \quad (5)$$

with these definitions we are able to define the corresponding derivatives.

Definition 3. Let N be a positive even integer. We define a centred derivative by:

$$D_e^\alpha f(t) = \lim_{h \rightarrow 0} \frac{(-1)^{N/2}}{h^N} \sum_{k=-N/2}^{N/2} (-1)^k \frac{\Gamma(N+1)}{\Gamma(N/2+k+1) \Gamma(N/2-k+1)} f(t - kh) \quad (6)$$

Definition 4. Let N be a positive odd integer. We define a centred derivative by:

$$D_o^\alpha f(t) = \lim_{h \rightarrow 0} \frac{(-1)^{(N+1)/2}}{h^N} \sum_{k=-(N-1)/2}^{(N+1)/2} \frac{(-1)^k \Gamma(N+1)}{\Gamma((N+1)/2-k+1) \Gamma((N-1)/2+k+1)} f(t - kh + h/2) \quad (7)$$

Both derivatives (6) and (7) coincide with the usual derivative.

2.2 Integral representations for the integer order centred differences

The result stated in (4) can be interpreted in terms of the residue theorem leading to the following theorem.

Theorem 1. Assume that $f(z)$ is analytic inside and on a closed integration path that includes the points $t = z - kh$, $h \in \mathbb{C}$, with $k = -N/2, -N/2+1, \dots, -1, 0, 1, \dots, N/2-1, N/2$. Then

$$\Delta_e^N f(z) = \frac{(-1)^{N/2} N!}{2\pi i h} \int_{C_c} f(z+w) \frac{\Gamma\left(\frac{-w}{h}+1\right)}{\Gamma\left(\frac{-w}{h}+\frac{N}{2}+1\right)} \frac{\Gamma\left(\frac{w}{h}\right)}{\Gamma\left(\frac{w}{h}+\frac{N}{2}+1\right)} dw \quad (8)$$

Proof. Equation (4) can be considered as $\frac{1}{2\pi i} \sum$ residues in the computation of the integral of a function with poles at $t = z - kh$. We can make a translation and consider poles at kh . As it can be seen by direct verification, we have [2, 5]:

$$\sum_{k=-N/2}^{N/2} \frac{N! (-1)^{N/2-k}}{(N/2+k)! (N/2-k)!} f(t - kh) = \frac{N!}{2\pi i h} \int_{C_c} \frac{f(z+w)}{\prod_{k=0}^{N/2} \left(\frac{w}{h} - k\right) \prod_{k=1}^{N/2} \left(\frac{w}{h} + k\right)} dw \quad (9)$$

Introducing the Pochhammer symbol, we can rewrite the above formula as:

$$\Delta_e^N f(z) = \frac{(-1)^{N/2} N!}{2\pi i h} \int_{C_c} \frac{f(z+w) \left(\frac{-w}{h}\right)}{\left(\frac{w}{h}\right)_{N/2+1} \left(\frac{-w}{h}\right)_{N/2+1}} dw \quad (10)$$

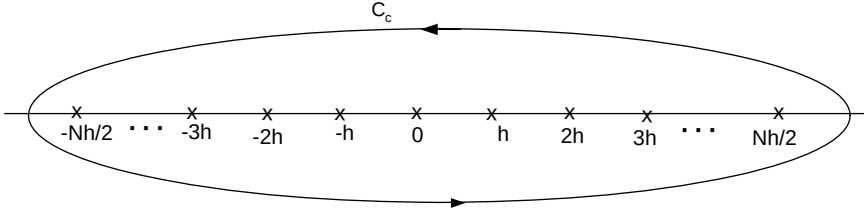


Fig. 1. Integration path and poles for the integral representation of integer even-order differences.

Attending to the relation between the Pochhammer symbol and the gamma function:

$$\Gamma(z+n) = (z)_n \Gamma(z) \quad (11)$$

we can write (8). $\square$

It is easy to test the coherency of (8) relatively to (4), by noting that the gamma function $\Gamma(z)$ has poles at the negative integers ($z = -n, n \in \mathbb{Z}^+$). The corresponding residues are equal to $(-1)^n/n!$ [9]. Both the gamma functions have infinite poles, but outside the integration path they cancel out and the integrand is analytic.

Theorem 2. In the conditions similar to the above theorem, we have²:

$$\Delta_0^N f(z) = \frac{(-1)^{(N+1)/2} N!}{2\pi i h} \int_{C_c} f(z+w) \frac{\Gamma\left(-\frac{w}{h} + \frac{1}{2}\right)}{\Gamma\left(-\frac{w}{h} + \frac{N}{2} + 1\right)} \frac{\Gamma\left(\frac{w}{h} + \frac{1}{2}\right)}{\Gamma\left(\frac{w}{h} + \frac{N}{2} + 1\right)} dw \quad (12)$$

To prove this, we proceed as above. By direct verification, we have

$$\sum_{k=-N/2}^{N/2} (-1)^{N/2-k} \binom{N}{N/2-k} f(z - kh) = \frac{N!}{2\pi i h} \int_{C_c} \frac{f(z+w)}{\prod_{k=0}^{(N-1)/2} \left(\frac{w}{h} - k - \frac{1}{2}\right) \prod_{k=0}^{(N-1)/2} \left(\frac{w}{h} + k + \frac{1}{2}\right)} dw \quad (13)$$

and

$$\Delta_0^N f(z) = \frac{(-1)^{(N+1)/2} N!}{2\pi i h} \int_{C_c} \frac{f(z+w)}{\left(\frac{w}{h} + \frac{1}{2}\right)_{(N+1)/2} \left(-\frac{w}{h} + \frac{1}{2}\right)_{(N+1)/2}} dw \quad (14)$$

that leads immediately to (12)

² Figure 2 shows the integration path and corresponding poles.

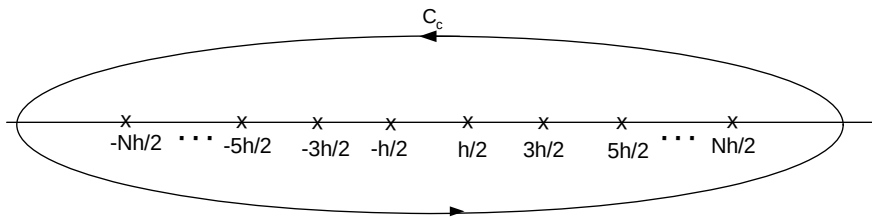


Fig. 2. Integration path and poles for the integral representation of integer odd-order differences.

2.3 The Cauchy derivative

To obtain derivatives from (8) and (12) we have to perform the computation of the limit as h goes to zero. However to obtain the integral formulae for the derivatives we must permute there the limit and integral operations. With this permutation we must compute the limit of two quotients of gamma functions.

As it is well known, the quotient of two gamma functions $\frac{\Gamma(s+a)}{\Gamma(s+b)}$ has an interesting expansion [9, 10]:

$$\frac{\Gamma(s+a)}{\Gamma(s+b)} = s^{a-b} \left[1 + \sum_1^N c_k s^{-k} + O(s^{-N-1}) \right] \quad (15)$$

as $|s| \rightarrow \infty$, uniformly in every sector that excludes the negative real half-axis. The coefficients in the series can be expressed in terms of Bernoulli polynomials, but their knowledge is not important here.

When h is very small,

$$\Delta_e^N f(z) = \frac{(-1)^{N/2} N!}{2\pi i h} \int_{C_c} f(z+w) \frac{1}{\left(\frac{w}{h}\right)^{N/2+1} \left(\frac{-w}{h}\right)^{N/2}} dw + g(h) \quad (16)$$

The $g(h)$ term is proportional to h^{N+2} . Dividing by h^N :

$$\frac{\Delta_e^N f(z)}{h^N} = \frac{N!}{2\pi i} \int_{C_c} f(z+w) \frac{1}{w^{N+1}} dw + \frac{g(h)}{h^N} \quad (17)$$

and allowing $h \rightarrow 0$, we obtain

$$f^{(N)}(z) = \frac{N!}{2\pi i} \int_{C_c} f(z+w) \frac{1}{w^{N+1}} dw \quad (18)$$

that is the Cauchy derivative. Similarly,

$$\Delta_o^N f(z) = \frac{(-1)^{(N+1)/2} N!}{2\pi i h} \int_{C_c} f(z+w) \frac{1}{\left(\frac{w}{h}\right)^{(N+1)/2} \left(\frac{-w}{h}\right)^{(N+1)/2}} dw + g(h) \quad (19)$$

and

$$\frac{\Delta_o^N f(z)}{h^N} = \frac{N!}{2\pi i} \int_{C_c} f(z+w) \frac{1}{w^{N+1}} dw + \frac{g(h)}{h^N} \quad (20)$$

leading to the Cauchy derivative again.

With these results we can state:

Theorem 3. In the conditions of theorems 1 and 2 we have:

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{(-1)^{N/2}}{h^N} \sum_{k=-N/2}^{N/2} \frac{(-1)^k \Gamma(N+1)}{\Gamma(N/2+k+1) \Gamma(N/2-k+1)} f(t-kh) \\ = \frac{N!}{2\pi i} \int_{C_c} f(z+w) \frac{1}{w^{N+1}} dw \end{aligned} \quad (21)$$

if N is a positive even integer and

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{(-1)^{(N+1)/2}}{h^N} \sum_{k=-(N-1)/2}^{(N+1)/2} (-1)^k \frac{\Gamma(N+1)}{\Gamma((N+1)/2-k+1) \Gamma((N-1)/2+k+1)} f(t-kh+h/2) \\ = \frac{N!}{2\pi i} \int_{C_c} f(z+w) \frac{1}{w^{N+1}} dw \end{aligned} \quad (22)$$

if N is a positive odd integer. Relations (21) and (22) show that both derivative definitions lead to the usual Cauchy formula.

3 Fractional Centred Differences

3.1 Type 1 and type 2 differences

Here we follow the steps of the previous section and introduce two types of fractional centred differences. Let $\alpha > -1$, $h \in \mathbb{R}^+$ and $f(t)$ a complex variable function.

Definition 5. We define a type 1 fractional difference by:

$$\Delta_{c_1}^\alpha f(t) = \sum_{k=-\infty}^{+\infty} \frac{(-1)^k \Gamma(\alpha+1)}{\Gamma(\alpha/2-k+1) \Gamma(\alpha/2+k+1)} f(t-kh) \quad (23)$$

Let $\alpha = 2M$, $M \in \mathbb{Z}^+$. We obtain:

$$\Delta_{c_1}^{2M} f(t) = \sum_{k=-M}^{+M} \frac{(-1)^k (2M)!}{(M-k)! (M+k)!} f(t-kh) \quad (24)$$

that aside a factor $(-1)^M$ it is the current $2M$ order centred difference.

Definition 6. We define a type 2 fractional difference by:

$$\Delta_{c_2}^\alpha f(t) = \sum_{k=-\infty}^{+\infty} \frac{(-1)^k \Gamma(\alpha+1)}{\Gamma[(\alpha+1)/2-k+1] \Gamma[(\alpha-1)/2+k+1]} f(t-kh+h/2) \quad (25)$$

Similarly, if α is odd ($\alpha = 2M+1$), it is, aside the factor $(-1)^{M+1}$, equal to current centred difference. In fact, we have:

$$\Delta_{c_2}^{2M+1} f(t) = \sum_{k=-M}^{M+1} \frac{(-1)^k (2M+1)!}{(M+1-k)! (M+k)!} f(t-kh+h/2) \quad (26)$$

In particular, with $M = 0$, we obtain: $\Delta_{c_2}^1 f(t) = f(t+h/2) - f(t-h/2)$.

With the following relation [8]³:

$$\sum_{k=-\infty}^{+\infty} \frac{1}{\Gamma(a-k+1) \Gamma(b-k+1) \Gamma(c+k+1) \Gamma(d+k+1)} = \frac{\Gamma(a+b+c+d+1)}{\Gamma(a+c+1) \Gamma(b+c+1) \Gamma(a+d+1) \Gamma(b+d+1)} \quad (27)$$

valid for $a+b+c+d > -1$, it is not very hard to show that:

$$\Delta_{c_1}^\beta \left\{ \Delta_{c_1}^\alpha f(t) \right\} = \Delta_{c_1}^{\alpha+\beta} f(t) \quad (28)$$

and

³ Page 123 of [8].

$$\Delta_{c_2}^\beta \left\{ \Delta_{c_2}^\alpha f(t) \right\} = - \Delta_{c_1}^{\alpha+\beta} f(t) \quad (29)$$

while

$$\Delta_{c_2}^\beta \left\{ \Delta_{c_1}^\alpha f(t) \right\} = \Delta_{c_2}^{\alpha+\beta} f(t) \quad (30)$$

provided that $\alpha+\beta > -1$. In particular, $\alpha+\beta = 0$, and the relations (28) and (29) show that when $|\alpha| < 1$ and $|\beta| < 1$ the inverse differences exist and can be obtained by using formulae (23) and (25).

We must remark that the zero-order difference is the identity operator and is obtained from (23). It is interesting to remark also that the combination of equal types of differences gives a type 1 difference, while the combination of different types gives a type 2 difference.

When comparing these differences with (4) and (5) we see that a power of -1 was removed. Later we will understand why.

3.2 Integral representations for the fractional centred differences

Let us assume that $f(z)$ is analytic in a region of the complex plane that includes the real axis. Assume that α is not an integer.

To obtain the integral representations for the previous differences we follow here the procedure used above [1, 2, 5, 6]. Essentially, it is a mere substitution of α for N in (8) and (12). In the first case, this leads easily to

$$\Delta_{c_1}^\alpha f(t) = \frac{\Gamma(\alpha+1)}{2\pi i h} \int_{C_c} f(z+w) \frac{\Gamma\left(\frac{-w}{h}+1\right) \Gamma\left(\frac{w}{h}\right)}{\Gamma\left(\frac{-w}{h}+\frac{\alpha}{2}+1\right) \Gamma\left(\frac{w}{h}+\frac{\alpha}{2}+1\right)} dw \quad (31)$$

The integrand has infinite poles at every nh , with $n \in \mathbb{Z}$. The integration path must consist of infinite lines above and below the real axis closing at the infinite. The easiest situation is obtained by considering two straight lines near the real axis, one above and the other below (Fig. 3).

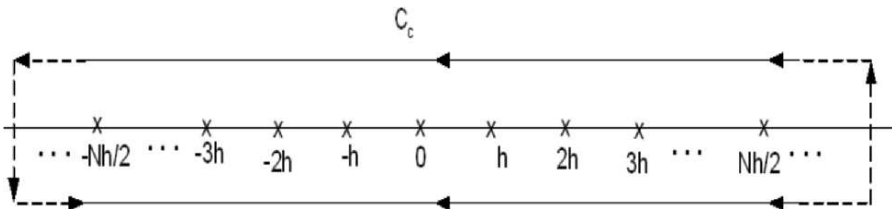


Fig. 3. Path and poles for the integral representation of type 1 differences.

Regarding to the second case, the poles are located now at the half-integer multiples of h (see Fig. 4), which leads to

$$\Delta_{c_2}^\alpha f(t) = \frac{\Gamma(\alpha+1)}{2\pi i h} \int_{C_c} f(z+w) \frac{\Gamma\left(-\frac{w}{h} + \frac{1}{2}\right) \Gamma\left(\frac{w}{h} + \frac{1}{2}\right)}{\Gamma\left(-\frac{w}{h} + \frac{\alpha}{2} + 1\right) \Gamma\left(\frac{w}{h} + \frac{\alpha}{2} + 1\right)} dw \quad (32)$$

These integral formulations will be used in the following section to obtain the integral formulae for the centred derivatives.

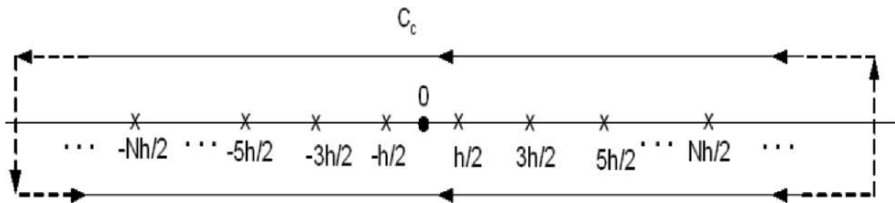


Fig. 4. Path and poles for the integral representation of type 2 differences.

4 Fractional Centred Derivatives

4.1 Definitions

To obtain fractional centred derivatives we proceed as usually [1, 2, 5, 6, 10]: divide the fractional differences by h^α ($h \in \mathbb{R}^+$) and let $h \rightarrow 0$.

Definition 7. For the first case and assuming again that $\alpha > -1$, we define the type 1 fractional centred derivative by:

$$D_{c_1}^\alpha f(t) = \lim_{h \rightarrow 0} \frac{\Gamma(\alpha+1)}{h^\alpha} \sum_{k=-\infty}^{+\infty} \frac{(-1)^k}{\Gamma(\alpha/2 - k + 1) \Gamma(\alpha/2 + k + 1)} f(t - kh) \quad (33)$$

Definition 8. For the second case, we define the type 2 fractional centred derivative given by

$$D_{c_2}^\alpha f(t) = \lim_{h \rightarrow 0} \frac{\Gamma(\alpha+1)}{h^\alpha} \sum_{k=-\infty}^{+\infty} \frac{(-1)^k}{\Gamma[(\alpha+1)/2 - k + 1] \Gamma[(\alpha-1)/2 + k + 1]} f(t - kh + h/2) \quad (34)$$

Formulae (33) and (34) generalise the positive integer order centred derivatives to the fractional case, although there should be an extra factor $(-1)^{\alpha/2}$ in the first case

and $(-1)^{(\alpha+1)/2}$ in the second case that we removed. It is a simple task to obtain the derivative analogues to (28), (29), and (30):

$$D_{c_1}^\beta \left\{ D_{c_1}^\alpha f(t) \right\} = D_{c_1}^{\alpha+\beta} f(t) \quad (35)$$

and

$$D_{c_2}^\beta \left\{ D_{c_2}^\alpha f(t) \right\} = - D_{c_1}^{\alpha+\beta} f(t) \quad (36)$$

while

$$D_{c_2}^\beta \left\{ D_{c_1}^\alpha f(t) \right\} = D_{c_2}^{\alpha+\beta} f(t) \quad (37)$$

again with $\alpha+\beta > -1$.

4.2 Integral formulae

To obtain the integral formulae for the centred fractional derivatives, we follow the same procedure used in the integer order case. We start from (31) and permute the limit and integral operations. As we saw before, when h is very small

$$\frac{\Gamma(w/h+a)}{\Gamma(w/h+b)} \approx (w/h)^{a-b} [1 + h.\varphi(w/h)] \quad (38)$$

where φ is regular near the origin. Accordingly to the above statement, the branch cut line used to define a function on the right-hand side in (38) is the negative real-half axis. Similarly, we

$$\frac{\Gamma(-w/h+a)}{\Gamma(-w/h+b)} \approx (-w/h)^{a-b} [1 + h.\varphi(-w/h)] \quad (39)$$

but now, the branch cut line is the positive real axis. With these results, we obtain

Theorem 5. In the above conditions, the integral formulation for the type 1 derivative is

$$D_{c_1}^\alpha f(t) = \frac{\Gamma(\alpha+1)}{2\pi i} \int_{C_c} f(z+w) \frac{1}{(w)_l^{\alpha/2+1} (-w)_r^{\alpha/2}} dw \quad (40)$$

while for the type 2 derivative is

$$D_{c_2}^\alpha f(t) = \frac{\Gamma(\alpha+1)}{2\pi i} \int_{C_c} f(z+w) \frac{1}{(w)_l^{(\alpha+1)/2} (-w)_r^{(\alpha+1)/2}} dw \quad (41)$$

The subscripts “l” and “r” mean respectively that the power functions have the left and right-half real axis as branch cut lines. These integrals generalise the Cauchy derivative.

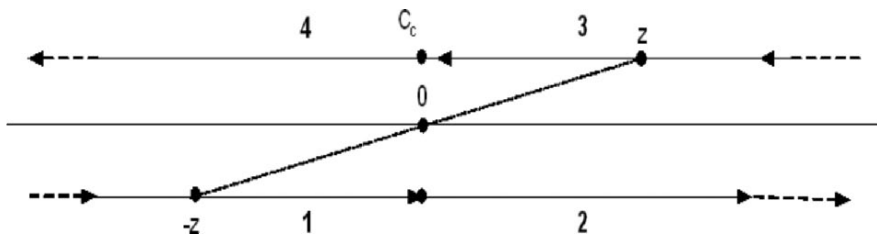


Fig. 5. Segments for the computation of the integrals (40) and (41).

Now, we are going to compute the above integrals for the special case of straight line paths. Let us assume that the distance between the horizontal straight lines in Figs. 1 and 2 is $2\varepsilon(h)$ that decreases to zero with h . In Fig. 5 we show the different segments used for the computation of the above integrals. Assuming that the straight lines are infinitely near the real axis, we obtain for the type 1 derivative:

$$\begin{aligned} \int_1 &= -\frac{\Gamma(\alpha+1)}{2\pi i} \int_0^\infty f(z-x) \frac{1}{x^{\alpha+1}} \frac{1}{e^{-i\alpha\pi/2} e^{-i\pi}} e^{i\pi} dx, \\ \int_2 &= \frac{\Gamma(\alpha+1)}{2\pi i} \int_0^\infty f(z+x) \frac{1}{x^{\alpha+1}} \frac{1}{e^{i\alpha\pi/2}} dx, \\ \int_3 &= -\frac{\Gamma(\alpha+1)e^{-i\alpha\pi/2}}{2\pi i} \int_0^\infty f(z+x) \frac{1}{x^{\alpha+1}} \frac{1}{e^{-i\alpha\pi/2}} dx \\ \int_4 &= \frac{\Gamma(\alpha+1)}{2\pi i} \int_0^\infty f(z-x) \frac{1}{x^{\alpha+1}} \frac{1}{e^{i\alpha\pi/2} e^{i\pi}} e^{i\pi} dx \end{aligned}$$

where the integers refer the straight line segment used in the computation. Joining the four integrals, we obtain:

$$D_{C_1}^\alpha f(t) = -\frac{\Gamma(\alpha+1)\sin(\alpha\pi/2)}{\pi} \int_0^\infty f(z-x) \frac{1}{x^{\alpha+1}} dx - \frac{\Gamma(\alpha+1)\sin(\alpha\pi/2)}{\pi} \int_0^\infty f(z+x) \frac{1}{x^{\alpha+1}} dx$$

or

$$D_{C_1}^\alpha f(t) = -\frac{\Gamma(\alpha+1) \sin(\alpha\pi/2)}{\pi} \int_{-\infty}^{\infty} f(z-x) \frac{1}{|x|^{\alpha+1}} dx \quad (42)$$

As α is not an odd integer and using the reflection formula of the gamma function, we obtain

$$D_{C_1}^\alpha f(t) = \frac{1}{2\Gamma(-\alpha) \cos(\alpha\pi/2)} \int_{-\infty}^{\infty} f(z-x) \frac{1}{|x|^{\alpha+1}} dx \quad (43)$$

that is the so called Riesz potential.

For the type 2 case, we compute again the integrals corresponding to the four segments to obtain:

$$\begin{aligned} \int_1 &= -\frac{\Gamma(\alpha+1)}{2\pi i} \int_0^{\infty} f(z-x) \frac{1}{x^{\alpha+1}} e^{-i(\alpha+1)\pi/2} dx \\ \int_2 &= \frac{\Gamma(\alpha+1)}{2\pi i} \int_0^{\infty} f(z+x) \frac{1}{x^{\alpha+1}} e^{i(\alpha+1)\pi/2} dx, \\ \int_3 &= -\frac{\Gamma(\alpha+1)e^{-i\alpha\pi/2}}{2\pi i} \int_0^{\infty} f(z+x) \frac{1}{x^{\alpha+1}} e^{-i(\alpha+1)\pi/2} dx \\ \int_4 &= \frac{\Gamma(\alpha+1)}{2\pi i} \int_0^{\infty} f(z-x) \frac{1}{x^{\alpha+1}} e^{i\pi} dx \end{aligned}$$

Joining the four integrals, we obtain:

$$D_{C_2}^\alpha f(t) = \frac{\Gamma(\alpha+1)\sin[(\alpha+1)\pi/2]}{\pi} \left[\int_0^{\infty} f(z-x) \frac{1}{x^{\alpha+1}} dx - \int_0^{\infty} f(z+x) \frac{1}{x^{\alpha+1}} dx \right]$$

As the last integral can be rewritten as:

$$\int_0^{\infty} f(z+x) \frac{1}{x^{\alpha+1}} dx = \int_{-\infty}^0 f(z-x) \frac{1}{(-x)^{\alpha+1}} dx$$

we obtain

$$D_{C_2}^\alpha f(t) = -\frac{1}{2\Gamma(-\alpha)\sin(\alpha\pi/2)} \int_{-\infty}^{\infty} f(z-x) \frac{\text{sgn}(x)}{|x|^{\alpha+1}} dx \quad (44)$$

that is the modified Riesz potential [10]. Both potentials (43) and (44) were studied also by Okikiolu [7]. These are essentially convolutions of a given function

with two acausal⁴ operators and are suitable for dealing with functions defined in $\mathbb{R}$ and that are not necessarily equal to zero at $\pm\infty$.

5 Coherence of the Results

5.1 Type 1 derivative

We want to test the coherence of the results by considering functions with Fourier transform. To perform this study, we only have to study the behaviour of the defined derivatives for $f(t) = e^{-i\omega t}$, $t, \omega \in \mathbb{R}$. In the following we will consider non integer orders greater than -1. We start by considering the type 1 derivative. From (23) we obtain

$$\Delta_{c_1}^\alpha e^{i\omega t} = e^{-i\omega t} \sum_{-\infty}^{+\infty} \frac{(-1)^n \Gamma(\alpha+1)}{\Gamma(\alpha/2-n+1) \Gamma(\alpha/2+n+1)} e^{i\omega n h} \quad (45)$$

where we recognize the discrete-time Fourier transform⁵ of $R_h(n)$ given by:

$$R_h(n) = \frac{(-1)^n \Gamma(\alpha+1)}{\Gamma(\alpha/2-n+1) \Gamma(\alpha/2+n+1)} \quad (46)$$

This sequence is the discrete autocorrelation of

$$h_n = \frac{(-\alpha/2)_n}{n!} u_n \quad (47)$$

where u_n is the discrete unit step Heaviside function [11]. As the discrete-time Fourier transform of h_n is:

$$H(e^{i\omega}) = \text{FT}[h_n] = (1 - e^{-i\omega h})^{\alpha/2} \quad (48)$$

the Fourier transform of $R_h(n)$ is

$$\begin{aligned} S(e^{i\omega}) &= \lim_{z \rightarrow e^{i\omega h}} (1 - z^{-1})^{\alpha/2} (1 - z)^{\alpha/2} = (1 - e^{-i\omega h})^{\alpha/2} (1 - e^{i\omega h})^{\alpha/2} \\ &= |e^{i\omega h/2} - e^{-i\omega h/2}|^\alpha = |2 \sin(\omega h/2)|^\alpha \end{aligned} \quad (49)$$

So,

$$|2 \sin(\omega h/2)|^\alpha = \sum_{-\infty}^{+\infty} \frac{(-1)^n \Gamma(\alpha+1)}{\Gamma(\alpha/2-n+1) \Gamma(\alpha/2+n+1)} e^{i\omega n h} \quad (50)$$

⁴ We name acausal the operators that are neither causal nor anti-causal.

⁵ In purely mathematical terms it is a Fourier series with $R_h(n)$ as coefficients.

We conclude that the Fourier series expansion of $|2 \sin(\omega h/2)|^\alpha$ has $R_h(n)$ as Fourier coefficients. Returning to (45) we write, then:

$$\Delta_{c_1}^\alpha e^{-i\omega t} = e^{-i\omega t} |2 \sin(\omega h/2)|^\alpha \quad (51)$$

So, there is a linear system with frequency response given by:

$$H_{\Delta 1}(\omega) = |2 \sin(\omega h/2)|^\alpha \quad (52)$$

that acts on a signal giving its centred fractional difference. Dividing by h^α ($h \in \mathbb{R}^+$) and computing the limit as $h \rightarrow 0$, (52) gives:

$$H_{D1}(\omega) = |\omega|^\alpha \quad (53)$$

that is the frequency response of the linear system that implements the type 1 centred fractional derivative. As α is not an even integer:

$$|\omega|^\alpha = \lim_{h \rightarrow 0} \frac{1}{h^\alpha} \sum_{n=-\infty}^{+\infty} \frac{(-1)^n \Gamma(\alpha+1)}{\Gamma(\alpha/2-n+1) \Gamma(\alpha/2+n+1)} e^{i\omega n h} \quad (54)$$

valid for $\alpha > -1$. The inverse Fourier transform of $|\omega|^\alpha$ is given by [7]:

$$\text{FT}^{-1}[|\omega|^\alpha] = \frac{1}{2\Gamma(-\alpha)\cos(\alpha\pi/2)} |t|^{-\alpha-1} \quad (55)$$

and we obtain the impulse response:

$$h_{D1}(t) = \frac{1}{2\Gamma(-\alpha)\cos(\alpha\pi/2)} |t|^{-\alpha-1} \quad (56)$$

leading to

$$D_{c_1}^\alpha f(t) = \frac{1}{2\Gamma(-\alpha)\cos(\alpha\pi/2)} \int_{-\infty}^{+\infty} f(\tau) |t-\tau|^{-\alpha-1} d\tau \quad (57)$$

that is coincides with (43).

5.2 Type 2 derivative

A similar procedure allows us to obtain from (25)

$$\Delta_{c_2}^\alpha e^{-i\omega t} = e^{-i\omega t} e^{-i\omega h/2} \sum_{k=-\infty}^{+\infty} \frac{(-1)^k \Gamma(\alpha+1)}{\Gamma[(\alpha+1)/2-k+1] \Gamma[(\alpha-1)/2+k+1]} e^{i\omega k h} \quad (58)$$

In order to maintain the coherence with the usual definition of discrete-time Fourier transform, we change the summation variable, obtaining

$$\Delta_{c_2}^\alpha e^{-i\omega t} = e^{-i\omega t} e^{-i\omega h/2} \sum_{k=-\infty}^{+\infty} \frac{(-1)^k \Gamma(\alpha+1)}{\Gamma[(\alpha+1)/2+k+1] \Gamma[(\alpha-1)/2-k+1]} e^{-i\omega k h} \quad (59)$$

Now, the coefficients of the above Fourier series are the cross-correlation, $R_{bc}(k)$, between h_n and g_n given by

$$h_n = \frac{(-a)_n}{n!} u_n, \quad g_n = \frac{(-b)_n}{n!} u_n \quad (60)$$

with $a = (\alpha+1)/2$ and $b = (\alpha-1)/2$. Let $S_{bc}(e^{i\omega})$ be the discrete-time Fourier transform of the cross-correlation, $R_{bc}(k)$. We conclude easily that $S_{bc}(e^{i\omega})$ is given by:

$$\begin{aligned} S_{bc}(e^{i\omega}) &= \lim_{z \rightarrow e^{i\omega h}} (1-z^{-1})^{(\alpha+1)/2} (1-z)^{(\alpha-1)/2} = (1-e^{-i\omega h})^{(\alpha+1)/2} (1-e^{i\omega h})^{(\alpha-1)/2} \\ &= (1-e^{-i\omega h})^{(\alpha+1)/2} (1-e^{i\omega h})^{(\alpha+1)/2} (1-e^{i\omega h})^{-1} \end{aligned} \quad (61)$$

We write, then:

$$\begin{aligned} \Delta_{c_2}^\alpha e^{-i\omega t} &= e^{i\omega t} (1-e^{-i\omega h})^{(\alpha+1)/2} (1-e^{i\omega h})^{(\alpha+1)/2} (1-e^{i\omega h})^{-1} e^{i\omega h/2} \\ &= e^{i\omega t} |2 \sin(\omega h/2)|^{\alpha+1} [2i \sin(\omega h/2)]^{-1} \end{aligned} \quad (62)$$

So, there is a linear system with frequency response given by:

$$H_{\Delta_2}(\omega) = |2 \sin(\omega h/2)|^{\alpha+1} [2i \sin(\omega h/2)]^{-1} \quad (63)$$

that acts on a signal giving its fractional centred difference. We can write also

$$\begin{aligned} &|2 \sin(\omega h/2)|^{\alpha+1} [2i \sin(\omega h/2)]^{-1} = \\ &\sum_{k=-\infty}^{+\infty} \frac{(-1)^k \Gamma(\alpha+1)}{\Gamma[(\alpha+1)/2+k+1] \Gamma[(\alpha-1)/2-k+1]} e^{-i\omega k h} \end{aligned} \quad (64)$$

Dividing the left-hand side in (64) by h^α ($h \in \mathbb{R}^+$) and computing the limit as $h \rightarrow 0$, it gives:

$$H_{D_2}(\omega) = -i |\omega|^\alpha \text{sgn}(\omega) \quad (65)$$

As

$$\frac{d|\omega|^{\alpha+1}}{d\omega} = (\alpha+1) |\omega|^\alpha \text{sgn}(\omega) \quad (66)$$

we obtain from (55) using a well-known property of the Fourier transform:

$$h_{D_2}(t) = \frac{-\text{sgn}(t)}{(\alpha+1)2\Gamma(-\alpha-1)\cos[(\alpha+1)\pi/2]} |t|^{-\alpha-1} \quad (67)$$

or, using the properties of the gamma function

$$h_{D_2}(t) = -\frac{\text{sgn}(t)}{2\Gamma(-\alpha)\sin(\alpha\pi/2)} |t|^{-\alpha-1} \quad (68)$$

and as previously:

$$D_{c_1}^\alpha f(t) = -\frac{1}{2\Gamma(-\alpha)\sin(\alpha\pi/2)} \int_{-\infty}^{+\infty} f(\tau) |t-\tau|^{-\alpha-1} \operatorname{sgn}(t-\tau) d\tau \quad (69)$$

5.3 On the existence of a inverse Riesz potential

In current literature [7,10], the Riesz potentials are only defined for negative orders verifying $-1 < \alpha < 0$. However, our formulation is valid for every $\alpha > -1$. This means that we can define those potentials even for positive orders. However, we cannot guaranty that there is always an inverse for a given potential. The theory presented in section 4.1 allows us to state that:

- The inverse of a given potential, when existing, is of the same type: the inverse of the type k ($k = 1, 2$) potential is a type- k potential.
- The inverse of a given potential exists iff its order α verifies $|\alpha| < 1$.
- The order of the inverse of an α order potential is a $-\alpha$ order potential.
- The inverse can be computed both by (33) [respectively (34)] and by (43) [respectively (44)].

This is in contradiction with the results stated in [10], about this subject and will have implications in the solution of differential equations involving centred derivatives.

5.4 An “analytic” derivative

An interesting result can be obtained by combining (53) with (65) to give a complex function

$$H_D(\omega) = H_{D1}(\omega) + iH_{D2}(\omega) \quad (70)$$

We obtain a function that is null for $\omega < 0$. This means that the operator defined by (44) is the Hilbert transform of that defined in (43). The inverse Fourier transform of (70) is an “analytic signal” and the corresponding “analytic” derivative is given by the convolution of the function at hand with the operator:

$$H_D(t) = \frac{|t|^{-\alpha-1}}{2\Gamma(-\alpha)\cos(\alpha\pi/2)} - i \frac{|t|^{-\alpha-1} \operatorname{sgn}(t)}{2\Gamma(-\alpha)\sin(\alpha\pi/2)} \quad (71)$$

This leads to a convolution integral formally similar to the Riesz–Feller potentials [10].

We can give this formula another aspect by noting that

$$\frac{1}{2\Gamma(-\alpha)\cos(\alpha\pi/2)} = -\frac{\Gamma(\alpha+1).\sin(\alpha\pi)}{2\pi\cos(\alpha\pi/2)} = -\frac{\Gamma(\alpha+1)\sin(\alpha\pi/2)}{\pi} \quad (72)$$

and

$$-\frac{1}{2\Gamma(-\alpha)\sin(\alpha\pi/2)} = \frac{\Gamma(\alpha+1).\sin(\alpha\pi)}{2\pi\sin(\alpha\pi/2)} = \frac{\Gamma(\alpha+1)\cos(\alpha\pi/2)}{\pi} \quad (73)$$

We obtain easily:

$$H_D(t) = -\frac{\Gamma(\alpha+1)}{\pi} \left[|t|^{-\alpha-1} \sin(\alpha\pi/2) - i|t|^{-\alpha-1} \operatorname{sgn}(t) \cos(\alpha\pi/2) \right] \quad (74)$$

that can be rewritten as

$$H_D(t) = \frac{i\Gamma(\alpha+1)|t|^{-\alpha-1}}{\pi} \operatorname{sgn}(t) e^{i\alpha\pi/2 \operatorname{sgn}(t)} \quad (75)$$

This impulse response leads to the following potential:

$$D_D^\alpha f(t) = \frac{\Gamma(\alpha+1)}{\pi} \int_{-\infty}^{+\infty} f(t-\tau) |\tau|^{-\alpha-1} \operatorname{sgn}(\tau) e^{i\alpha\pi/2 \operatorname{sgn}(\tau)} d\tau \quad (76)$$

Of course, the Fourier transform of this potential is zero for $\omega < 0$. Similarly, the function

$$H_D(\omega) = H_{D1}(\omega) - iH_{D2}(\omega) \quad (77)$$

is zero for $\omega > 0$. Its inverse Fourier transform is easily obtained, proceeding as above.

5.5 The integer order cases

It is interesting to use the centred type 1 derivative with $\alpha = 2M + 1$ and the type 2 with $\alpha = 2M$.

For the first, $\alpha/2$ is not integer and we can use formulae (49) to (54). However, they are difficult to manipulate. We found better to use (55), but we must avoid the product $\Gamma(-\alpha).\cos(\alpha\pi/2)$, because the first factor is ∞ and the second is zero. To solve the problem, we use (72) to obtain a factor equal to $-\frac{(2M+1)!(-1)^M}{\pi}$. We obtain finally

$$\text{FT}^{-1}[|\omega|^{2M+1}] = -\frac{(2M+1)!(-1)^M}{\pi} |t|^{-2M-2} \quad (78)$$

and the corresponding impulse response:

$$h_{D1}(t) = - \frac{(2M+1)! (-1)^M}{\pi} |t|^{-2M-2} \quad (79)$$

Concerning the second case, $\alpha = 2M$, we use formula (65). As above, we have the product $\Gamma(-\alpha) \cdot \sin(\alpha\pi/2)$ that is again a $\infty \cdot 0$ situation. Using (73) we obtain a factor $\frac{(2M)! (-1)^M}{\pi}$. We obtain then:

$$FT^{-1}[|\omega|^{2M} \text{sgn}(\omega)] = \frac{\text{sgn}(t) (2M)! (-1)^M}{\pi} |t|^{-2M-1} \quad (80)$$

and

$$h_{D2}(t) = \frac{\text{sgn}(t) (2M)! (-1)^M}{\pi} |t|^{-2M-1} \quad (81)$$

As we can see, the formulae (78) and (80) allow us to generalise the Riesz potentials for integer orders. However, they do not have inverse.

6 Conclusions

We introduced a general framework for defining the fractional centred differences and consider two cases that are generalisations of the usual even and odd integer orders centred differences. These new differences led to centred derivatives similar to the usual Grwald–Letnikov ones.

For those differences, we proposed integral representations from where we obtained the derivative integrals, similar to the ordinary Cauchy formula, by limit computations inside the integrals and using the asymptotic property of the quotient of two gamma functions. We obtained an integrand that is a multivalued expression needing two branch cut lines to define a function.

For the computation of those integrals we used a special path consisting of two straight lines lying immediately above and below the real axis. These computations led to generalisations of the well known Riesz potentials.

The most interesting feature of the presented theory lies in the equality between two different formulations for the Riesz potentials. As one of them is based on a summation formula it will be suitable for numerical computations.

To test the coherence of the proposed definitions we applied them to the complex exponential. The results show that they are suitable for functions with Fourier transform, meaning that every function with Fourier transform has a centred derivative.

References

1. Ortigueira, MD, (2005) Fractional Differences Integral Representation and its Use to Define Fractional Derivatives, In: Proceedings of the ENOC-2005, Fifth EUROMECH Nonlinear Dynamics Conference, Eindhoven University of Technology, The Netherlands, August, 7–12.
2. Ortigueira MD, Coito F (2004) From Differences to Differintegrations, *Fract. Calc. Appl. Anal.* 7(4).
3. Ortigueira MD (2006) A coherent approach to non integer order derivatives, *Signal Processing*, special issue on Fractional Calculus and Applications.
4. Diaz IB, Osler TI (1974) Differences of fractional order, *Math. Comput.* 28 (125).
5. Ortigueira MD (2006) Fractional Centred Differences and Derivatives, to be presented at the IFAC FDA Workshop to be held at Porto, Portugal, 19–21 July, 2006.
6. Ortigueira MD (2005) Riesz potentials via centred derivatives submitted for publication in the *Int. J. Math. Math. Sci.* December 2005.
7. Okikiolu GO (1966) Fourier Transforms of the operator $H\alpha$, In: Proceedings of Cambridge Philosophy Society 62, 73–78.
8. Andrews GE, Askey R, Roy R (1999) *Special Functions*, Cambridge University Press, Cambridge.
9. Henrici P (1974) *Applied and Computational Complex Analysis*, Vol. 1. Wiley, pp. 270–271.
10. Samko SG, Kilbas AA, Marichev OI (1987) *Fractional Integrals and Derivatives – Theory and Applications*. Gordon and Breach Science, New York.
11. Ortigueira MD (2000) Introduction to Fractional Signal Processing. Part 2: Discrete-Time Systems, In: *IEE Proceedings on Vision, Image and Signal Processing*, No.1, February 2000, pp. 71–78.

Part 2

Classical Mechanics and Particle Physics

ON FRACTIONAL VARIATIONAL PRINCIPLES

Dumitru Baleanu¹ and Sami I. Muslih²

¹ Department of Mathematics and Computer Sciences, Faculty of Arts and Sciences, Çankaya University, 06530 Ankara, Turkey; E-mail: dumitru@cankaya.edu.tr; Institute of Space Sciences, P.O. Box MG-36, R 76900, Magurele-Bucharest, Romania; E-mail: baleanu@venus.nipne.ro

² Department of Physics, Al-Azhar University, Gaza, Palestine; E-mail: smuslih@ictp.it

Abstract

The paper provides the fractional Lagrangian and Hamiltonian formulations of mechanical and field systems. The fractional treatment of constrained system is investigated together with the fractional path integral analysis. Fractional *Schrödinger* and Dirac fields are analyzed in details.

Keywords

Fractional calculus, fractional variational principles, fractional Lagrangian and Hamiltonian, fractional Schrödinger field, fractional Dirac field.

1 Introduction

Variational principles play an important role in physics, mathematics, and engineering science because they bring together a variety of fields, lead to novel results and represent a powerful tool of calculation.

It has been observed that in physical sciences the methodology has changed from complete confidence on the tools of linear, analytic, quantitative mathematical physics towards a combination of nonlinear, numerical, and qualitative techniques.

Derivatives and integrals of fractional order [1–5] have found many applications in recent studies in various fields [6–18]. Several important results in numerical analysis [19], various areas of physics [5], and engineering have been reported. For example, in fields as viscoelasticity [20–22], electrochemistry, diffusion processes [23], the analysis is formulated with respect to fractional-order derivatives and integrals. The fractional derivative accurately describes natural phenomena that occur in such common engineering problems as heat transfer, electrode/electrolyte behavior, and subthreshold nerve propagation [24]. Also, the fractional calculus found many

many applications in recent studies of scaling phenomena [25] as well as in classical mechanics [26].

Although many laws of nature can be obtained using certain functionals and the theory of calculus of variations, not all laws can be obtained by using this procedure. For example, almost all systems contain internal damping, yet traditional energy-based approach cannot be used to obtain equations describing the behavior of a nonconservative system [27]. For these reasons during the last decade huge efforts were dedicated to apply the fractional calculus to the variational problems [28–31]. Riewe has applied the fractional calculus to obtain a formalism which can be used for describing both conservative and nonconservative systems [28–29]. By using this approach, one can obtain the Lagrangian and the Hamiltonian equations of motion for the nonconservative systems.

The fractional variational problem of Lagrange was studied in [32]. A new application of a fractal concept to quantum physics has been reported in [33–34]. The issue of having nonconservative equations from the use of a variational principle was investigated recently in [35]. The simple solution of the fractional Dirac equation of order $2/3$ was investigated recently in [36]. Even more recently, the fractional calculus technique was applied to the constrained systems [37–38] and the path integral quantization of fractional mechanical systems with constraints was analyzed in [39].

The aim of this paper is to present some of the latest developments in the field of fractional variational principles. The fractional Euler–Lagrange equations, the fractional Hamiltonian equations, and the fractional path integral formulation are discussed for both discrete systems and field theory.

The paper is organized as follows:

Euler–Lagrange equations for discrete systems are briefly reviewed in section 2. In section 3 the fractional Euler–Lagrange equations of field systems are presented and the fractional Schrödinger equation is obtained from a fractional variational principle. Section 4 is dedicated to the fractional Hamiltonian analysis. Section 5 is dedicated to the fractional path integral of Dirac field and nonrelativistic particle interacting with external electromagnetism field. Finally, section 6 is devoted to our conclusions.

2 Fractional Euler–Lagrange Equations

2.1 Riemann–Liouville fractional derivatives

One of the main advantages of using Riemann–Liouville fractional derivatives within the variational principles is the possibility of defining the integration by parts as well as the fractional Euler–Lagrange equations become the classical ones when α is an integer.

In the following some basic definitions and properties of Riemann–Liouville fractional derivatives are presented.

The left Riemann–Liouville fractional derivative is defined as follows

$${}_a\mathbf{D}_t^\alpha f(t) = \frac{1}{\Gamma(n-\alpha)} \left(\frac{d}{dt} \right)^n \int_a^t (t-\tau)^{n-\alpha-1} f(\tau) d\tau, \quad (1)$$

and the form of the right Riemann–Liouville fractional derivative is given below

$${}_t\mathbf{D}_b^\alpha f(t) = \frac{1}{\Gamma(n-\alpha)} \left(-\frac{d}{dt} \right)^n \int_t^b (\tau-t)^{n-\alpha-1} f(\tau) d\tau. \quad (2)$$

Here the order α fulfills $n-1 \leq \alpha < n$ and Γ represents the Euler's gamma function. If α becomes an integer, these derivatives become the usual derivatives

$${}_a\mathbf{D}_t^\alpha f(t) = \left(\frac{d}{dt} \right)^\alpha, \quad {}_t\mathbf{D}_b^\alpha f(t) = \left(-\frac{d}{dt} \right)^\alpha, \quad \alpha = 1, 2, \dots \quad (3)$$

Let us consider a function f depending on n variables, $x_1, x_2, \dots, x_n$. A partial left Riemann–Liouville fractional derivative of order α_k , $0 < \alpha_k < 1$, in the k -th variable is defined as [2]

$$(\mathbf{D}_{a_k+}^{\alpha_k} f)(x) = \frac{1}{\Gamma(1-\alpha)} \frac{\partial}{\partial x_k} \int_{a_k}^{x_k} \frac{f(x_1, \dots, x_{k-1}, u, x_{k+1}, \dots, x_n)}{(x_k - u)^{\alpha_k}} du \quad (4)$$

and a partial right Riemann–Liouville fractional derivative of order α_k has the form

$$(\mathbf{D}_{a_k-}^{\alpha_k} f)(x) = \frac{1}{\Gamma(1-\alpha)} \frac{\partial}{\partial x_k} \int_{x_k}^{a_k} \frac{f(x_1, \dots, x_{k-1}, u, x_{k+1}, \dots, x_n)}{(-x_k + u)^{\alpha_k}} du. \quad (5)$$

If the function f is differentiable we obtain

$$\begin{aligned} (\mathbf{D}_{a_k+}^{\alpha_k} f)(x) &= \frac{1}{\Gamma(1-\alpha_k)} \left[\frac{f(x_1, \dots, x_{k-1}, a_k, x_{k+1}, \dots, x_n)}{(x_k - a_k)^{\alpha_k}} \right] \\ &\quad + \int_{a_k}^{x_k} \frac{\frac{\partial f}{\partial u}(x_1, \dots, x_{k-1}, u, x_{k+1}, \dots, x_n)}{(x_k - u)^{\alpha_k}} du. \end{aligned} \quad (6)$$

2.2 Fractional Euler–Lagrange equations for mechanical systems

Many applications of fractional calculus amount to replacing the time derivative in an evolution equation with a derivative of fractional order.

For a given classical Lagrangian the first issue is to construct its fractional generalization. The fractional Lagrangian is not unique because there are several possibilities to replace the time derivative with fractional ones. One of

the requirements is to obtain the same Lagrangian expression if the order α becomes 1.

The most general case was investigated in [32], namely the fractional lagrangian was considered as $L\left(t, q^\rho, {}_a\mathbf{D}_t^\alpha q^\rho, {}_t\mathbf{D}_b^\beta q^\rho\right)$, where $\rho = 1, \dots, n$. Let $J[q^\rho]$ be a functional as given below

$$\int_a^b L\left(t, q^\rho, {}_a\mathbf{D}_t^\alpha q^\rho, {}_t\mathbf{D}_b^\beta q^\rho\right) dt, \quad (7)$$

where $\rho = 1 \dots n$ defined on the set of n functions which have continuous left Riemann–Liouville fractional derivative of order α and right Riemann–Liouville fractional derivative of order β in $[a, b]$ and satisfy the boundary conditions $q^\rho(a) = q_a^\rho$ and $q^\rho(b) = q_b^\rho$.

In [32] it was proved that a necessary condition for $J[q^\rho]$ to admit an extremum for given functions $q^\rho(t)$, $\rho = 1, \dots, n$ is that $q^\rho(t)$ satisfies the following fractional Euler–Lagrange equations

$$\frac{\partial L}{\partial q^\rho} + {}_t\mathbf{D}_b^\alpha \frac{\partial L}{\partial {}_a\mathbf{D}_t^\alpha q^\rho} + {}_a\mathbf{D}_t^\beta \frac{\partial L}{\partial {}_t\mathbf{D}_b^\beta q^\rho} = 0, \rho = 1, \dots, n. \quad (8)$$

3 Fractional Lagrangian Treatment of Field Theory

3.1 Fractional classical fields

A covariant form of the action would involve a Lagrangian density $\mathcal{L}$ via $S = \int \mathcal{L} d^3x dt$ where $\mathcal{L} = \mathcal{L}(\phi, \partial_\mu \phi)$ and with $L = \int \mathcal{L} d^3x$. The classical covariant Euler–Lagrange equation are given below

$$\frac{\partial \mathcal{L}}{\partial \phi} - \partial_\mu \frac{\partial \mathcal{L}}{\partial (\partial_\mu \phi)} = 0. \quad (9)$$

Here ϕ denotes the field variable.

In the following the fractional generalization of the above Lagrangian density is developed. Let us consider the action function of the form

$$S = \int \mathcal{L}\left(\phi(x), (\mathbf{D}_{a_k-}^{\alpha_k})\phi(x), (\mathbf{D}_{a_k+}^{\alpha_k})\phi(x), x\right) d^3x dt, \quad (10)$$

where $0 < \alpha_k \leq 1$ and a_k correspond to x_1, x_2, x_3 and t respectively. Let us consider the ϵ finite variation of the functional $S(\phi)$, that we write with explicit dependence from the fields and their fractional derivatives, namely

$$\begin{aligned} \Delta_\epsilon S(\phi) = \int & [\mathcal{L}(x^\mu, \phi + \epsilon \delta \phi, (\mathbf{D}_{a_k-}^{\alpha_k})\phi(x) + \epsilon (\mathbf{D}_{a_k-}^{\alpha_k})\delta \phi, (\mathbf{D}_{a_k+}^{\alpha_k})\phi(x) \\ & + \epsilon (\mathbf{D}_{a_k+}^{\alpha_k})\delta \phi) - \mathcal{L}(x^\mu, \phi, (\mathbf{D}_{a_k-}^{\alpha_k})\phi(x), (\mathbf{D}_{a_k+}^{\alpha_k})\phi(x))] d^3x dt. \end{aligned} \quad (11)$$

We develop the first term in the square brackets, which is a function on ϵ , as a Taylor series in ϵ and we retain only the first order. By using (11) we obtain

$$\begin{aligned} \Delta_\epsilon S(\phi) = & \int [\mathcal{L}(x, \phi, (\mathbf{D}_{\infty-}^{\alpha_k})\phi(x), (\mathbf{D}_{\infty+}^{\alpha_k})\phi(x)) + (\frac{\partial \mathcal{L}}{\partial \phi} \delta \phi) \epsilon \\ & + \sum_{k=1}^4 \frac{\partial \mathcal{L}}{\partial (\mathbf{D}_{\infty-}^{\alpha_k} \phi)} \delta (\mathbf{D}_{\infty-}^{\alpha_k} \phi) \epsilon + \sum_{k=1}^4 \frac{\partial \mathcal{L}}{\partial (\mathbf{D}_{\infty+}^{\alpha_k} \phi)} \delta (\mathbf{D}_{\infty+}^{\alpha_k} \phi) \epsilon + O(\epsilon) \\ & - \mathcal{L}(x, \phi, (\mathbf{D}_{\infty-}^{\alpha_k})\phi(x), (\mathbf{D}_{\infty+}^{\alpha_k})\phi(x))] d^3 x dt. \end{aligned} \quad (12)$$

Taking into account (12) the form of (11) becomes

$$\begin{aligned} \Delta_\epsilon S(\phi) = & \epsilon \int [\frac{\partial \mathcal{L}}{\partial \phi} \delta \phi + \sum_{k=1}^4 \frac{\partial \mathcal{L}}{\partial (\mathbf{D}_{a_k-}^{\alpha_k} \phi)} (\mathbf{D}_{a_k-}^{\alpha_k} \delta \phi) \\ & + \sum_{k=1}^4 \frac{\partial \mathcal{L}}{\partial (\mathbf{D}_{a_k+}^{\alpha_k} \phi)} (\mathbf{D}_{a_k+}^{\alpha_k} \delta \phi) + O(\epsilon)] d^3 x dt. \end{aligned} \quad (13)$$

The next step is to perform a fractional integration by parts of the second term in (13) by making use of the following formula [2]

$$\int_{-\infty}^{\infty} f(x) (\mathbf{D}_{a_k+}^{\alpha_k} g)(x) dx = \int_{-\infty}^{\infty} g(x) (\mathbf{D}_{a_k-}^{\alpha_k} f)(x) dx. \quad (14)$$

As a result we obtain

$$\begin{aligned} \Delta_\epsilon \mathcal{L}(\phi) = & \epsilon \int [\frac{\partial \mathcal{L}}{\partial \phi} \delta \phi + \sum_{k=1}^4 \{(\mathbf{D}_{a_k+}^{\alpha_k}) \frac{\partial \mathcal{L}}{\partial (\mathbf{D}_{a_k-}^{\alpha_k} \phi)}\} \delta \phi \\ & + \sum_{k=1}^4 \{(\mathbf{D}_{a_k-}^{\alpha_k}) \frac{\partial \mathcal{L}}{\partial (\mathbf{D}_{a_k+}^{\alpha_k} \phi)}\} \delta \phi] d^3 x dt + \int O(\epsilon) d^3 x dt. \end{aligned} \quad (15)$$

After taking the limit $\lim_{\epsilon \rightarrow 0} \frac{\Delta_\epsilon S(\phi)}{\epsilon}$ we obtain the fractional Euler-Lagrange equations as given before

$$\frac{\partial \mathcal{L}}{\partial \phi} + \sum_{k=1}^4 \{(\mathbf{D}_{a_k+}^{\alpha_k}) \frac{\partial \mathcal{L}}{\partial (\mathbf{D}_{a_k-}^{\alpha_k} \phi)} + (\mathbf{D}_{a_k-}^{\alpha_k}) \frac{\partial \mathcal{L}}{\partial (\mathbf{D}_{a_k+}^{\alpha_k} \phi)}\} = 0. \quad (16)$$

We observe that for $\alpha_k \rightarrow 1$, the equations (16) are the usual Euler-Lagrange equations for classical fields.

3.2 Fractional Schrödinger equation

Let us consider the Schrödinger wave mechanics for a single particle in a potential $V(x)$. The classical Lagrangian to start with is given as follows

$$\mathcal{L} = \frac{i\hbar}{2}(\psi^\dagger \dot{\psi} - \dot{\psi}^\dagger \psi) - \frac{\hbar^2}{2m} \nabla \psi^\dagger \nabla \psi - V(x) \psi \psi^\dagger. \quad (17)$$

The most general fractional generalization of (12) becomes

$$\mathcal{L} = \frac{i\hbar}{2}(\psi^\dagger \mathbf{D}_{a_t+}^{\alpha_t} \psi - \psi \mathbf{D}_{a_t+}^{\alpha_t} \psi^\dagger) - \frac{\hbar^2}{2m} \mathbf{D}_{a_x+}^{\alpha_x} \psi \mathbf{D}_{a_x+}^{\alpha_x} \psi^\dagger - V(x) \psi \psi^\dagger. \quad (18)$$

Let us consider now that all terminal points are equal to $-\infty$ and denote $\mathbf{D}_{-\infty+}^{\alpha_k}$ by $\mathbf{D}_+^{\alpha_k}$ and $\mathbf{D}_{-\infty-}^{\alpha_k}$ by $\mathbf{D}_-^{\alpha_k}$, respectively. As a result the Euler–Lagrange equations for ψ and $\psi^\dagger$ become

$$\frac{i\hbar}{2}(\mathbf{D}_+^{\alpha_t} \psi - \mathbf{D}_-^{\alpha_t} \psi) - \frac{\hbar^2}{2m} (\mathbf{D}_-^{\alpha_x} \mathbf{D}_+^{\alpha_x}) \psi - V(x) \psi = 0, \quad (19)$$

$$\frac{i\hbar}{2}(-\mathbf{D}_+^{\alpha_t} \psi^\dagger + \mathbf{D}_-^{\alpha_t} \psi^\dagger) - \frac{\hbar^2}{2m} (\mathbf{D}_-^{\alpha_x} \mathbf{D}_+^{\alpha_x}) \psi^\dagger - V(x) \psi^\dagger = 0. \quad (20)$$

We observe that if $\alpha_k \rightarrow 1$ the usual Schrödinger equation is obtain.

4 Fractional Hamiltonian Formulations

4.1 Riewe approach

In the following we briefly review Riewe’s formulation of fractional generalization of Lagrangian and Hamiltonian equations of motion. The starting point is the following action

$$S = \int_a^b L(q_n^r, Q_{n'}^r, t) dt. \quad (21)$$

Here the generalized coordinates are defined as

$$q_n^r = ({}_a \mathbf{D}_t^\alpha)^n x_r(t), \quad Q_{n'}^r = ({}_t \mathbf{D}_b^\alpha)^{n'} x_r(t), \quad (22)$$

and $r = 1, 2, \dots, R$ represents the number of fundamental coordinates, $n = 0, \dots, N$, is the sequential order of the derivatives defining the generalized coordinates q , and $n' = 1, \dots, N'$ denotes the sequential order of the derivatives in definition of the coordinates Q .

A necessary condition for S to posses an extremum for given functions $x_r(t)$ is that $x_r(t)$ fulfill the Euler–Lagrange equations

$$\frac{\partial L}{\partial q_0^r} + \sum_{n=1}^N ({}_t\mathbf{D}_b^\alpha)^n \frac{\partial L}{\partial q_n^r} + \sum_{n'=1}^{N'} ({}_a\mathbf{D}_t^\alpha)^{n'} \frac{\partial L}{\partial Q_{n'}^r} = 0. \quad (23)$$

The generalized momenta have the following form

$$\begin{aligned} p_n^r &= \sum_{k=n+1}^N ({}_t\mathbf{D}_b^\alpha)^{k-n-1} \frac{\partial L}{\partial q_k^r}, \\ \pi_{n'}^r &= \sum_{k=n'+1}^{N'} ({}_a\mathbf{D}_t^\alpha)^{k-n'-1} \frac{\partial L}{\partial Q_k^r}. \end{aligned} \quad (24)$$

Thus, the canonical Hamiltonian is given by

$$H = \sum_{r=1}^R \sum_{n=0}^{N-1} p_n^r q_{n+1}^r + \sum_{r=1}^R \sum_{n'=0}^{N'-1} \pi_{n'}^r Q_{n'+1}^r - L. \quad (25)$$

The Hamilton's equations of motion are given below

$$\frac{\partial H}{\partial q_N^r} = 0, \quad \frac{\partial H}{\partial Q_{N'}^r} = 0. \quad (26)$$

For $n = 1, \dots, N$, $n' = 1, \dots, N'$ we obtain the following equations of motion

$$\frac{\partial H}{\partial q_n^r} = {}_t\mathbf{D}_b^\alpha p_n^r, \quad \frac{\partial H}{\partial Q_{n'}^r} = {}_a\mathbf{D}_t^\alpha \pi_{n'}^r, \quad (27)$$

$$\frac{\partial H}{\partial q_0^r} = -\frac{\partial L}{\partial q_0^r} = {}_t\mathbf{D}_b^\alpha p_0^r + {}_a\mathbf{D}_t^\alpha \pi_0^r. \quad (28)$$

The remaining equations are given by

$$\frac{\partial H}{\partial p_n^r} = q_{n+1}^r = {}_a\mathbf{D}_t^\alpha q_n^r, \quad \frac{\partial H}{\partial \pi_{n'}^r} = Q_{n'+1}^r = {}_t\mathbf{D}_b^\alpha Q_{n'}^r, \quad (29)$$

$$\frac{\partial H}{\partial t} = -\frac{\partial L}{\partial t}, \quad (30)$$

where, $n = 0, \dots, N$, $n' = 1, \dots, N'$.

4.2 Fractional Hamiltonian formulation of constrained systems

Let us consider the action (21) in the presence of constraints

$$\Phi_m(t, q_0^1, \dots, q_0^R, q_n^r, Q_{n'}^r) = 0, m < R. \quad (31)$$

In order to obtain the Hamilton's equations for the the fractional variational problems presented by Agrawal in [32], we redefine the left and the right canonical momenta as :

$$\begin{aligned}
p_n^r &= \sum_{k=n+1}^N ({}_t\mathbf{D}_b^\alpha)^{k-n-1} \frac{\partial \bar{L}}{\partial q_k^r}, \\
\pi_{n'}^{r'} &= \sum_{k=n'+1}^{N'} ({}_a\mathbf{D}_t^\alpha)^{k-n'-1} \frac{\partial \bar{L}}{\partial Q_k^{r'}}.
\end{aligned} \tag{32}$$

Here

$$\bar{L} = L + \lambda_m \Phi_m(t, q_0^1, \dots, q_0^R, q_n^r, Q_{n'}^{r'}), \tag{33}$$

where λ_m represents the Lagrange multiplier and $L(q_n^r, Q_{n'}^{r'}, t)$.

Using (32), the canonical Hamiltonian becomes

$$\bar{H} = \sum_{r=1}^R \sum_{n=0}^{N-1} p_n^r q_{n+1}^r + \sum_{r=1}^R \sum_{n'=0}^{N'-1} \pi_{n'}^{r'} Q_{n'+1}^{r'} - \bar{L}. \tag{34}$$

Then, the modified canonical equations of motion are obtained as

$$\{q_n^r, \bar{H}\} = {}_t\mathbf{D}_b^\alpha p_n^r, \quad \{Q_{n'}^{r'}, \bar{H}\} = {}_a\mathbf{D}_t^\alpha \pi_{n'}^{r'}, \tag{35}$$

$$\{q_0^r, \bar{H}\} = {}_t\mathbf{D}_b^\alpha p_0^r + {}_a\mathbf{D}_t^\alpha \pi_0^{r'}, \tag{36}$$

where, $n = 1, \dots, N$, $n' = 1, \dots, N'$.

The other set of equations of motion are given by

$$\{p_n^r, \bar{H}\} = q_{n+1}^r = {}_a\mathbf{D}_t^\alpha q_n^r, \quad \{\pi_{n'}^{r'}, \bar{H}\} = Q_{n'+1}^{r'} = {}_t\mathbf{D}_b^\alpha Q_{n'}^{r'}, \tag{37}$$

$$\frac{\partial \bar{H}}{\partial t} = -\frac{\partial \bar{L}}{\partial t}. \tag{38}$$

Here, $n = 0, \dots, N$, $n' = 1, \dots, N'$ and the commutator $\{, \}$ is the Poisson's bracket defined as

$$\{A, B\}_{q_n^r, p_n^r, Q_{n'}^{r'}, \pi_{n'}^{r'}} = \frac{\partial A}{\partial q_n^r} \frac{\partial B}{\partial p_n^r} - \frac{\partial B}{\partial q_n^r} \frac{\partial A}{\partial p_n^r} + \frac{\partial A}{\partial Q_{n'}^{r'}} \frac{\partial B}{\partial \pi_{n'}^{r'}} - \frac{\partial B}{\partial Q_{n'}^{r'}} \frac{\partial A}{\partial \pi_{n'}^{r'}}, \tag{39}$$

where, $n = 0, \dots, N$, $n' = 1, \dots, N'$.

5 Fractional Path Integral Formulation

In this section we define the fractional path integral as a generalization of the classical path integral for fractional field systems. The fractional path integral for unconstrained systems emerges as follows

$$K = \int d\phi \, d\pi^\alpha \, d\pi^\beta \exp i \left\{ \int d^4x \left(\pi^\alpha {}_a\mathbf{D}_t^\alpha \phi + \pi^\beta {}_t\mathbf{D}_b^\beta \phi - \mathcal{H} \right) \right\}. \tag{40}$$

5.1 Dirac field

Lagrangian density for Dirac fields of order $2/3$ is proposed as follows [36]

$$\mathcal{L} = \bar{\psi} \left(\gamma^\alpha \mathbf{D}_{\alpha-}^{2/3} \psi(x) + (m)^{2/3} \psi(x) \right). \quad (41)$$

By using (41) the generalized momenta become

$$(\pi_{t-})_\psi = \bar{\psi} \gamma^0, \quad (\pi_{t-})_{\bar{\psi}} = 0. \quad (42)$$

From (41) and (42) we construct the canonical Hamiltonian density as

$$\mathcal{H}_T = -\bar{\psi} \left(\gamma^k \mathbf{D}_{k-}^{2/3} \psi(x) + (m)^{2/3} \psi(x) \right) + \lambda_1 [(\pi_{t-})_\psi - \bar{\psi} \gamma^0] + \lambda_2 [(\pi_{t-})_{\bar{\psi}}]. \quad (43)$$

Making use of (43), the canonical equations of motion have the following forms

$$\mathbf{D}_{t+}^{2/3} (\pi_{t-})_\psi = -(m)^{2/3} \bar{\psi}(x) - \mathbf{D}_{k-}^{2/3} \gamma^k \bar{\psi}(x), \quad (44)$$

$$\mathbf{D}_{t+}^{2/3} (\pi_{t-})_{\bar{\psi}} = -(m)^{2/3} \psi(x) - \gamma^k \mathbf{D}_{k-}^{2/3} \psi(x) - \gamma^0 \lambda_1 = 0, \quad (45)$$

$$\mathbf{D}_{t+}^{2/3} \psi = \frac{\partial \mathcal{H}_T}{\partial (\pi_{t-})_\psi} = \lambda_1, \quad (46)$$

$$\mathbf{D}_{t+}^{2/3} \bar{\psi} = \frac{\partial \mathcal{H}_T}{\partial (\pi_{t-})_{\bar{\psi}}} = \lambda_2, \quad (47)$$

which lead us to the following equation of motion

$$\mathbf{D}_+^{2/3} \gamma^\alpha \bar{\psi}(x) + (m)^{2/3} \bar{\psi}(x) = 0, \quad (48)$$

$$\gamma^\alpha \mathbf{D}_+^{2/3} \psi(x) + (m)^{2/3} \psi(x) = 0. \quad (49)$$

The path integral for this system is given by

$$\begin{aligned} K &= \int d(\pi_{t-})_\psi d(\pi_{t-})_{\bar{\psi}} d\psi d\bar{\psi} \delta[(\pi_{t-})_\psi - \bar{\psi} \gamma^0] \delta[(\pi_{t-})_{\bar{\psi}}] \\ &\times \exp i \left[\int d^4x \left\{ (\pi_{t-})_\psi \mathbf{D}_{t-}^{2/3} \psi + (\pi_{t-})_{\bar{\psi}} \mathbf{D}_{t-}^{2/3} \bar{\psi} - \mathcal{H} \right\} \right]. \end{aligned} \quad (50)$$

Integrating over $(\pi_{\alpha-})_\psi$ and $(\pi_{\alpha-})_{\bar{\psi}}$, we arrive at the result

$$K = \int d\psi d\bar{\psi} \exp i \left[\int d^4x \mathcal{L} \right]. \quad (51)$$

5.2 Nonrelativistic particle interacting with external electromagnetism field

Let us consider the Lagrangian for a nonrelativistic particle of mass m and charge e in an external field as

$$S = \int_a^b \left(\frac{m}{2} \left(\frac{dx_k}{dt} \right)^2 - eA_k(x)\dot{x}_k \right) dt, \quad k = 1, 2, 3. \quad (52)$$

The corresponding action in fractional mechanics looks as follows:

$$S = \int_a^b \left(\frac{m}{2} ({}_a\mathbf{D}_t^{\alpha_m} x_k)^2 - eA_k(x)({}_a\mathbf{D}_t^{\alpha_m} x_k) \right) dt. \quad (53)$$

If we assume $0 < \alpha_m < 1$ and take the limit $\alpha_m \rightarrow 1^+$ we recover the classical model.

The path integral for this system is given by

$$k = \int \prod_{i=0}^{m-1} d({}_a\mathbf{D}_t^{\alpha_i} x_k) \exp^{i\left\{ \int_a^b \left(\frac{m}{2} ({}_a\mathbf{D}_t^{\alpha_m} x_k)^2 - A_k(x)({}_a\mathbf{D}_t^{\alpha_m} x_k) \right) dt \right\}} \quad \alpha_0 = 0. \quad (54)$$

For all $\alpha_m \rightarrow 1^+$, we obtain the path integral for the classical system.

6 Conclusions

We have presented the fractional extensions of the usual Euler–Lagrange equations of motion for both discrete and field theories. As an example the fractional Schrödinger equation for a single particle moving in a potential $V(x)$ was obtained from a fractional variational principle. The fractional Hamiltonian was constructed by using the Riewe’s formulation and the extension of Agrawal’s approach for the case of fractional constrained systems was presented. The classical results are recovered under the limit $\alpha \rightarrow 1$. The existence of various definitions of fractional derivatives and the nonlocality property of fractional Lagrangians make the notion of fractional mechanical constrained systems not an easy notion to be defined. Therefore we have to take into account the nonlocality property during the fractional quantization procedure. For a given fractional constrained mechanical system a Poisson bracket was defined and it reduces to the classical case under certain limits. The fractional path integral approach was analyzed and the fractional actions for Dirac’s field and nonrelativistic particle interacting with external electromagnetism field were found. We mention that in this manuscript the fractional path integral formulation represents the fractional generalization of the classical case. We stress on the fact the fractional path integral formulation depends on the definitions of the momenta and the fractional Hamiltonian.

Acknowledgments

Dumitru Baleanu would like to thank O. Agrawal and J. A. Tenreiro Machado for interesting discussions. Sami I. Muslih would like to thank the Abdus Salam International Center for Theoretical Physics, Trieste, Italy, for support and hospitality during the preliminary preparation of this work. The authors would like to thank ASME for allowing them to republish some results which were published already in proceedings of IDETC/CIE 2005, the ASME 2005 Inter-International Design Engineering Technical Conference and Computers and information in Engineering Conference, September 24–28, 2005, Long Beach, California, USA. This work was done within the framework of the Associateship Scheme of the Abdus Salam ICTP.

References

1. Miller KS, Ross B (1993) *An Introduction to the Fractional Calculus and Fractional Differential Equations*. Wiley, New York.
2. Samko SG, Kilbas AA, Marichev OI (1993) *Fractional Integrals and Derivatives – Theory and Applications*. Gordon and Breach, Linghorne, PA.
3. Oldham KB, Spanier J (1974) *The Fractional Calculus*. Academic Press, New York.
4. Podlubny I (1999) *Fractional Differential Equations*. Academic Press, New York.
5. Hilfer I (2000) *Applications of Fractional Calculus in Physics*. World Scientific, New Jersey.
6. Vinagre BM, Podlubny I, Hernandez A, Feliu V (2000) Some approximations of fractional order operators used in control theory and applications. *Fract. Calc. Appl. Anal.*, 3(3):231–248.
7. Silva MF, Machado JAT, Lopes AM (2005) Modelling and simulation of artificial locomotion systems; *Robotica*, 23(5):595–606.
8. Mainardi F (1996) Fractional relaxation-oscillation and fractional diffusion-wave phenomena. *Chaos, Solitons and Fractals*, 7(9):1461–1477.
9. Zaslavsky GM (2005) *Hamiltonian Chaos and Fractional Dynamics*. Oxford University Press, Oxford.
10. Mainardi F (1996) The fundamental solutions for the fractional diffusion-wave equation. *Appl. Math. Lett.*, 9(6):23–28.
11. Tenreiro Machado JA (2003) A probabilistic interpretation of the fractional order differentiation. *Fract. Calc. Appl. Anal.*, 1:73–80.
12. Raberto M, Scalas E, Mainardi F (2002) Waiting-times and returns in high-frequency financial data: an empirical study. *Physica A*, 314(1–4):749–755.
13. Ortigueira MD (2003) On the initial conditions in continuous-time fractional linear systems. *Signal Processing*, 83(11):2301–2309.
14. Agrawal OP (2004) Application of fractional derivatives in thermal analysis of disk brakes. *Nonlinear Dynamics*, 38(1–4):191–206.
15. Tenreiro Machado JA (2001) Discrete-time fractional order controllers. *Fract. Calc. Appl. Anal.*, 4(1):47–68.
16. Lorenzo CF, Hartley TT (2004) Fractional trigonometry and the spiral functions. *Nonlinear Dynamics*, 38(1–4):23–60.
17. Baleanu D, Avkar T (2004) Lagrangians with linear velocities within Riemann-Liouville fractional derivatives. *Nuovo Cimento*, B119:73–79.

18. Baleanu D (2005) About fractional calculus of singular Lagrangians. JACIII, 9(4):395–398.
19. Diethelm K, Ford NJ, Freed AD, Luchko Yu (2005) Algorithms for the fractional calculus: a selection of numerical methods. Comput. Methods Appl. Mech. Engrg., 194:743–773.
20. Blutzer RL, Torvik PJ (1996) On the fractional calculus model of viscoelastic behaviour. J. Rheology, 30:133–135.
21. Chatterjee A (2005) Statistical origins of fractional derivatives in viscoelasticity. J. Sound Vibr., 284:1239–1245.
22. Adolfsson K, Enelund M, Olsson P (2005) On the fractional order model of viscoelasticity. Mechanic of Time-Dependent Mater, 9:15–34.
23. Metzler R, Joseph K (2000) Boundary value problems for fractional diffusion equations. Physica A, 278:107–125.
24. Magin RL (2004) Fractional calculus in bioengineering. Crit. Rev. Biom. Eng., 32(1):1–104.
25. Zaslavsky GM (2002) Chaos, fractional kinetics, and anomalous transport. Phys. Rep., 371(6):461–580.
26. Rabei EM, Alhalholy TS (2004) Potentials of arbitrary forces with fractional derivatives. Int. J. Mod. Phys. A, 19(17–18):3083–3092.
27. Bauer PS (1931) Dissipative dynamical systems I. Proc. Natl. Acad. Sci., 17:311–314.
28. Riewe F (1996) Nonconservative Lagrangian and Hamiltonian mechanics. Phys. Rev., E53:1890–1899.
29. Riewe F (1997) Mechanics with fractional derivatives, Phys. Rev. E 55:3581–3592.
30. Klimek M (2001) Fractional sequential mechanics-models with symmetric fractional derivatives. Czech. J. Phys., 51, pp. 1348–1354.
31. Klimek M (2002) Lagrangean and Hamiltonian fractional sequential mechanics. Czech. J. Phys. 52:1247–1253.
32. Agrawal OP (2002) Formulation of *Euler – Lagrange* equations for fractional variational problems. J. Math. Anal. Appl., 272:368–379.
33. Laskin N (2002) Fractals and quantum mechanics. Chaos, 10(4):780–790.
34. Laskin N (2000) Fractional quantum mechanics and *Lévy* path integrals, Phys. Lett., A268(3):298–305.
35. Dreisigmeyer DW, Young PM (2003) Nonconservative Lagrangian mechanics: a generalized function approach. J. Phys. A. Math. Gen., 36:8297–8310.
36. Raspini A (2001) Simple solutions of the fractional Dirac equation of order $2/3$. Physica Scr., 4:20–22.
37. Muslih S, Baleanu D (2005) Hamiltonian formulation of systems with linear velocities within Riemann-Liouville fractional derivatives. J. Math. Anal. Appl., 304(3):599–606.
38. Baleanu D, Muslih S (2005) Lagrangian formulation of classical fields within Riemann-Liouville fractional derivatives, Physica Scr., 72(2–3):119–121.
39. Muslih SI, Baleanu D (2005) Quantization of classical fields with fractional derivatives. Nuovo Cimento, 120:507–512.

FRACTIONAL KINETICS IN PSEUDOCHAOTIC SYSTEMS AND ITS APPLICATIONS

George M. Zaslavsky

Courant Institute of Mathematical Sciences and Department of Physics,
New York University, 251 Mercer Street, New York, NY 10012;
E-mail: zaslav@cims.nyu.edu

Abstract

The phenomenon of stickiness of the dynamical trajectories to the domains of periodic orbits (islands), or simply to periodic orbits, can be considered a primary source of the fractional kinetic equation (FKE). An additional condition for the FKE occurrence is a property of the corresponding sticky domains to have space-time invariance under the space-time renormalization transform. The dynamics in some class of polygonal billiards is pseudochaotic (i.e., dynamics is random but the Lyapunov exponent is zero), and the corresponding features of the self-similarity are reflected in the discrete space-time renormalization invariance. We consider an example of such a billiard and its dynamical and kinetic properties that leads to the FKE.

Keywords

Fractional kinetics, pseudochaos, recurrences, billiards.

1 Introduction

In this paper we would like to focus on a class of dynamical systems for which one can use the equations with fractional derivatives as a natural way to describe the most significant features of the dynamics. The first characteristic property of the systems under consideration is that their dynamics is chaotic, or random, or mixed. Chaotic dynamics means the existence of a nonzero Lyapunov exponent. Random dynamics means nonpredictable motion with zero Lyapunov exponents. Mixed dynamics means an alternation of the finite time Lyapunov exponent between almost zero and nonzero values. The second case is called pseudochaos and the last case can be close to either chaos or to pseudochaos, depending on the situation. Additional insight into chaos and pseudochaos is given in the review paper [1]. It becomes clear that the last

two cases correspond to a random dynamics that cannot be described by the processes of the Gaussian or Poissonian, or similar types with all finite moments. A more adequate description of chaos and pseudochaos corresponds to the process of the Lévy type, with infinite second and higher moments, due to the nonuniformity of the phase space of dynamical systems. The so-called fractional kinetic equation (FKE) was introduced in [2–4] in which the ideas of Lévy flights and fractal time [5] were applied to the specific characteristic of the randomness generated by the instability of the dynamics, rather than by the presence of external random forces.

A typical FKE has the form

$$\frac{\partial^\beta F(y, t)}{\partial t^\beta} = \mathcal{D} \frac{\partial^\alpha F(y, t)}{\partial |y|^\alpha}, \quad (0 < \beta \leq 1, \quad 0 < \alpha \leq 2) \quad (1)$$

where $F(y, t)$ is the probability density function, and fractional derivatives could be of arbitrary type, specifically depending on the physical situation of the initial-boundary conditions, etc. More discussions on this subject and different modifications of (1) can be found in [6]. The general type of literature related to the FKE is fairly large (see references in [1] and [7]). This work will be restricted to specific dynamical systems.

The most important issue of application of (1) to the dynamical systems is that exponents (α, β) are defined by the dynamics only and, in some way, they characterize the local property of instability of trajectories. This provides a possibility to find the values of (α, β) from the first principles, and this will be the subject of this paper where the dynamics in some rectangular billiards will be considered, and a review of some previous results, as well as new ones, will be presented.

2 Definition of Pseudochaos

Consider a standard definition of the finite-time Lyapunov exponent σ_t [8]:

$$\sigma_t = \frac{1}{t} \ln[d(t)/d(0)] \quad (2)$$

where $d(t)$ is a distance between two trajectories started in a very small domain A , such that $d_0 \leq \text{diam } A$. The function σ_t is fairly complicated and depends on the choice of A in the full phase space Γ and on d_0 . To simplify the approach one can consider a coarse-graining (smoothing) of σ_t over arbitrary small volume $\delta\Gamma(A) \rightarrow 0$. Consider the measure

$$\begin{aligned} dP(\sigma_t; t_{\max}, \delta\Gamma) &\rightarrow P_{\sigma_t} d\sigma_t, \\ t_{\max} &\rightarrow \infty, \quad \delta\Gamma(A) \rightarrow 0 \end{aligned} \quad (3)$$

that characterizes a distribution function of σ_t . This system is called pseudochaotic if

$$\lim_{t \rightarrow \infty} P_{\sigma_t}(\sigma_t = 0) \neq 0, \quad (4)$$

i.e., for fairly large t there exists a finite domain near the $\sigma_t = 0$ where the probability to find almost zero Lyapunov exponent is nonzero. Such a situation reveals the so-called stickiness of trajectories to the borders of domains of regular or periodic dynamics [2–4], and it was, for example, explicitly demonstrated in [9] for tracers in 3-vortex system.

The sticky domain can be of zero volume. It also can be that

$$P_{\sigma_\infty} = \delta(\sigma_\infty), \quad (5)$$

where $\delta(x)$ is δ -function, i.e., the system has only zero Lyapunov exponent and at the same time the trajectories are nonintegrable. In this case, initially close trajectories diverge (for unstable systems) in the subexponential or polynomial way. There are many examples of this type of pseudochaos: interval exchange transformation [10–13]; polygonal billiards [14–16]; round-off error dynamics [17–19]; isometry transformation [20, 21]; overflow in digital filters [22–25]; and others.

Related to the behavior of P_{σ_t} is the distribution of Poincaré recurrences. Consider a small domain A in phase space with the volume $\delta\Gamma(A)$. Then $P_{\text{rec}}(t; A)dt$ in the limit $\delta\Gamma(A) \rightarrow 0$ is a probability of trajectories, started at A , to return to A within time $t \in (t, t + dt)$. This probability depends on A and it is normalized as

$$\int_0^\infty P_{\text{rec}}(t; A)dt = 1 \quad (6)$$

for all positions A . In the uniform phase space $P_{\text{rec}}(t; A) = P_{\text{rec}}(t)$. In many typical Hamiltonian systems with mixed phase space for the major part of phase space

$$P_{\text{rec}}(t; A) \sim 1/t^\gamma, \quad (t \rightarrow \infty) \quad (7)$$

where γ is called recurrence exponent. For the Anosov-type systems

$$P_{\text{rec}}(t; A) = P_{\text{rec}}(t) \sim e^{-ht}, \quad (t \rightarrow \infty) \quad (8)$$

where h is the metric (Kolmogorov-Sinai) entropy.

Recurrence Conjecture: For pseudochaotic systems of the type shown in Fig. 1 $P_{\text{rec}}(t, A)$ follows (7) with A from the major part of phase space.

The connection between P_{σ_t} and $P_{\text{rec}}(t; A)$ is not known well and the study of pseudochaotic dynamics meets numerous difficulties [26–28]. The polygonal billiards have zero Lyapunov exponent and they are a good example of pseudochaos to be studied [1, 11, 12]. There are two important properties of pseudochaotic billiards that are subjects of this paper: (a) trajectories in some polygonal billiards can be presented on compact invariant surfaces (see Fig. 1) [16, 29]; (b) kinetic description of trajectories in some polygonal billiards can be described by the FKE of the type (1) or similar equation.

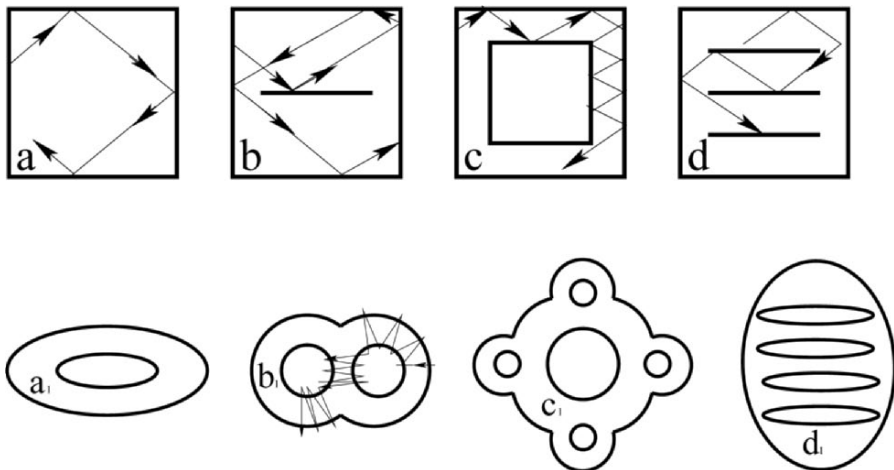


Fig. 1. Four examples of billiards and their corresponding invariant iso-surfaces.

3 The Origin of Fractional Kinetics

In this section we discuss some general principles of the origin of the FKE instead of the regular diffusion equation (see also [2–4]). Let x be a coordinate in the phase space of a system, and for simplicity, $x \in \mathbf{R}^2$. A typical kinetic description of the system evolution appears with a p.d.f. $F(y, t)$ in the reduced space $y \in \mathbf{R}$. For the regular diffusion equation y is a slow variable, usually the action variable. An additional condition is that the renormalization invariance: kinetic equation is invariant with respect to the renormalization group (RG) transform:

$$(\text{RG}): \quad t' = \lambda_t t, \quad y' = \lambda_y y, \quad \lambda_t / \lambda_y^2 = 1 \quad (9)$$

The RG-invariance can be continuous or discrete. The regular diffusion equation

$$\frac{\partial F(y, t)}{\partial t} = \mathcal{D} \frac{\partial^2 F(y, t)}{\partial y^2} \quad (10)$$

satisfies the continuous RG transform (9), i.e., λ_t, λ_y can be arbitrary within the constraint (9). In other words, solutions of (10) can be considered as $F = F(y^2/t)$.

More general situation than (9) implies the following RG-invariance under the transform

$$(\text{RG})_{\alpha\beta}: \quad t' = \lambda_t t, \quad y' = \lambda_y y, \quad \lambda_y^\alpha / \lambda_t^\beta = 1, \quad (11)$$

where (α, β) are fractional in general. Comparing with (9) and (10), the new result appears as an outcome due to two reasons: the specific structure of the

dynamics in phase space, and a coarse-graining or averaging procedure that effectively can reduce the space-time dimensions. In the case $\beta = 1$ and $\alpha = 2$, we arrive at (9). Typically, the dynamics possesses fixed values of the scaling parameters λ_t, λ_y . In this case the RG is discrete and (11) can be replaced by the transform:

$$\begin{aligned} (\text{DRG})_{\alpha\beta}: \quad t' &= \lambda_t t, & y' &= \lambda_y y, \\ \lambda_y^\alpha / \lambda_t^\beta &= \exp(2\pi i m), & (m &= 0, 1, \dots) \end{aligned} \quad (12)$$

As with (10), one can consider FKE (1) and verify that it is invariant with the respect to the $(\text{RG})_{\alpha\beta}$ or $(\text{DRG})_{\alpha\beta}$ transforms. This means that the solution of (1) can be written as

$$F(y, t) = F(|y|/t^{\mu/2}) \quad (13)$$

in the continuous case. The existence of the DRG-invariance implies another form of the distribution function

$$\begin{aligned} F(y, t) &= F_0(|y|/t^{\mu/2}) \\ &\times \left\{ 1 + \sum_{m=1}^{\infty} C_m \cos(2\pi m \ln t / \ln \lambda_t) \right\}, \end{aligned} \quad (14)$$

$(t > 0)$

where we consider only real symmetric functions and

$$\mu = 2\beta/\alpha \quad (15)$$

is the so-called transport exponent. The coefficients C_n are defined by the initial condition. The corresponding equation for (14) will appear later.

The last term in (14) represents the so-called log-periodicity, i.e., periodicity with respect to $\ln t$ with a period

$$T_{\log} = \ln \lambda_t. \quad (16)$$

Its appearance is due to the discreteness of the RG transform (see more in reviews [30] and [1]). The meaning of μ can also be understood from integrating (1) in moments. Let us multiply (1) by $|y|^\alpha$ and integrate it with respect to y . Then it gives

$$\langle |y|^\alpha \rangle = \text{const. } t^\beta \quad (17)$$

for the case (13) or

$$\begin{aligned} \langle |y|^\alpha \rangle &= \text{const. } t^\beta \\ &\times \left\{ 1 + \sum_{m=1}^{\infty} C_m \cos(2\pi m \ln t / \ln \lambda_t) \right\}, \end{aligned} \quad (18)$$

$(t > 0)$

for the case (14). It is assumed that the moments $\langle |y|^\alpha \rangle$ are finite. In fact they have a weak divergence and the average $\langle |y|^\alpha \rangle$ should be performed over $F(y, t)$ that is truncated for $y > y_{\max}$ and $y_{\max}$ depends on the considered t . Our following steps are to show that the introduced cases of FKE (1) with solution of the type (13) or the generalized case (14) can appear in some models related to realistic physical systems.

4 Bar-in-Square Billiard

This simple shape of the billiard (Fig. 1(b)), also called “square-with-slit billiard”, was considered as a model for different applications in plasma and fluids (see [31–34] and references therein). The main results for this billiard can be applied also to the square-in-square billiard (Fig. 1(c)) due to its symmetry.

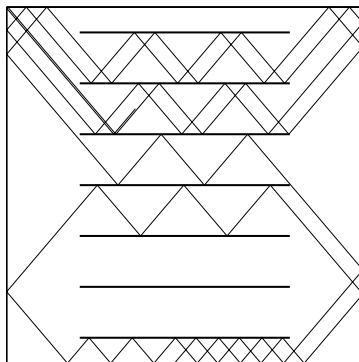


Fig. 2. Double-periodic continuation of the bar-in-square billiard makes a kind of Lorentz gas.

Let us parameterize trajectories in the billiard by coordinates $(x(t), y(t))$. The conservative variables are $|\rho| = (\dot{x}^2 + \dot{y}^2)^{1/2}$ and $\bar{\xi} \equiv |\tan \vartheta| = |\dot{y}(t)/\dot{x}(t)|$. The trajectory is called rational if $\bar{\xi}$ is rational and irrational if $\bar{\xi}$ is irrational. Rational trajectories are periodic and irrational ones perform a random walk along y with weak mixing [14, 15]. Consider an ensemble of irrational trajectories with initial conditions $x_0 \in (x_0 - \Delta x/2, x_0 + \Delta x/2)$, $y_0 \in (y_0 - \Delta y/2, y_0 + \Delta y/2)$, and $F(y, t)$ is a p.d.f. to find a trajectory (a particle) at time t within the interval $(y, y + dy)$:

$$\int_0^\infty F(y, t) dy = 1 \quad (19)$$

in the lifted space, i.e., in the space where the bar-in-square billiard is periodically continued along x and y (see Fig. 2). The function $F(y, t)$ appears as a

result of integrating $F(x, y; t)$ over x [33, 34]. It was shown that there are two scaling parameters for the dynamics of trajectories:

$$\lambda_T = \exp(\pi^2/12 \ln 2), \quad \lambda_a = 2.685 \dots \quad (20)$$

Their origin comes from the theory of continued fractions [35].

Denote

$$\begin{aligned} \bar{\xi} &= \xi_0 + \xi, \\ \xi &= 1/(a_1 + 1/(a_2 + \dots)) \equiv [a_1, a_2, \dots] < 1 \\ \xi_n &= [a_1, a_2, \dots, a_n] = p_n/q_n \end{aligned} \quad (21)$$

where ξ_0 is integer part of $\bar{\xi}$ and p_n, q_n are co-prime. Then there exist two scaling properties

$$\begin{aligned} \lim_{n \rightarrow \infty} \left(\frac{1}{n} \ln q_n \right) &= \ln \lambda_q = \pi^2/12 \ln 2 \approx 1.18 \dots \\ \lim_{n \rightarrow \infty} (a_1 \dots a_n)^{1/n} &= \prod_{k=1}^{\infty} \left(1 + \frac{1}{k^2 + 2k} \right)^{\ln k / \ln 2} \\ &= \lambda_a \approx 2.685 \dots \end{aligned} \quad (22)$$

Since denominator q_n of the n -th approximant defines the period of some rational orbit with the corresponding ϑ_n , we can rewrite (22) in the form

$$\begin{aligned} T_n &\sim \lambda_T^n q_T(n), \quad (n \rightarrow \infty), \\ \prod_{k=1}^n a_k &\sim \lambda_a^n g_a(n), \quad (n \rightarrow \infty), \end{aligned} \quad (23)$$

where g_T, g_a are slow function of n , i.e.,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \ln g_T(n) = \lim_{n \rightarrow \infty} \frac{1}{n} \ln g_a(n) = 0 \quad (24)$$

T_n is a period of rational trajectories for the n -th approximant ξ_n , and

$$\lambda_T = \lambda_q \quad (25)$$

It was shown in [33, 34] that the transport exponent μ can be expressed as

$$\mu = \gamma - 1 \approx 1 + \ln \lambda_a / \ln \lambda_T, \quad (26)$$

where γ was introduced in (7). Our next step is to show how this result can be obtained from the RG approach. From the Eqs. (22) and (20) the expression (26) gives

$$\mu \approx 1.75 \pm 0.1, \quad \gamma \approx 2.75 \pm 0.1 \quad (27)$$

while the simulations give almost the same results.

5 Renormalization Group Equation

Let

$$P_{\text{int}}(t) = \int_0^t P_{\text{rec}}(t) dt \quad (28)$$

is the probability of return to some small domain after time $t' \leq t$. Let also $\{t_n\}$

$$\{t_n\}_1^\infty = \{t_1, t_2, \dots\}, \quad t_k < t_{k+1} \quad (29)$$

is a set of ordered return times and

$$\{T_n\}_1^\infty = \{T_1, T_2, \dots\}, \quad T_k < T_{k+1} \quad (30)$$

is a set of ordered periods of rational trajectories. We assume that

$$P_{\text{int}}(t_n) = P_{\text{int}}(T_n), \quad (n \rightarrow \infty) \quad (31)$$

and the same is for the moments of $P_{\text{int}}(t_n)$. This gives a possibility to apply the scaling properties (23) to the construction of the DRG equation. The main idea of this derivation follows [2–4] and [32].

Consider integrated probability for recurrences $P_{\text{int}}(t_n)$ defined on the discrete set (29) with a boundary condition

$$P_{\text{int}}(\infty) = 1 \quad (32)$$

that follows from (28) and (6). Then we can write for fairly large n :

$$P_{\text{int}}(t_{n+1}) = J_{n,n+1} P_{\text{int}}(t_n) + \Delta P(t_n) \quad (33)$$

where ΔP is a slow function of t_n and $J_{n,n+1}$ is the corresponding Jacobian for the transform of variables. Due to the condition (31), Eq. (33) can be rewritten as

$$P_{\text{int}}(T_{n+1}) = J_{n,n+1} P_{\text{int}}(T_n) + \Delta P(T_n) \quad (34)$$

The most sensitive part of (34) is the Jacobian $J_{n,n+1}$ which depends on the choice of ensemble, the coarse-graining procedure, and the phase space variables. In the considered model the effective phase space element can be defined as

$$d\Gamma_n = dx_n dy_n = dt_n dy_n = dT_n dy_n \quad (35)$$

since it is used for ensemble of trajectories with different y_n , i.e., ϑ_n , and since $d\Gamma_n$ is renormalizable. Then

$$\begin{aligned} J_{n,n+1} &= \frac{d\Gamma_n}{d\Gamma_{n+1}} = \frac{dT_n}{dT_{n+1}} \frac{dy_n}{dy_{n+1}} \\ &= \frac{dT_n}{dT_{n+1}} \frac{d \prod_1^n a_k}{d \prod_1^{n+1} a_k}. \end{aligned} \quad (36)$$

From (23) and (24), it follows

$$J_{n,n+1} = \frac{1}{(\lambda_T \lambda_a)}, \quad (n \rightarrow \infty) \quad (37)$$

and the Eq. (34) arrives to

$$P_{\text{int}}(T_{n+1}) = \frac{1}{(\lambda_T \lambda_a)} P_{\text{int}}(T_{n+1}/\lambda_T) + \Delta P(T_{n+1}) \quad (38)$$

where we use the slowness of $\Delta P(T_n) \approx \Delta P(T_n \lambda_T) = \Delta P(T_{n+1})$.

The obtained Eq. (38) is a typical RG equation in statistical physics, and its solution can be written using the Mellin transform [5,36] (see also [37]), as

$$P_{\text{int}}(t) = t^{-\kappa} \left\{ B_0 + \sum_{m=1}^{\infty} B_m \cos(2\pi m \ln t / \ln \lambda_T) \right\} \quad (39)$$

where

$$\kappa = 1 + \ln \lambda_a / \ln \lambda_T \quad (40)$$

and constants B_n can be obtained using the Mellin transform as in [36] and where we have replaced the discrete variable T_n by t . The leading term of the expansion (38)

$$P_{\text{int}}(t) \sim \text{const. } t^{-\kappa} \quad (41)$$

can be easily understood by substituting it into (38). Omitting the term ΔP we have for the singular part of the solution to (38):

$$t^{-\kappa} = (1/\lambda_T \lambda_a)(t/\lambda_T)^{-\kappa} \quad (42)$$

that leads to (40).

Comparing (40) to (26) we obtain

$$\begin{aligned} \gamma &= 2 + \ln \lambda_a / \ln \lambda_T \\ \mu &= \gamma - 1 = \kappa = 1 + \ln \lambda_a / \ln \lambda_T \end{aligned} \quad (43)$$

Substitutions of values (20) into (43) defines

$$\mu \approx 1.88, \quad \gamma \approx 2.88 \quad (44)$$

in a good agreement with the simulation data from [33,34] (see also [38]).

6 Conclusion

The obtained solution $P_{\text{int}}(t)$ gives for the distribution of Poincaré recurrences

$$\begin{aligned} P_{\text{rec}}(t) &= dP_{\text{int}}(t)/dt \\ &= t^{-\kappa} \left\{ \bar{B}_0 + \sum_{m=1}^{\infty} \bar{B}_m \cos(2\pi m \ln t / \ln \lambda_T) \right\} \end{aligned} \quad (45)$$

similar to (38) but with different coefficients $\bar{B}_m$. One can compare this expression to (14) to see the similarity. Nevertheless, one more step is necessary in order to link these two expressions. At the moment we can state only that the FKE (1) is the simplest equation that satisfies the conditions of RG-invariance (12). In addition to this, the FKE (1) has the same DRG-invariance as the RG equation (38) for the Poincaré recurrences. This result indicates a deep link between the renormalization space-time invariance of the dynamics to that of the renormalization properties of the kinetic equation with space-time fractional derivatives.

In conclusion, let us mention some recent developments for the multi-bar billiards [32] with application of the results to the anomalous transport properties in the tokamaks.

Acknowledgment

This work was supported by the Office of Naval Research, Grant No. N00014-02-1-0056 and by the Department of Energy, Grant No. DE-FG02-92ER54184.

References

1. Zaslavsky GM (2002) *Chaos, fractional kinetics, and anomalous transport*. Phys. Rep., 371:461–580.
2. Zaslavsky GM (1992) Anomalous transport and fractal kinetics. In: *Topological Aspects of the Dynamics of Fluids and Plasmas* (edited by Moffatt HK, Zaslavsky GM, Comte P, Tabor M), pp. 481–492. Kluwer, Dordrecht.
3. Zaslavsky GM (1994a) Fractional kinetic-equation for Hamiltonian chaos. Physica D, 76:110–122.
4. Zaslavsky GM (1994b) Renormalization group theory of anomalous transport in systems with Hamiltonian chaos. Chaos, 4:25–33.
5. Montroll EW, Shlesinger MF (1984) On the wonderful world of random walk. In: *Studies in Statistical Mechanics*, Vol. 11 (edited by Lebowitz J, Montroll EW), pp. 1–121. North-Holland, Amsterdam.
6. Saichev AI, Zaslavsky GM (1997) Fractional kinetic equations: solutions and applications. Chaos, 7:753–764.
7. Metzler R, Klafter J (2000) The random walk's guide to anomalous diffusion: a fractional dynamics approach. Phys. Rep. 339:1–77.
8. Ott E (1993) *Chaos in Dynamical Systems*. Cambridge University Press, Cambridge.
9. Leoncini X, Zaslavsky GM (2002) Jets, stickiness, and anomalous transport. Phys. Rev. E, 65:Art. No. 046216.
10. Adler R, Kitchens B, Tresser C (2001) Dynamics of non-ergodic piece-wise affine maps of the torus. Ergodic Theory Dynam. Sys. 21:959–999.
11. Gutkin E, Katok A (1989) Weakly mixing billiards. In: *Holomorphic Dynamics*, Lecture Notes in Mathematics 1345:163–176. Springer, Berlin.

12. Katok A (1987) The growth-rate for the number of singular and periodic-orbits for a polygonal billiard. *Commun. Math. Phys.*, 111:151–160.
13. Zorich A (1997) Deviation for interval exchange transformations. *Ergodic Theory Dynam. Sys.*, 17:1477–1499.
14. Gutkin E (1986) Billiards in polygons. *Physica D*, 19:311–333.
15. Gutkin, E (1996) Billiards in polygons: Survey of recent results. *J. Stat. Phys.* 83:7–26.
16. Richens PJ, Berry MV (1981) Pseudointegrable systems in classical and quantum-mechanics. *Physica D*, 2:495–512.
17. Kouptsov KL, Lowenstein JH, Vivaldi F (2002) Quadratic rational rotations of the torus and dual lattice maps. *Nonlinearity*, 15:1795–1842.
18. Lowenstein JH, Hatjispyros S, Vivaldi F (1997) Quasi-periodicity, global stability and scaling in a model of Hamiltonian round-off. *Chaos*, 7:49–66.
19. Lowenstein JH, Vivaldi F (1998) Anomalous transport in a model of Hamiltonian round-off. *Nonlinearity*, 11:1321–1350.
20. Lowenstein JH, Kouptsov KL, Vivaldi F (2004) Recursive tiling and geometry of piecewise rotations by $\pi/7$. *Nonlinearity*, 17:1–25.
21. Lowenstein JH, Poggiaspalla G, Vivaldi F. Sticky orbits in a kicked-oscillator model. Not published.
22. Ashwin P (1997) Elliptic behaviour in the sawtooth standard map. *Phys. Lett. A*, 232:409–416.
23. Ashwin P, Chambers W, Petrov G (1997) Lossless digital filter overflow oscillations; approximation of invariant fractals. *Intern. J. Bifur. Chaos*, 7:2603–2610.
24. Chua LO, Lin T (1988) Chaos in digital filters. *IEEE Trans.*, CAS-35:648–658.
25. Chua LO, Lin T (1990) Fractal pattern of 2nd order nonlinear digital filters - a new symbolic analysis. *Int. J. Cir. Theory Appl.*, 18:541–550.
26. Bernardo M, Courbage M, Truong TT (2003) Random walks generated by area preserving maps with zero Lyapunov exponents. *Commun. Nonlin. Sci. and Numerical Simul.*, 8:189–199.
27. Courbage M, Hamdan D (1995) Unpredictability in some nonchaotic dynamical systems. *Phys. Rev. Lett.*, 74:5166–5169.
28. Courbage M, Hamdan D (1997) Decay of correlations and mixing properties in a dynamical system with zero K-S entropy. *Ergodic Theory Dynam. Sys.*, 17:87–103.
29. Galperin GA, Zemlyakov AN (1990) *Mathematical Billiards*. Nauka, Moscow.
30. Sornette D (1998) Discrete-scale invariance and complex dimensions. *Phys. Rep.*, 297:239–270.
31. Carreras BA, Lynch VE, Garcia L, Edelman M, Zaslavsky GM (2003) Topological instability along filamented invariant surfaces. *Chaos*, 13:1175–1187.
32. Zaslavsky GM, Carreras BA, Lynch VE, Garcia L, Edelman M. Topological instability along invariant surfaces and Pseudochaotic Transport. *Phys. Rev.*, E 72:Art. No. 026227.
33. Zaslavsky GM, Edelman M (2001) Weak mixing and anomalous kinetics along filamented surfaces. *Chaos*, 11:295–305.
34. Zaslavsky GM, Edelman M (2004) Fractional kinetics: from pseudochaotic dynamics to Maxwell's Demon. *Physica D*, 193:128–147.
35. Khinchin AY (1964) *Continued Fractions*. University of Chicago Press, Chicago.

36. Hughes BD, Shlesinger MF, Montroll EW (1981) Random walks with self-similar clusters (stochastic processes – stable distributions – fractals – nondifferentiable functions). *Proc. Nat. Acad. Sci. USA*, 78:3287–3291.
37. Niemeijer Th, van Leeuwen MJ (1976) Renormalization theory for Ising-like spin systems. In: *Phase Transition and Critical Phenomena*, Vol. 6 (edited by Domb C, Green MS), pp. 425–501. Academic Press, London.
38. Lyubomudrov O, Edelman M, Zaslavsky GM (2003) Pseudochaotic systems and their fractional kinetics. *Int. J. Mod. Phys., B* 17:4149–4167.

SEMI-INTEGRALS AND SEMI-DERIVATIVES IN PARTICLE PHYSICS

Peter W. Krempl

AVL List GmbH, A-8020 Graz, Austria; E-mail: peter.kremp1@avl.com

Abstract

This paper gives a short review about the application of semi-integrals to the Abel-type integral transform with the kernel $(t^2 - \hat{x})^{-1/2}$ and the properties of this important integral transforms. Its practical application for the instrumentation in accelerator physics to determine the particle beam densities in the transversal phase space in a synchrotron is demonstrated for the CERN proton synchrotron booster (PSB) Beamscope. This device allows the direct observation of the amplitude distribution of the betatron oscillations. It deals further with a space-like description of the wave function of spin-half particles within the Schrödinger picture, one of the most famous non-integer phenomena in physics. It will be shown, that assuming the existence of half-integer derivatives, wave functions for spin-1/2-particles can be derived in just the same way as for the normal angular momentum. These functions satisfies the eigenvalue equations for spin $\pm 1/2$, as well as the change of the spin state applying the creation and annihilation operators. These wave functions display directly the observed 4 Pi symmetry of such particles. This description is complementary to the common description using Pauli matrices and spinors.

Keywords

Fractional calculus, integral equations, particle accelerators, beam, diagnostics, quantum mechanics, spin.

1 Introduction

Semi-integrals and semi-derivatives are defined as fractional integrals or derivatives of the order 1/2. They have been the first objects of fractional calculus considered in the history, as we know from letters between Leibnitz and De L'Hôpital, more than 300 years ago, in which the latter asked for the meaning of the derivative if the order is 1/2, and the former answered: "*Il y a de l'apparence qu'on tirera un jour des conséquences bien utiles de ces paradoxes,*

*car il n'y a guère de paradoxes sans utilité.*Ž Today, there exist numerous applications of this kind of calculus in nearly all fields of natural science, thus Oldham and Spanier devoted a special chapter to this type of fractional operators in the first book on fractional calculus [1].

This contribution will therefore be restricted to two applications in particle physics, like the determination of transversal density distribution in particle beams from measurements of the amplitude distribution of their betatron oscillations. For this purpose, the solution of Abel-type integral equations with the kernel $(t^2 - x^2)^{-1/2}$ with the help of fractional calculus will be considered first. The second application, using only the existence of semi-derivatives, will give a complementary description of the wave function of spin-1/2-particles, to which belong all the known fundamental elementary particles, like leptons and quarks, and which displays directly the 4π of symmetry of these particles.

2 Abel-Type Integral Equations

2.1 Abel's integral equation

Semi-integrals and semi-derivatives are defined as fractional integrals or one of the most famous works, and perhaps the first approach towards application of fractional calculus, was Abel's solution of the tautochrone (or brachistochrone) problem [2], which is recapitulated in [1]. The integral equation, to which his name was given, is

$$g(x) = \int_0^x \frac{f(t)dt}{\sqrt{x-t}} \quad (1)$$

and Abel found its solution

$$f(t) = \frac{1}{\pi} \frac{d}{dt} \int_0^t \frac{g(x)dx}{\sqrt{t-x}} \quad (2)$$

without conscious use of fractional calculus, but he showed as first [3], that it could be written as a fractional derivative. Nevertheless, Laurent [4] solved the integral Eq. (2) using fractional operators. Today, a lot of different definitions for Abel's integral equations are given in the literature and we have to distinguish them carefully. Let us start with the following integral

equation, which is also called Abel equation, with has variable lower integration limit and fixed upper limit $R > 0$:

$$g(x) = \int_x^R \frac{f(t)dt}{(t-x)^{1-\alpha}} = \Gamma(\alpha) {}_x I_R^\alpha f(x) \quad (3)$$

and assume the more general case with $0 < \alpha < 1$, where ${}_x I_R^\alpha f(x)$ denotes the fractional integral:

$${}_x I_R^\alpha f(x) = \frac{1}{\Gamma(\alpha)} \int_x^R (t-x)^{\alpha-1} f(t)dt \quad (4)$$

according to Riemann–Liouville’s definition. Applying fractional calculus we immediately obtain the relation:

$$f(x) = \frac{1}{\Gamma(\alpha)} {}_x I_R^{-\alpha} g(x) = -\frac{1}{\Gamma(\alpha)} {}_x D_R^\alpha g(x) \quad (5)$$

yielding the well known solution:

$$f(x) = -\frac{\sin(\pi\alpha)}{\pi} \frac{d}{dx} \int_x^R \frac{g(\xi)d\xi}{(\xi-x)^{1-\alpha}} \quad (6)$$

to this integral equation. For the special case of $\alpha = 1/2$ we have:

$$g(x) = \int_x^R \frac{f(t)dt}{\sqrt{t-x}} \quad (7)$$

with the inversion formula:

$$f(t) = -\frac{1}{\pi} \frac{d}{dt} \int_t^R \frac{g(x)x dx}{\sqrt{x-t}} \quad (8)$$

2.2 Abel-Type integral equation

Many problems in natural sciences ask for the inversion of the following homogeneous Volterra-type integral equation of the first kind:

$$g(x) = \int_x^R \frac{f(t)dt}{\sqrt{t^2 - x^2}} \equiv A(x, t)f(t) \quad (9)$$

which is denominated as Abel-type integral equation and abbreviated by the operator $A(x, t)$. The inversion of this integral equation by the methods of classical functional analysis has the problem of the non-Hermitian kernel which is not analytic over the whole interval $(0, R)$ due to the two branch points at $t = \pm x$. The substitutions

$$x^2 = \xi, \quad t^2 = \tau, \quad F(\tau) = \frac{f(\sqrt{\tau})}{2\sqrt{\tau}} \quad \text{and} \quad G(\xi) = g(\sqrt{\xi}) \quad (10)$$

give the proper Abel integral equation

$$G(\xi) = \int_{\xi}^{R^2} \frac{F(\tau) d\tau}{\sqrt{\tau - \xi}} \quad (11)$$

Re-substituting of (10) into the solution (8) yields directly the inverse operator:

$$f(t) = -\frac{2}{\pi} \frac{d}{dt} \int_x^R \frac{g(x)x dx}{\sqrt{t^2 - x^2}} = A^{-1}(t, x)g(x) \quad (12)$$

Alternative methods to invert (9), making no explicit use of the fractional calculus are much more sophisticated and are given elsewhere, together with a table of important transformation pairs [5,6]. Further pairs of this transform are listed in [7].

2.3 General properties of the Abel-Type integral operator $A(x, t)$

Before considering the application, some useful formulae concerning the operators defined in (9) should be demonstrated. Obviously, this operator is a linear one. Applying the inverse operator on both sides of

$$\psi(x) = A(x, t) \frac{d}{dt} f(t) \quad (13)$$

and comparing the two differentials yields after a second application of $A^{-1}(t, x)$:

$$\psi(x) = \frac{1}{x} \frac{d}{dx} A(x, t) t f(t) \quad (14)$$

This gives together with (13) the important identity:

$$A(x, t) \frac{d}{dt} = \frac{1}{x} \frac{d}{dx} A(x, t) t \quad (15)$$

which improves in many cases the numerical computation of the Abel inversion. This improvement may also be achieved using

$$A(x, t) \frac{df}{dt} = \int_x^R \frac{f(t) - f(x)}{(t^2 - x^2)^{3/2}} t dt \quad (16)$$

recommended in [8] to avoid the numerical differentiation of (frequently noisy) functions $f(t)$. We can further demonstrate that:

$$-\frac{2}{\pi} \frac{d}{dx} A(x, t) t A(t, x) = I \quad (17)$$

where I denotes the identity operator.

Finally we show the *normalisation theorem*: Let $g(x) = A(x, t)f(t)$ be the transform of $f(t)$ according to (9), then their integrals are related by:

$$\int_0^R g(x) dx = \frac{\pi}{2} \int_0^R f(t) dt \quad (18)$$

which can be proved by entering $f(t) = -(2/\pi)(d/dt)A(t, x) \times g(x)$ into the right-hand side of (18). Thus we get:

$$\frac{\pi}{2} \int_0^R f(t) dt = - \int_0^R \frac{d}{dt} [A(t, x) x g(x)] dt = - \int_{F(0)}^{F(R)} dF(t) \quad (19)$$

$$F(t) = \int_t^R \frac{g(x) x dx}{\sqrt{x^2 - t^2}} \quad (20)$$

Then (18) follows, since $F(R) = 0$, and $F(0)$ becomes according to (20)

$$F(0) = \int_0^R g(x) dx \quad (21)$$

(15) delivers immediately the relation:

$$\frac{d}{dx} A(x, t) = x A(x, t) \frac{d}{dt} \frac{1}{t} \quad (22)$$

The *scaling law* denotes:

$$A(\lambda x, t) = \int_{\lambda x}^{\lambda R} \frac{dt}{\sqrt{t^2 - (\lambda x)^2}} = \int_x^R \frac{d(t/\lambda)}{\sqrt{(t/\lambda)^2 - x^2}} = A(x, t/\lambda) \quad (23)$$

2.4 Geometrical interpretation

A simple geometrical picture can be used to explain the meaning of the integral Eq. (9) in the cases of radial symmetry. Let $P(r)$ denote a radial symmetric surface density or probability distribution of any quantity Q , and $q(r)$ the corresponding radial distribution function, which means:

$$P(r) = \frac{\partial^2 Q}{\partial x \partial y} \Big|_{r=\sqrt{x^2+y^2}} = \frac{1}{2\pi r} \frac{dQ}{dr} = \frac{1}{2\pi r} q(r) \quad (24)$$

then the integral over a strip parallel to the y -axis with the width dx or the so-called projected density distribution $p(x)$ normal to the x -axis is given by:

$$p(x) = \int_{-Y(x)}^{+Y(x)} P(r) dy = 2 \int_x^R \frac{P(r) r dr}{\sqrt{r^2 - x^2}} \quad (25)$$

(where $Y(x) = +(R^2 - x^2)^{1/2}$) as shown in Fig. 1

$$G(\xi) = \int_{\xi}^{R^2} \frac{F(\tau) d\tau}{\sqrt{\tau - \xi}} \quad (26)$$

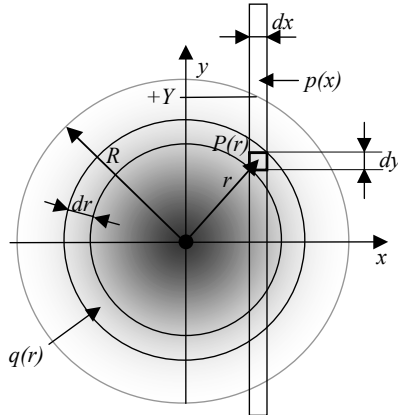


Fig. 1. Definition of surface $P(r)$, radial $q(r)$, and projected $p(x)$ density distributions.

From this it follows that the projected density $p(x)$ is the Abel-type transform of $2rP(r)$ according to the definition (9). Table 1 shows the relations between these three density distributions:

Table 1. Relations between surface, radial, and projected density distributions

	Surface density $P(r)$	Radial density $q(r)$	Projected density $p(x)$
$P(r) =$		$\frac{1}{2\pi r} q(r)$	$-\frac{1}{\pi} A(r, x) \frac{dp}{dx}$
$q(r) =$	$2\pi r P(r)$		$-2r A(r, x) \frac{dp}{dx}$
$p(x) =$	$2A(x, r) r P(r)$	$\frac{1}{\pi} A(x, r) q(r)$	

3 Application to Density Distributions in Circular Particle Accelerators

3.1 Betatron oscillations

In a circular particle accelerator like a synchrotron, the particles perform oscillations around the so-called closed orbit, which would be the orbit for particles with no transversal momentum. This orbit normally does not coincide with the geometrical circumference of the accelerator. It is the only trajectory along which a particle could completely circulate arriving at the same coordinates after each circulation, which causes the name “closed orbit”. In reality, the particle trajectories will deviate from this closed orbit due to their transversal distance x from this orbit into the x -direction and their corresponding transversal momentum. Thus they will arrive at different distances and transverse momenta after each circulation. These transverse oscillations around the closed orbit are called betatron oscillations, and are separated into the two transverse directions (horizontal and vertical) with respect to the closed orbit.

Let $x(s)$ denote the distance of a particle from its closed orbit at the longitudinal (circular) coordinate s . According to the theory of these oscillations [9] (Bruck 1966), this distance is a function of the longitudinal coordinate s , and the particles normalised amplitude a and phase ϕ . Its function:

$$x(s) = a\sqrt{\beta(s)} \cos[\Psi(s) + \phi] \quad (27)$$

describes pseudo-harmonic oscillations around the closed orbit at which $x(s)=0$, where $\beta(s)$ denotes the so-called betatron function which gives the envelope of the trajectories of all particles with the amplitude $a=1$, but different phases ϕ . Introducing the local amplitudes $r(s)$ at the location s :

$$r(s) = a\sqrt{\beta(s)} \quad (28)$$

we find for the projected density distribution $p(x)$ into the x -direction the expression:

$$p(x) = \frac{1}{\pi} A(x, r) q(r) \quad (29)$$

where $q(r)$ denotes the amplitude distribution of these oscillations. An algorithm for the numerical computation of $p(x)$ from the observed amplitude distribution $q(r)$ is given in [10]. If we observe the projected density distribution $p(x)$, we can calculate the amplitude distribution $q(r)$ by the inversion of (29):

$$q(r) = -2rA(r, x) \frac{dp}{dx} \quad (30)$$

The phase space distribution $P(r)$, which is the density in the normalised rotational symmetric transversal phase space, can be obtained by:

$$P(r) = \frac{1}{2\pi r} q(r) = -\frac{1}{\pi} A(r, x) \frac{dp}{dx} \quad (31)$$

From (31) it can be seen, that the amplitude distribution $q(r)$ has to become zero at $r=0$, i.e., for the closed orbit, which means that in reality, no particle can follow this orbit, because there is no place in the phase space.

3.2 Determination of the particle distributions

There are several possibilities to determine one of the two distributions $q(r)$ or $p(x)$. The projected density $p(x)$ can be observed measuring the charge density of the free electrons originating from particle interactions with the residual gas. This method has the advantage that its operation is non-destructive, but due to experimental difficulties and the necessary differentiation in relation (30) the amplitude distribution cannot be determined with sufficient accuracy. The most precise measurements are based on the direct observation of the amplitude distribution by continuous removal of all particles having an amplitude larger than $r_{\max}(t)$ and determination of the remaining beam current and its differential $dI/dr_{\max}$ as a function of time. There are two principal methods to perform such

measurements. The first method consists in a controlled motion of the beam towards a scraping obstacle, and was introduced at the CERN proton synchrotron booster (PSB) under the acronym “Beamscope”, which means “betatron amplitude scraping by closed orbit perturbation” [11]. A local deformation of the closed orbit is produced by three magnetic dipoles, and moves the beam towards a fixed scraping target. The density of the remaining beam current can be measured with a beam current transformer. The other method (fast blade scanner, FBS) consists in the fast injection of a scraper into the beam. The Beamscope has the advantage, that it can be also operated as a

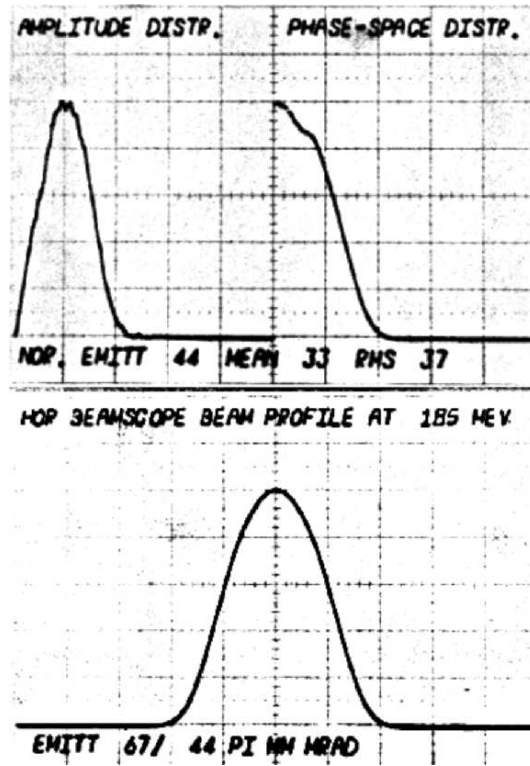


Fig. 2. The three density distributions of a proton beam in the CERN PSB observed with the “Beamscope”. The amplitude distribution was directly observed by analogue differentiation of the slow beam transformer signal, and the other two distributions have been calculated with a simple algorithm running on a digital processing oscilloscope [11]. Only one half of the symmetric phase space density is shown on the upper right side. The projected density is completely displayed to show the calculated beam profile.

beam shaping device, which scrapes only such particles whose amplitudes exceed a certain limit. However, for a complete determination of the amplitude distribution, the beam becomes destroyed. Measuring directly the amplitude distribution $q(a)$, the projected density and the phase space density can be deduced very precisely using (29) or (31), as shown in Fig. 2.

4 Wave Functions of Spin-1/2-Particles

We will now turn towards another application of semi-derivatives which is given in the field of non-relativistic quantum mechanics.

4.1 Quantum description of angular momentum

Let us start with a short recapitulation of quantum mechanics of the angular momentum of any system, for which the classical expression $\mathbf{L}$ is given by:

$$\mathbf{L} = \mathbf{x} \times \mathbf{p} \quad (32)$$

$\mathbf{x}$ denoting the coordinate vector and $\mathbf{p}$ the momentum vector of the system. Replacing the classical variables by their corresponding operators, we obtain its description within the framework of the Schrödinger picture. Thus $\mathbf{x}$ becomes replaced by $\mathbf{x}$, and $\mathbf{p}$ by the momentum operator $\mathbf{p} = -i\hbar\nabla$ (where $\hbar$ denotes Planck's quantum number $h/2\pi$) yielding the angular momentum operator $\mathbf{L}$

$$\mathbf{L} = -i\hbar\mathbf{x} \times \nabla \quad (33)$$

with the Cartesian components in spherical coordinates φ and ϑ :

$$\begin{aligned} \mathbf{L}_x &= i\hbar \left(\sin \varphi \frac{\partial}{\partial \vartheta} + \operatorname{ctg} \vartheta \cos \varphi \frac{\partial}{\partial \varphi} \right) \\ \mathbf{L}_y &= -i\hbar \left(\cos \varphi \frac{\partial}{\partial \vartheta} - \operatorname{ctg} \vartheta \sin \varphi \frac{\partial}{\partial \varphi} \right) \\ \mathbf{L}_z &= -i\hbar \frac{\partial}{\partial \varphi} \end{aligned} \quad (34)$$

These operators do not commute:

$$[\mathbf{L}_j, \mathbf{L}_k] = i\hbar \varepsilon_{jkl} \mathbf{L}_l \quad (35)$$

Its dot product can be written as

$$\mathbf{L}^2 = \mathbf{L}_x^2 + \mathbf{L}_y^2 + \mathbf{L}_z^2 = -\hbar^2 \Delta_{\vartheta, \varphi} = -\hbar^2 \left[\frac{1}{\sin \vartheta} \frac{\partial}{\partial \vartheta} \left(\sin \vartheta \frac{\partial}{\partial \vartheta} \right) + \frac{1}{\sin^2 \vartheta} \frac{\partial^2}{\partial \varphi^2} \right] \quad (36)$$

and commutes with all three components of $\mathbf{L}$:

$$[\mathbf{L}^2, \mathbf{L}_j] = 0 \quad (37)$$

These commutation rules tell us, that the magnitude of the angular momentum can always be exactly determined, but only one of its components can be measured at the same time with arbitrary accuracy. The other two remain indeterminable. Usually, L_z is taken as the measurable component. For a stationary state, the part of the wave function Ψ_L which describes the angular momentum has to be an eigenfunction of the angular momentum operators, thus it has to satisfy the two equations:

$$\mathbf{L}^2 \Psi_L = L^2 \Psi_L \quad (38)$$

and

$$\mathbf{L}_z \Psi_L = L_z \Psi_L \quad (39)$$

where L^2 and L_z represents the associated eigenvalues. The spherical harmonics $\Psi_{l,m}(\vartheta, \varphi)$:

$$Y_{l,m}(\vartheta, \varphi) = \sqrt{\frac{(l-|m|)!(2l+1)}{4\pi(l+|m|)!}} P_l^{|m|}(\cos \vartheta) \exp(im\varphi) \quad (40)$$

with the associated Legendre polynomials satisfy the relations:

$$\mathbf{L}^2 Y_{l,m}(\vartheta, \varphi) = l(l+1)\hbar^2 Y_{l,m}(\vartheta, \varphi) \quad (41)$$

$$\mathbf{L}_z Y_{l,m}(\vartheta, \varphi) = m\hbar Y_{l,m}(\vartheta, \varphi) \quad (42)$$

yielding the eigenvalues

$$L^2 = l(l+1)\hbar^2 \quad \text{and} \quad L_z = m\hbar \quad (43)$$

for L^2 and L_z . The magnitude of the angular momentum

$$L = \hbar \sqrt{l(l+1)} \quad (44)$$

as well as its z -component L_z are quantised with l and m as quantum numbers. The latter can be changed applying the ladder operators

$$\mathbf{L}_{\pm} = \mathbf{L}_x \pm i\mathbf{L}_y \quad (45)$$

to create (+) or annihilate (−) a magnetic quantum according to

$$\mathbf{L}_z(\mathbf{L}_{\pm}\Psi_L) = \hbar(m \pm 1)\mathbf{L}_{\pm}\Psi_L \quad (46)$$

but do not change the magnitude:

$$\mathbf{L}^2(\mathbf{L}_{\pm}\Psi_L) = \hbar^2 l(l+1)\mathbf{L}_{\pm}\Psi_L \quad (47)$$

Since L_x and L_y are both Hermitian (which implies) and $|\mathbf{L}_{\pm}\Psi_L|^2 > 0$, we have the well known restrictions $-l < m < l$ between the so-called angular l and magnetic m quantum numbers.

4.2 Spin-1/2-particles in fractional description

Since the discovery of the electron spin, 1925, this quantum state keeps some mystery, since it cannot be described like an angular momentum, but shows experimental evidences to be something like an angular momentum due to the facts that spin

- acts experimentally like something rotating in space having an intrinsic angular momentum
- has the physical dimension of an angular momentum
- couples with the normal orbital angular momentum to the resulting total angular momentum like an angular momentum
- is conserved as part of angular momentum
- exhibits a magnetic moment expected from circulating currents
- and some other things remembering on angular momentum

Since all fundamental particles have spin 1/2, this quantum phenomenon is very important. The usual description of spin 1/2 is based on the Pauli matrices, avoiding any space-like imagination. People who are used to work with fractional order calculus will ask why physicists do not try to describe the spin similar to the well-known angular momentum, but only with a further degree of freedom. Indeed, this can easily be done. Let us look on the spherical harmonics $Y_{l,m}(\vartheta, \varphi)$ given in (40) and the associated Legendre polynomials

$$P_l^{|m|}(x) = (1-x^2)^{\frac{|m|}{2}} \frac{d^{|m|}}{dx^{|m|}} P_l(x) \quad (48)$$

with

$$P_l(x) = \frac{1}{2^l l!} \frac{d^l}{dx^l} (x^2 - 1)^l \quad (49)$$

and extend them to fractional order $l = 1/2$, $m = \pm 1/2$ i.e., to evaluate the semi-derivatives in the expressions:

$$P_{1/2}^{1/2}(x) = (1 - x^2)^{\frac{1}{4}} \frac{d^{1/2}}{dx^{1/2}} P_{1/2}(x) \quad (50)$$

$$P_{1/2}(x) = \sqrt{\frac{2}{\pi}} \frac{d^{1/2}}{dx^{1/2}} \sqrt{x^2 - 1} \quad (51)$$

We can merge these two equations to:

$$P_{1/2}^{1/2}(x) = \sqrt{\frac{2}{\pi}} (1 - x^2)^{\frac{1}{4}} \frac{d^{1/2+1/2}}{dx^{1/2+1/2}} (x^2 - 1)^{1/2} \quad (52)$$

and need only to assume the existence of the fractional order derivatives, but fortunately, do not need to evaluate them by fractional calculus. Thus we obtain the two independent solutions (+ or -):

$$\tilde{\Psi}_{\pm 1/2} = Y_{1/2, \pm 1/2}(\vartheta, \varphi) = -i \frac{1}{\pi} \cot \vartheta \sqrt{\sin \vartheta} \exp\left(\pm i \frac{\varphi}{2}\right) \quad (53)$$

satisfying the operator equations:

$$\mathbf{S}^2 \tilde{\Psi}_{\pm 1/2} = -\hbar^2 \Delta_{\vartheta, \varphi} \tilde{\Psi}_{\pm 1/2} = \frac{1}{2} \left(\frac{1}{2} + 1\right) \hbar^2 \tilde{\Psi}_{\pm 1/2} \quad (54)$$

$$\mathbf{S}_z \tilde{\Psi}_{\pm 1/2} = -i\hbar \frac{\partial}{\partial \varphi} \tilde{\Psi}_{\pm 1/2} = \pm \frac{\hbar}{2} \tilde{\Psi}_{\pm 1/2} \quad (55)$$

for the spin operator $\mathbf{S}$ which is in complete analogy to the orbital angular momentum $\mathbf{L}$. It might be noticed that usage of the generalised Legendre polynomials will also yield such solutions. The imaginary factor $(-i)$ is due to the selected branches of the roots and can be omitted, because if Ψ is an eigenfunction, then $c\Psi$ with arbitrary constant c is also an eigenfunction of the differential operator.

However, the physical interpretation of the wave function asks for a normalisation. Thus, the integral of its norm $|\Psi^* \Psi|$ over the whole space has to be 1. But what is the “whole” space in our case? To answer this question we have to look on solution (53). Omitting for simplicity the unitary factor $(-i)$ we have

the phases $\pm\varphi/2$, which do not affect the magnitude of these functions, but which differ monumentally from their analogues in the orbital angular momentum functions.

(a) **4 π -Periodicity**

We immediately see, that there is a periodicity of 4π in the phase φ . This means that a spin-1/2-particle has to turn twice in the space to return in its initial state. This has intrigued physicists since the first experimental evidences of this phenomenon [12,13]. The interpretation of these experiments was taken into doubt by many physicists, but recently confirmed again with modified experiments avoiding the reasons for criticism [14]. Today, this 4π -periodicity of spin-1/2-particles becomes generally accepted. If any theorist would have introduced the fractional description of the spin just after his discovery in the twenties of the last century, his model would have been made ridiculous due to this 4π -periodicity. Perhaps, this was one of the reasons, why one did not believe in a description similar to the orbital angular momentum. Nowadays, “dynamical phases” or similar concepts [15] are proposed to save the “classical” spinor picture.

(b) **Wave function of spin-1/2-particles**

Let us now return to our question about the “whole space”. Since the norm of the wave function is the probability for the particle to occupy the volume element of the space, this space has to contain all possible configurations. This means that the “whole” space is twice our civilian space, i.e., $\mathfrak{R}^3 \otimes 2$. We have to extend the integral for φ from 0 to 4π . This means that our spin wave functions Ψ_S for spin $\pm 1/2$ have to look like:

$$\Psi_{\pm 1/2} = \frac{1}{\pi\sqrt{2}} \cot \vartheta \sqrt{\sin \vartheta} \exp\left(\pm i \frac{\varphi}{2}\right) \quad (56)$$

These functions form an orthonormal basis for all spin wave functions over the $\mathfrak{R}^3 \otimes 2$. The complete wave function of spin-1/2-particles is the product of the wave function $\Psi(\mathbf{r})$ describing the location of the particle in the $\mathfrak{R}^3$, and the spin wave function Ψ_S over the spin space $\mathfrak{R}^3 \otimes 2$

$$\Psi = \Psi(\mathbf{r})\Psi_S \quad (57)$$

(c) **Ladder operators**

The ladder operators $\mathbf{S}_{\pm}$

$$\mathbf{S}_{\pm} = \mathbf{S}_x \pm i\mathbf{S}_y \quad (58)$$

create (+) or annihilate (−) a magnetic spin quantum according to

$$\mathbf{S}_z(\mathbf{S}_{\pm}\Psi_S) = \hbar(m_s \pm 1)\mathbf{S}_{\pm}\Psi_S \quad (59)$$

but maintains the magnitude s of the spin

$$\mathbf{S}^2(\mathbf{S}_{\pm}\Psi_S) = \hbar^2 s(s+1)\mathbf{S}_{\pm}\Psi_S \quad (60)$$

which can be proved analogue to the proof for $\mathbf{L}_{\pm}$ previously given.

(d) Interpretation of the spin

The nice aspect of the fractional description of spin-1/2-particles is beside the direct evidence of their 4π -periodicity the possibility of an interpretation in space. In this picture, the spin-up and spin-down states differ only by the sign of the exponential term, which is also the sole complex part of the spin wave function yielding the phase $\phi = \pm\varphi/2$. In quantum mechanics, the “mean velocity” of a particle with the mass m_p is proportional to the gradient of the phase ϕ of its wave function:

$$\mathbf{v} = \frac{\hbar}{m_p} \nabla \phi \quad (61)$$

Applying this to the spin wave functions (56) we get in polar coordinates:

$$\mathbf{v} = \frac{\hbar}{m_p r \sin \vartheta} \{0, 0, \pm 1/2\} \quad (62)$$

which tells us that we have a rotation around the chosen z -axes. Spin $+1/2$ corresponds to a right-handed rotation, spin $-1/2$ to a left-handed rotation.

$$\mathbf{L}_z \Psi_L = L_z \Psi_L \quad (63)$$

Since this fractional description of spin-1/2-particles allows its interpretation in the real space, it is complementary to Pauli’s spinor picture. It has to be proven if a rigorous application of this fractional description of the spin can yield all the other observable results like the standard spinor description.

5 Conclusions

These examples show that semi-integrals and semi-derivatives are appropriate to describe natural phenomena. Today, the application of semi-integrals in

connection with Abel-type integral equations pervades all natural and technical sciences, as well as modern medicine. The fractional description of spin-1/2-particles can perhaps contribute to enlighten the mystery of the spin.

References

1. Oldham KB, Spanier J (1974) *The Fractional Calculus*. Academic Press, New York.
2. Abel NH (1823) Solution de quelques problèmes à l'aide d'intégrales définies. Mag. Naturvidenskaberne.
3. Abel NH (1826) Auflösung einer mechanischen Aufgabe. J. für die Reine Angew. Math, 1:153–157.
4. Laurent MH (1884) Sur le calcul des dérivées à indices quelconques. Nouv. Ann. Math. 3(3):240–252.
5. Krempl PW (1974) The Abel-type integral transformation with the kernel $(t^2 - x^2)^{-1/2}$ and its application to density distributions of particle beams. CERN MPS/Int.BR/74-1, pp. 1–31.
6. Krempl PW (2005) Some Applications of Semi-Derivatives and Semi-Integrals in Physics. Proc. ENOC 05, Eindhoven, ID 11-363, 10 pp.
7. Deans SR (1996) Radon and Abel Transforms. in Poularikas AD (ed.), *The Transforms and Applications Handbook*. CRC Press, Boca Raton, pp. 631–717.
8. Yuan Z-G (2003) The Filtered Abel Transform and its Application in Combustion Diagnostics, NASA/CR-2003-212121, pp 1–11.
9. BruckH (1966) *Accélérateurs circulaires de particules*. Press Universitaires de France, Paris.
10. Krempl PW (1974) TMIBS—un programme pour le calcul de la densité projetée à partir des mesures effectuées avec les cibles. CERN MPS/BR Note 74-16, pp. 1–9.
11. Krempl PW (1975) Beamscope. CERN PSB/Machine Experiment News 126b.
12. Rauch H, Zeilinger A, Badurek G, Wilfing A (1975) Verification of coherent spinor rotation of fermions. Phys. Lett., 54A(6):425–427.
13. Werner SA, Colella R, Overhauser AW, Eagen CF (1975). Observation of the phase shift of a neutron due to precession in a magnetic field. Phys. Rev. Lett., 35(16):1053–1055.
14. Ioffe A, Mezei F (2001) 4π -symmetry of the neutron wave function under space rotation, Physica B, 297:303–306.
15. Hasegawa Y, Badurek G (1999) Noncommuting spinor rotation due to balanced geometrical and dynamical phases. Phys. Rev. A, 59(3):4614–4622.

Part 2

Classical Mechanics and Particle Physics

MESOSCOPIC FRACTIONAL KINETIC EQUATIONS VERSUS A RIEMANN–LIOUVILLE INTEGRAL TYPE

Raoul R. Nigmatullin¹ and Juan J. Trujillo²

¹ Theoretical Physics Department, Kazan State University, Kremlevskaya 15, 420008, Kazan, Tatarstan, Russian Federation; E-mail: nigmat@knet.ru

² Departamento de Análisis Matemático, University of La Laguna, 38271, La Laguna, Tenerife, Spain; E-mail: JTrujill@ull.es

Abstract

It is proved that kinetic equations containing noninteger integrals and derivatives are appeared in the result of reduction of a set of micromotions to some *averaged* collective motion in the mesoscale region. In other words, it means that after a proper statistical average the microscopic dynamics is converted into a collective complex dynamics in the mesoscopic regime. A fractal medium containing strongly correlated relaxation units has been considered. It is shown that in the most cases the original of the memory function recovers the Riemann–Liouville fractional integral. For a strongly correlated fractal medium a *generalization* of the Riemann–Liouville fractional integral is obtained. For the fractal-branched processes one can derive the stretched exponential law of relaxation that is widely used for description of relaxation phenomena in disordered media. It is shown that the generalized stretched-exponential function describes the averaged collective motion in the fractal-branched complex systems. The application of the fractional kinetic equations for description of the dielectric relaxation phenomena is also discussed. These kinetic equations containing non integer integral and derivatives with real and complex exponents and their possible generalizations can be applicable for description of different relaxation or diffusion processes in the intermediate (mesoscale) region.

Keywords

Generalized Riemann–Liouville fractional integral, universal decoupling procedure.

1 Introduction

Recently much attention has been paid to existence of equations containing non integer operators with *real* fractional exponent [1–7]. But in papers related to consideration of the fractional equations containing noninteger operators of integration or differentiation are realized on an “intuitive” level in the form of some postulates/suppositions *imposed* on a structure or model considered. At the present time a systematic deduction of kinetic equations containing noninteger integration/differentiation operators based on the given structure of a disordered medium with the usage of the modern methods of nonequilibrium statistical mechanics is *absent*. So, there is a barest necessity to derive kinetic equations with noninteger operators of differentiation and integration from the first principles of the statistical mechanics, based on the consideration of an infinite chain of equations for a set of correlation functions. It becomes evident that equations with fractional derivatives can play a crucial role in description of kinetic and transfer phenomena in the mesoscale region. From our point of view this necessary mathematical instrument should lie in deep understating of the “physics” of the fractional calculus.

In present time the interest in application of the mathematical apparatus of the fractional calculus in different branches of techniques and natural sciences is considerably increased. Here one can remind the applications of the fractional calculus in

1. Fractional control of engineering systems.
2. Advancement of calculus of variations and optimal control to fractional dynamic systems.
3. Analytical and numerical tools and techniques.
4. Fundamental explorations of the mechanical, electrical, and thermal constitutive relations and other properties of various engineering materials such as viscoelastic polymers, foam, gel, and animal tissues, and their engineering and scientific applications.
5. Fundamental understanding of wave and diffusion phenomenon, their measurements and verifications.
6. Bioengineering and biomedical applications

Detailed references can be found in the recent review, in the proceedings of the conference and in papers [2, 4, 8–14].

The first attempt to understand the result of averaging of a smooth function over the given fractal (Cantor) set has been undertaken in [15]. In the note and later in paper some doubts were raised to the reliability of the previously obtained result [15–17]. The criticism expressed in these publications forced one of the authors of this paper (RRN) to reconsider the former result, and the detailed study of this problem showed that the doubts had some grounds and were directly linked with the relatively delicate procedure of averaging a smooth function over fractal sets, in particular, on Cantor set and its generalizations.

In order to dissipate these doubts and realize mathematically correct averaging procedure over fractal sets it was necessary to carry out a special study. Complete investigation has been given in the book [18], where the correct averaging procedure was considered in detail. One can prove that the previous result [15] is correct for *random* fractals, for regular fractals the procedure of averaging of a smooth function over fractal sets leads to the memory function expressed in terms of the Riemann–Liouville integral containing the *complex* power-law exponent. The further generalization for the modified Cantor sets has been realized in papers of Prof. Fu-Yao Ren with coauthors [19–21]. Another approach leading to the fractional integral and related to coarse graining time averaging is considered by R. Hilfer in the recent book [1]. Independent analysis of above-cited papers could lead to a conclusion that the physical meaning of the fractional integral with *real* exponent has been understood. Temporal fractional integral can be interpreted as a conservation of part of states localized on a self-similar (fractal) object if the physical system considered has at least two parts of different states. One part is distributed *inside* of a fractal set (the conserved part of states) and another part of states is located *outside* of the fractal set (the lost part of states). That’s why it is easy to understand the fractional integral of one-half order, when for its understanding the consideration of a fractal object is *not* necessary. Half of states are lost automatically in diffusion process with semi-infinite boundary conditions [22]. From the geometrical point of view the temporal fractional integral is associated with Cantor set or its generalizations, occupying an intermediate position between the classical Euclidean point and continuous line. But the meaning of fractional integral with real fractional exponent is *not* complete in the light of papers [8, 23–26], where the correct understanding of different self-similar objects with the *complex* fractal dimensions is discussed. These interesting ideas forced one of the authors of this paper (RRN) to reconsider their previous results obtained in [18] and give a possibility to understand the geometrical/physical meaning of mathematical operator with the *complex* fractional exponent [4]. So the basic question, which we are going to solve and discuss in this paper, can be formulated as follows:

Is it possible to suggest a “universal” decoupling procedure for a memory function in order to obtain noninteger operator with real or complex exponent in the mesoscale region from a kinetic equation with memory?

We are going to show that details of the averaging procedure developed in [18] will help us to find the proper answer for the question formulated above.

The following content of this article obeys the next structure. In the section 2 we present the basic equations of statistical mechanics containing a memory function.

In section 3 we derive the memory function for a strongly correlated fractal medium. In this section we show also *how* it is possible to generalize the Riemann–Liouville integral. The general solutions containing log-periodic function help to understand the geometrical/physical meaning of noninteger operator containing an imaginary part of the complex fractional exponent.

In sections 4 and 5 we consider the basic mathematical properties of the generalized RL integral. In section 6 we determine the strong-correlated fractal-branched structures, which, in turn, help to derive the stretched-exponential law of relaxation. The basic results are summarized in final section 7.

2 Different Kinetic Equations with Memory

In many branches of physics the relationship between physical values are related by means of memory function. For example, in the theory of linear response the deviation from the mean physical value evoked by the applied external field can be expressed as [27]

$$\Delta A(t) = \int_{-\infty}^t K(t-t_1) F(t_1) dt_1. \quad (1)$$

Here $\Delta A(t) = \langle A(t) \rangle - \langle A \rangle_0$. The kernel $K(t) = i \langle [A(t), B]_{\pm} \rangle$ with $[A(t), B]_{\pm} = A(t)B \pm BA(t)$ defines the corresponding correlation function, the value $F(t_1)$ defines an amplitude of the external field, entering into the perturbation Hamiltonian

$$H_1(t) = BF(t). \quad (2)$$

Here B corresponds to a quantum-mechanical operator, which determines the interaction of the many-body Hamiltonian with external field. For example, in the case of interaction with electric field B coincides with the operator of total polarization P , for magnetic field the operator B corresponds to the magnetization operator M and etc.

If one can try to consider the dynamics of the system in the Mori–Zwanzig formalism [28] then the relationship between the autocorrelation function of the second order $M_1(t)$ with correlation functions of higher orders can be written as

$$\frac{dM_1(t)}{dt} = \lambda_1 M_1(t) + \Omega_1 \int_0^t K(t-\tau) M_1(\tau) d\tau. \quad (3)$$

Here λ_1 and Ω_1 are some characteristic parameters, $K(t)$ is the correlation function of the next order, which plays a role of a memory function for the initial correlation function $M_1(t)$. An analog of Eq. (3) with specific memory function and similar variant of this equation was derived in papers [29].

Based on the Zubarev kinetic formalism one can derive the diffusion equation with memory [27,30,31]

$$\frac{\partial n(r, t)}{\partial t} = \int_0^t \sum_{\alpha=x,y,z} K_{\alpha\alpha}(t-t_1) \nabla_{\alpha}^2 n(r, t), \quad (4)$$

where

$$K_{\alpha\alpha}(t) = \int d^3r r_{\alpha}^2 W(r, t), \quad K_{\alpha\beta} = 0 \quad (\alpha \neq \beta), \quad (5)$$

and $n(r, t)$ coincides with local density of electric dipoles or spins.

Other equations with memory in the framework of the generalized Zubarev formalism are considered in [32]. One can notice that the basic kinetic Eq. (1), (3), and (4) imply some decoupling procedure related to the further calculation of the memory function $K(t)$. Usual decoupling formalism based on integer derivatives *cannot* lead to the noninteger integration/differentiation operators. In order to derive possible “fractional” kinetic equations it is necessary to impose some *assumption* related with the calculation of a memory function. We are going to suggest rather general decoupling procedure for calculation of the memory function based on a self-similar structure of the medium considered.

3 Memory Function for a Strongly Correlated Fractal Medium

In this paper we want to derive a structure of a kernel $K(t)$ for a strongly correlated fractal medium. The structure of the kernel for a *weakly correlated* fractal medium has been considered recently in [33]. The special procedure for recognition of the “fractional” kinetics from dielectric spectroscopy data has been suggested in papers [34–36]. For further purpose it is useful to use the Laplace/Fourier transform applied to Eq. (1), (3), and (4) in order to have a possibility to consider the kernel $K(t)$ separately.

For the case of strongly correlated clusters one can *suppose* that the memory function forms a self-similar structure combining these clusters in the form of a *product*. Such formation is possible, for example, in the case of percolation phenomenon

$$K(s) = \prod_n N_n f(s\tau_n). \quad (6)$$

For the case of strongly correlated dipoles/relaxation units satisfying to the suppositions made in the previous section 3 the Laplace image of the memory function takes the form

$$K(z) \equiv P(z) = \prod_{n=-(N-1)}^{N-1} b^n f(z\xi^n) \equiv \prod_{n=-(N-1)}^{N-1} f(z\xi^n), \quad (N \gg 1). \quad (7)$$

In this section we are going to show that evaluation of expression (7) does not depend on the concrete form of the microscopic function $f(z)$. One can notice that the product satisfies to the following *exact* equation

$$P(z\xi) = \frac{f(z\xi^N)}{f(z\xi^{-N+1})} P(z). \quad (8)$$

Without loss of a generality we suppose that the Laplace-image of the function $f(z)$ depending on a complex variable z (the variable z defines the dimensionless Laplace parameter $z = s\tau_0$ with respect to some characteristic time τ_0) and describing the microscopic act of interaction of an electric dipole with thermostat has the following form:

$$f(z) = \frac{a_0 + a_1 z + \dots + a_K z^K}{b_0 + b_1 z + \dots + b_P z^P}, \quad (9)$$

with $K \leq P + 1$ and the polynomial in denominator has only negative and complex-conjugated roots.

Case (a): $\text{Re}(z) \ll 1$ ($c_0 = a_0/b_0$, $c_1 = a_1/b_0 - b_1/b_0^2$)

$$f(z) = c_0 - c_1 z + c_2 z^2 + \dots \quad (10a)$$

Case (b) $\text{Re}(z) \gg 1$ ($A_1 = a_K/b_P$, $A_2 = a_{K-1}b_P - a_K b_{P-1}/b_P^2$)

$$f(z) = \frac{A_1}{z^{P-K}} + \frac{A_2}{z^{P-K+1}} + \dots \quad (10b)$$

For $K = P + 1$ we define $f(z)$ as a *relaxation function* describing the process of interaction of a dipole with a thermostat. If $K < P + 1$ and the denominator of polynomial (9) has divisible roots then we define $f(z)$ as an *exchange function* describing the interaction process of a dipole with thermostat. The reason for such division is that the minimum value of the function $f(t) \stackrel{LT}{=} f(s)$ (we are using the same notation f for original $f(t)$ and Laplace image $f(s)$ and suppose that it does not evoke further misunderstanding) in the first case is $f(t=0) \neq 0$ and moreover $f(t)$ tends to zero as $t \rightarrow \infty$ monotonically. In the exchange case, however, the value $F(t=0) = 0$ and the microscopic function $f(t)$ has at least one maximum and so may tend to zero monotonically or nonmonotonically as $t \rightarrow \infty$.

For $K = P$ the process of interaction with thermostat has delta-like function $\delta(t)$ collision character.

Taking into account the asymptotic decompositions (10) at $\xi < 1$ in the limit $N \rightarrow \infty$ the last relationship for the fixed N is reduced to the scaling functional equation of the type

$$P(z\xi) = \frac{a_0}{b_0} \frac{\tilde{z}^{P-K}}{A_1} P(z), \quad (11)$$

where $\tilde{z} = z\xi^{-N+1}$.

This scaling equation is valid for the interval of complex variable z satisfying to conditions:

$$z_{\min} \ll |z| \ll z_{Mx}. \quad (12)$$

Here

$$z_{\min} = \left[\left| \frac{a_{K-1}b_P - a_K b_{P-1}}{b_P^2} \right| \right]^{\frac{1}{P-K+1}} \xi^{N-1}, \quad z_{Mx} = \frac{b_0^2 \xi^{-N}}{|a_1 b_0 - a_0 b_1|}. \quad (13)$$

The limiting parameters entering into the last expression are defined by expression (9).

At $P = K$ that physically corresponds the $\delta(t)$ -like collisions with thermostat the solution of the last functional Eq. (11) has the form

$$P(z) = \pi_v (\ln z) z^{-v}, \quad v = \frac{\ln(1/b)}{\ln(1/\xi)}, \quad b \equiv \frac{b_0 A_1}{A_0}. \quad (14)$$

For $P \neq K$ one can obtain the general solution of the scaling equation by the method of a free constant variation. Taking the natural logarithm from the both part of Eq. (11) we have

$$X(z\xi) = X(z) + (P - K) \ln(z) - \ln(\tilde{b}). \quad (15)$$

Here $X(z) = \ln(P(z))$, the constant $\tilde{b}$, which is supposed to be finite is defined by expression

$$\tilde{b} = b \xi^{(P-K)(N-1)} \equiv \frac{b_0 A_1}{a_0} \xi^{(P-K)(N-1)}. \quad (16)$$

The solution of the functional Eq. (16) we are presenting in the form

$$X(z) = \pi_0 (\ln(z)) + C_1 \ln^2(z) + C_2 \ln(z). \quad (17)$$

Here $\pi_0(\ln z \pm \ln \xi) = \pi_0(\ln z)$ is a complex log-periodic function with real period $\ln(\xi)$, $C_{1,2}$ are free variation constants, which are determined from Eq. (15). Putting solution (17) into (15) we have

$$2C_1 \ln(\xi) = P - K, \quad C_1 \ln^2(\xi) + C_2 \ln(\xi) = -\ln(\tilde{b}). \quad (18)$$

From Eq. (18) one can obtain for $\xi < 1$

$$C_1 = -\frac{P - K}{2 \ln(1/\xi)}, \quad C_2 = -\frac{\ln(1/\tilde{b})}{\ln(1/\xi)} - \frac{P - K}{2}. \quad (19)$$

The general solution for the product $P(z)$ can be written finally as

$$P(z) = z^{-v} \exp(\pi_0 \ln(z)) \exp(-\gamma \ln^2(z)) \equiv R(\ln(z)) z^{-v} \exp(-\gamma \ln^2(z)). \quad (20)$$

Here the power-law exponent v and the positive damping constant γ are defined correspondingly by expressions

$$v = \frac{\ln(1/\tilde{b})}{\ln(1/\xi)} + \frac{P - K}{2} = \frac{\ln\left(\frac{a_0}{b_0 A_1}\right)}{\ln\left(\frac{1}{\xi}\right)} + (P - K) \left(N - \frac{1}{2}\right), \quad \gamma = \frac{P - K}{2 \ln(1/\xi)}. \quad (21)$$

The original taken from expression (20) (as before $z = s\tau_0$ is the dimensionless complex variable)

$$K(t) = \frac{1}{2\pi\tau_0 j} \int_{a-j\infty}^{a+j\infty} P(z) \exp(zt / \tau_0) dz \quad (22)$$

generalizes the conventional definition [37] of the Riemann–Liouville integral for the values $\gamma \neq 0$. The properties of the kernel (22) need a special mathematical examination and considered in the next section. For $b > 1$ it is possible to check by direct calculations that the solution for $P(z)$ can be written in the same form (20) with constants determined by expression (21).

4 The Analytical Form of the Kernel $K(t)$

In this section it is convenient to give another definition of the generalized fractional integral operator (GFIO) $(K^{\nu,\gamma} f)(t)$, which depends on two parameters and acts on a smooth and arbitrary function $f(t)$ as:

$$(K^{\nu,\gamma} f)(t) \equiv \left[L^{-1} \left(e^{-[\nu \ln(s) + \gamma (\ln^2(s))]} \right) \right] (t) * f(t), \quad (23)$$

where the asterisk

$$f(t) * g(t) = \int_0^t f(t-x)g(x)dx, \quad (24)$$

determines the Laplace convolution operation. This GFIO can be written in the following implicit form in time domain:

$$(K^{\nu,\gamma} f)(t) = Q^{\nu,\gamma}(t) * f(t) = \int_0^t Q^{\nu,\gamma}(t-x)f(x)dx \quad (25)$$

where $Q^{\nu,\gamma}(t)$ is the solution of the following Volterra equation

$$tQ^{\nu,\gamma}(t) = [(\nu - 2\gamma E - 2\gamma \ln(t)) * Q^{\nu,\gamma}(t)]. \quad (26)$$

This integral equation follows from the conditions:

$$\begin{aligned} \text{(a)} \quad & \int_0^\infty e^{-u} Q(u) du = 1. \\ \text{(b)} \quad & \int_0^\infty Q(u) du = 0 \end{aligned}$$

and differential equation for the function $Y(s) = \exp[-\nu \ln(s) - \gamma \ln^2(s)]$ that it is obtained easily in s -complex plane

$$\frac{dY(s)}{ds} + \left(\frac{\nu}{s} + 2\gamma \frac{\ln(s)}{s} \right) Y(s) = 0. \quad (27)$$

Taking into account the relationships [38],

$$\begin{aligned} \frac{\ln(s)}{s} & \stackrel{LT}{=} -E - \ln(t), \\ -tQ(t) & \stackrel{LT}{=} \frac{dY(s)}{ds} \end{aligned} \quad (28)$$

($E = 0.5772156649\dots$ is the Euler constant) and initial conditions $Y(0) = 0$ or $Y(1) = 1$ one can obtain from (27) the desired Volterra equation (26).

5 Basic Properties of the Generalized Riemann–Liouville Integral

The new fractional Riemann–Liouville type operator which we have introduced above $K^{\nu, \gamma}$, have the following properties, which are very easy to proof, when it is used over suitable functions:

1. $K^{\nu, 0} = I_{0+}^{\nu}$. It recovers the fractional Riemann–Liouville integral operator.
2. $K(t) = \frac{1}{\tau_0} (K^{\nu, \gamma} g)(t / \tau_0)$, where $g(t) = L^{-1}(R(\ln(s)))(t)$.
3. $K^{\nu_1, \gamma_1} K^{\nu_2, \gamma_2} = K^{\nu_2, \gamma_2} K^{\nu_1, \gamma_1} = K^{\nu_1 + \nu_2, \gamma_1 + \gamma_2}$. This is the index law for our fractional operator, which proof hold from the definition (23).
4. $K^{0, 0} = I$, where I is the identity operator.
5. $D^{\nu} K^{\nu, \gamma} = K^{0, \gamma}$, where $D^{\nu} = D_{0+}^{\nu}$ is the fractional Riemann–Liouville derivative operator.

6 Consideration of Relaxation Processes in the Fractal-Branched Structures

Let us suppose that the Laplace-image of the initial memory function $K(t)$ has more general form

$$P(z) = \prod_{n=0}^{N-1} [f(z\xi^n)]^{n_0 b^n} = \exp \left[n_0 \sum_{n=0}^{N-1} b^n \ln [f(z\xi^n)] \right] \equiv \exp [n_0 S_N(z)]. \quad (29)$$

The physical meaning of this function is related to the fact that microscopic relaxation function $f(z)$ has *additional* branching in the self-similar volumes $V_n \propto n_0 b^n$. The variable z in (29) can coincide with a dimensionless frequency variable or a temporal variable t , respectively. The evaluation of the last expression depends essentially on the asymptotic behavior of the function $f(z)$ and from the interval of location of the scaling parameters ξ and b . We suppose that these scaling parameters satisfy to the following inequality

$$0 < \xi \leq \frac{1}{b} < 1. \quad (30)$$

By analogy with expression (11) one can show that the sum $S_N(z)$ figuring in expression (29) satisfies to the relationship

$$S_N(z\xi) = \frac{1}{b} S_N(z) + b^{N-1} \ln \left[f(z\xi^N) \right] - \frac{1}{b} \ln \left[f(z) \right]. \quad (31)$$

We suppose that the asymptotic behavior of the function $f(z)$ obeys the following decompositions:

at $|z| \ll 1$

$$f(z) = 1 + c_1 z + c_2 z^2 + \dots, \quad (32a)$$

at $|z| \gg 1$

$$f(z) = \frac{A}{z^\beta} \exp(-rz) + \frac{A_1}{z^{1+\beta}} \exp(-2rz) + \dots \quad (32b)$$

In the last relationship we combined the exponential and power-law asymptotic in order to consider them together. At $\beta = P - K \neq 0$ and $r = 0$ it describes the power-law asymptotic similar to (8); the case $\beta = 0$ and $r \neq 0$ corresponds to the exponential asymptotic. In the limit $N \rightarrow \infty$ one can obtain the following scaling equation for the limiting value of the sum $S(z)$

$$S(z\xi) = \frac{1}{b} S(z) + \left(c_1 + \frac{r}{b} \right) z + \frac{\beta}{b} \ln(z) - \frac{1}{b} \ln(A). \quad (33)$$

The limits of the intermediate asymptotic are determined from inequalities:

$$c_2 |z|^2 \ll 1, \quad A_1 \exp \left[-2r|z| - (1+\beta) \ln(|z|) \right] \ll 1. \quad (34)$$

Similar to calculations realized above the scaling equation (33) can be solved analytically based again on the method of variation of a free constant. We are giving the final solutions for the memory function $P(z)$ related with $S(z)$ by relationship (29).

The case ($b \neq 1$, $b > 1$, $\xi < 1$):

$$P(z) = P_0 z^\mu \exp \left[z^\nu \sigma_\nu (\ln(z) + \lambda z) \right]. \quad (35)$$

Here $\sigma_\nu (\ln(z) \pm \ln(\xi)) = \sigma_\nu (\ln(z))$ is a log-periodic oscillating function with period $\ln(\xi)$. In the one-mode approximation (OMA) this function can be presented in the form

$$\sigma_\nu (\ln(z)) = -\gamma(\nu) + \gamma_1 \exp(j \langle \Omega \rangle \ln(z)) + \gamma_1^* \exp(-j \langle \Omega \rangle \ln(z)). \quad (36)$$

Here the zero Fourier-component should accept the negative values. Other parameters in (35) are defined by expressions

$$P_0 = \exp(-S_0), \quad S_0 = n_0 \left(\frac{\ln(A)}{b-1} + \frac{\beta b \ln(1/\xi)}{(b-1)^2} \right), \quad \nu = \frac{n_0 \ln(b)}{\ln(1/\xi)}, \quad (37)$$

$$\mu = \frac{n_0 \beta}{b-1}, \quad \lambda = -\frac{n_0 (r + bc_1)}{1 - \beta \xi}.$$

The further investigations show that for the case $b < 1$, $\xi < 1$ the value of the constant $c_1 = 0$ in expansion (32a); the power-law exponents μ and ν can be

simultaneously negative satisfying to condition $\mu\nu > 0$. The damping constant λ can accept *positive* or negative values.

The case ($b = 1$, $\xi < 1$) generalizes expression (20) obtained above

$$P(z) = R(\ln(z)) \exp \left[-\gamma \left(\ln^2(z) \right) + \nu \ln(z) - \lambda z \right], \quad (38)$$

with parameters

$$\gamma = \frac{n_0 \beta}{2 \ln(1/\xi)}, \quad \nu = n_0 \left(\frac{\ln(A)}{\ln(1/\xi)} + \frac{\beta}{2} \right), \quad \lambda = n_0 \frac{r + c_1}{1 - \xi}. \quad (39)$$

Expression (38) generalizes the well-known Kohlrausch–Williams–Watts relaxation law suggested many years ago for description of nonexponential relaxation phenomena in many disordered systems [39, 40]. As before, we discovered and mathematically confirmed the *reduction* phenomenon, when a set of micromotions is averaged and transformed again into a collective motion. It is interesting to note that different partial cases (for some concrete forms of $f(t)$, describing different types of micromotions) leading to the “pure” stretched-exponential dependence in time domain has been considered by many authors in section 8 published in the Proceedings of the International Symposium [41]. These nonexponential functions have been applied for description of relaxation phenomena of statistical defects in condensed media, in glasses etc.

7 Results and Discussions

Approach developed in this paper helps to understand the general decoupling procedure applied to a memory function that can lead to equations containing non-integer integrals and derivatives with real or complex power-law exponents. These equations naturally explain temporal irreversibility phenomena which can be appeared in linear systems with “*remnant*” memory. For linear systems the so-called “partial” irreversibility is appeared in the result of reduction procedure, when a many-body system lost many microscopic states and only part of states in the form of collective motions are conserved on the following level of intermediate scales and expressed in the form of the fractional integral. This approach opens new possibilities for analysis of different kinetic equations with remnant memory, when the RL-operators can be modified by a damping constant defined by (21) or new convolution term appearing in the Laplace image appearing in (20). The temperature dependence of the power-law exponents ν and $\langle \Omega \rangle$, which can enter into the corresponding kinetic equation merits a special examination.

References

1. Hilfer R (ed.) (2000) *Applications of Fractional Calculus in Physics*. World Scientific, Singapore.
2. Zaslavsky GM (2002) Phys. Rep., 371:461.
3. Kupferman R (2002) J. Stat. Phys., 114:291.
4. Nigmatullin RR, Le Mehaute A (2005) J. Non-Crystalline Solids, 351:28–88.
5. Orsingher E, Beghin L (2004) Probab. Theory Relat. Fields, 128:141.
6. Del-Castillo-Negrete D, Carreras BA, Lynch VE (2003) Phys. Rev. Lett., 91(1):018302–1.
7. Zaburdaev VYu, Chukbar KV (2003) JETP Lett., 77:551.
8. Sornette D (1998) Phys. Rep., 297:239.
9. Agrawal OP (2002) Nonlinear Dynamics, 29:145; Agrawal OP (2002) Trans. ASME J. Vibration Acoustics, 124:454.
10. Narahari Achar BN, Hanneken John W, Clarke T (2004) Physica A, 339:311.
11. Deng R, Davies P, Bajaj AK (2003) J. Sound Vibration, 262/3:391.
12. Schmidt VH, Bohannan G, Arbogast D, Tuthill G (2000) J. Phys. Chem. Solids, 61:283.
13. Manabe S (2002) Nonlinear Dynamics, 29:251.
14. Battaglia J-L, Le Lay L, Bastale J-C, Oustaloup A, Cois O (2000) Int. J. Thermal Sci., 39:374.
15. Nigmatullin RR (1992) Theor. Math Phys., 90(3):354.
16. Rutman RS (1994) Theor. Math. Phys., 100(3):476 (in Russian).
17. Rutman RS (1995) Theor. Math. Phys., 105(3):393 (in Russian).
18. Le Mehaute A, Nigmatullin RR, Nivanen L (1998) *Fleches du Temps et Geometrie Fractale*. Hermès, Paris (in French).
19. Fu-Yao Ren, Zu-Guo Yu, Feng Su (1996) Phys. Lett. A, 219:59.
20. Zu-Guo Yu, Fu-Yao Ren, Ji Zhou (1997) J. Phys. A. Math. Gen. 30:55–69.
21. Fu-Yao Ren, Jin-Rong Liang (2000) Physica A, 286:45.
22. Babenko Yu I (1986) Heat and Mass Transfer. The method of calculation of heat and diffusion currents. Leningrad, Chemistry (in Russian).
23. Johansen A, Sornette D, Hansen AE (2000) Physica D, 138:302.
24. Johansen A, Sornette D (1998) Int. J. Mod. Phys., 9:433.
25. Stauffer D, Sornette D (1998) Physica A, 252:271.
26. Huang Y, Quillon G, Saleur H, Sornette D (1997) Phys. Rev E., 55(6):6433.
27. Zubarev DN (1971) *Non-Equilibrium Statistical Thermodynamic*, Nauka, Moscow (in Russian).
28. Mori H (1963) Progr. Theor. Phys. 30:578; Zwanzig R (1961) in: Brittin WE, Downs BW, Downs J, *Lectures in Theoretical Physics*, Interscience Publications, New York, Vol. III, pp. 106–141; Yulmetuev RM, Khusnutdinov NR (1994) J. Phys. A, 27:53–63; Shurygin VYu, Yulmetuev RM (1989) Zh. Eks. Theor. Fiz., 96:938 (Sov. JETP (1989) 69:532 and references therein).

29. Montroll EW, Shlesinger MF (1984) in: Lebowitz J, Montroll E (eds.) *Studies in Statistical Mechanics* Vol. 11, p. 1.; Saichev AI, Zaslavsky, GM (1997) *Chaos*, 7:753.
30. Zubarev DN, Morozov VG (1983) *Physica A.*, 120:11.
31. Nigmatullin RR (1984) *Phys. Stat. Sol. (b)* 123:739.
32. Nigmatullin RR, Tayurski DA (1991) *Physica A*, 175:275.
33. Nigmatullin RR (2005) *Physica B.: Phys. Condens. Matter*, 358:201.
34. Nigmatullin RR, Osokin SI (2003) *J. Signal Proc.*, 83:2433.
35. Nigmatullin RR, Osokin SI, Smith G (2003) *J. Phys D: Appl. Phys.* 36:2281.
36. Nigmatullin RR, Osokin SI, Smith G (2003) *Phys. C.: Condens. Matter* 15:3481.
37. Oldham K, Spanier J. (1974) *The Fractional Calculus*. Academic Press, New York.
38. Abramowitz M, Stegun I (1964) *Handbook of Mathematical Functions*, National Bureau of Standards, Applied Mathematics Series Vol. 55.
39. Kohlrausch R (1847) *Ann. Phys.*, 12:393.
40. Williams G, Watts D (1970) *Trans. Faraday Soc.*, 66:80.
41. *Fractals in Physics* (1985) *Proceedings of the 6th International Symposium*, Trieste, Italy, 9–12 July, Pietronero L, Tozatti E (eds.), Elsevier Science.

Part 3

Diffusive Systems

ENHANCED TRACER DIFFUSION IN POROUS MEDIA WITH AN IMPERMEABLE BOUNDARY

N. Krepyшева¹, L. Di Pietro², and M. C. Néel³

¹ Present address: Monogeskhnuru 78 A-122, Nizhny, Russia; E-mail: muguet@fromru.com

² UMR Climat, Sol et Environnement INRA-UAPV, Domaine St Paul, Site Agroparc, 84914 Avignon cedex 20, France; E-mail: lili@avignon.inra.fr

³ UMR Climat, Sol et Environnement INRA-UAPV, 74 rue Louis Pasteur, 84000 Avignon, France; E-mail: marie-christine-neel@univ-avignon.fr

Abstract

We show that the space-fractional advection-diffusion equation arising from the large-scale dynamics of Lévy walks has to be modified when a boundary condition is imposed. In particular, we study the case of a reflective barrier constraining the diffusing particles to a semi-infinite domain. We obtain a modified kernel for the Riesz–Feller derivative with respect to the corresponding operator in an infinite medium.

Keywords

Superdiffusion, Space fractional equation, Reflective boundary, Lévy walks.

1 Introduction

Fick’s law is extensively used as a model for describing tracer diffusion in porous media. In heterogeneous soils, the evolution of tracer concentration usually shows anomalous super-diffusion, and Fick’s law fails to adequately describe the observed concentration profiles [1–5].

Recently, models involving space-fractional derivatives [6–8] have been proposed to better describe the spreading of a solute dissolved in a filtrating fluid.

It has been shown [7, 9] that space-fractional diffusion equations arise as the limiting dynamics of continuous time random walks (CTRW) with long-range spatial correlations decaying as a power law. The underlying jump length probability distribution is assumed to be a symmetric α -stable Lévy law, while the mean waiting time is finite (similarly to Brownian motion), so that the process is Markovian [10, 11].

When considering a moving fluid with constant mean velocity, the space-fractional diffusion equation involves an advection or drift term [12].

Both analytical [8] and numerical solutions [13, 14] are available for the space-fractional diffusion equation subject to a Dirac-pulse initial condition for an infinite medium with and without a constant drift.

Boundary conditions need to be considered for experimental applications. Since Lévy statistics correspond, on the macroscopic level, to non-local (in space) operators accounting for long-range interactions, the introduction of boundary conditions is not straightforward [15]. Non-locality, in general, implies a coupling between the boundary condition and the fractional differential equation itself [11]. The case of an absorbing boundary was considered by [16], who obtained an expression for the propagator of Lévy flights in a quiescent medium.

In the present contribution, we model the influence of an impermeable boundary within the framework of space-fractional partial differential equations. We derive the macroscopic dynamics of a CTRW, based upon symmetric Lévy flights, in a semi-infinite medium with a reflecting boundary with and without an advective term. In the macroscopic limit, we obtain a space-fractional equation, which involves a modified non-local Riesz–Feller derivative, whose kernel incorporates the boundary condition. We further find a numerical solution of the obtained fractional model by modifying the numerical scheme that discretizes the space-fractional equation in an infinite domain [14].

2 A CTRW Model with an Impermeable Boundary

We consider particles performing a CTRW with independent jump lengths and waiting time probability densities, respectively denoted by $\varphi_1(x)$, and $\varphi_2(t)$. We assume that the disordered motions of the particles on the small scale can be represented by quasi-instantaneous Lévy flights, so $\varphi_1(x)$ is a symmetric α -stable Lévy’s law with $\alpha \in (1, 2]$. The waiting time p.d.f $\varphi_2(t)$ is assumed to follow a Poisson law with mean t_0 .

In an infinite statistically homogeneous medium, the transition probability density for a particle which is in x' at time t' to jump from x' to x during the interval $[t', t' + dt']$ and to stay there at least until time t is

$$\psi(x - x', t - t') = \varphi_1(x - x') \varphi_2(t - t'), \quad (1)$$

since the random walk is translation invariant.

Furthermore, the probability density $C(x, t)$ for a particle to be located in x at instant t given that it was initially at $x = 0$, satisfies the following Kolmogorov–Feller chain equation [17]:

$$C(x, t) = \phi(t)\delta(x) + \int_0^t \varphi_2(t-t') \left[\int_{-\infty}^{\infty} \varphi_1(x-x')C(x', t')dx' \right] dt' \quad (2)$$

where $\delta(\cdot)$ is the Dirac delta function, and $\phi(t) = \int_t^{\infty} \varphi_2(t') dt'$ is the survival function representing the probability that at instant t the particle is still at position $x = 0$.

We suppose now that the particles are constrained to stay on the right-hand side ($x > 0$) of an elastic barrier ($x = 0$). Due to the boundary presence, the transition probability density will not be spatially invariant.

We shall consider two cases: (i) the tracer performs random walks within a fluid at rest, and (ii) within a fluid moving at speed v with respect to the laboratory frame.

(i) The fluid is at rest

A particle which is in x' at time t' can jump to x directly or after hitting the barrier. We assume now that the jumps are ballistic motions in a uniform force field due to random impulses distributed in accordance with φ_1 . If we further assume that hitting the wall does not affect the kinetic energy of the particles, the indirect jumps from x' to x are distributed according to the p.d.f. $\varphi_1(-x-x')$. Hence, the transition p.d.f in the presence of the boundary is

$$\begin{aligned} \Lambda(x-x', t-t') &= \psi(x-x', t-t') + \psi(-x-x', t-t') \\ &= \varphi_2(t-t') [\varphi_1(x-x') + \varphi_1(-x-x')] \end{aligned} \quad (3)$$

(ii) The fluid moves with constant mean speed v

Advection modifies the transition probability density, depending on whether particles are advected all the time or not. A tracer dissolved in a free fluid, for instance, is advected all the time. But when the fluid is enclosed in a porous matrix, particles may not be sensitive to the general drift while being trapped and it may be considered that advection acts only during the jumps.

Advection restricted to jumps in an infinite one-dimensional medium was analysed by [12]. In the latter case, the transition p.d.f for a particle in x' to jump to x during time interval $[t', t' + dt']$ and then to stay in x until instant t is

$$\psi_1(x-x', t-t') = T_{\tau_a v} \varphi_1(x-x') \varphi_2(t-t'), \quad (4)$$

with T_l being the translation of amplitude $l = v\tau_a$ along x , and τ_a denoting the mean advection time. Since quasi-instantaneous jumps are considered here, τ_a must be much smaller than τ_0 . Nevertheless intermediate situations may be imagined, allowing for all possible values of τ_a in $(0, \tau_0)$ and corresponding to sce-

narios, where particles are advected during time intervals containing the jump durations.

In a free space, if the particles are advected all the time [18] indicated that the reasoning should be different. In this latter case, the transition p.d.f. becomes

$$\psi_2(x - x', t - t') = T_{v(t-t')} \varphi_1(x - x') \varphi_2(t - t') \quad (5)$$

Just as in case (i), we incorporate the boundary condition at $x = 0$.

When advection is restricted to jumps, or more generally to time intervals of mean τ_a containing all the jump durations, travelling from x' to x implies jumping from x' to $x - v\tau_a$ with respect to a frame moving only during the time intervals involving the jumps. If we disregard possible interactions between the shock with the wall and the necessary acceleration and deceleration of the temporarily moving frame, hitting the wall means that the initial impulsion of the jump would give a landing point at abscissa $-x + v\tau_a$ if there were no wall at $x = 0$. Hence the transition p.d.f from x' to x with the last jump having started during $[t', t' + dt']$ is

$$\begin{aligned} \Lambda_1(x - x', t - t') &= H(x - v\tau_a) [\psi_1(x - x', t - t') + \psi_1(-x - x', t - t')] \\ &= H(x - v\tau_a) T_{\tau_a v} \varphi_2(t - t') [\varphi_1(x - x') + \varphi_1(-x - x')], \end{aligned} \quad (6)$$

where H denotes the Heaviside step function.

When advection applies all the time, a particle which is in x at time t after the last jump occurred at time t' , performed a jump ending at $x - v(t - t')$. This is possible for $x > v(t - t')$ only. Hence the transition probability from x' to x with the last jump having started during $[t', t' + dt']$ is

$$\begin{aligned} \Lambda_2(x - x', t - t') &= H(x - v(t - t')) [\psi_2(x - x', t - t') + \psi_2(-x - x', t - t')] \\ &= H(x - v(t - t')) T_{v(t-t')} \varphi_2(t - t') [\varphi_1(x - x') + \varphi_1(-x - x')] \end{aligned} \quad (7)$$

3 The Macroscopic Limit

To obtain the macroscopic dynamics of the CTRW in an infinite medium, the method [9, 19] consists in transforming the Kolmogorov–Feller chain equation into the Fourier–Laplace domains and in taking the appropriate asymptotic limits. In the latter case, the following space-fractional equation is obtained:

$$\partial_t C(x, t) = K \nabla_x^\alpha C(x, t), \quad (8)$$

In the latter equation, ∇_x^α is the symmetric Riesz–Feller derivative [20], defined by the Fourier transform $F(\nabla_x^\alpha G)(k) = -|k|^\alpha FG(k)$. For $\alpha = 2$, it is the usual laplacian, while for $\alpha \neq 2$, it satisfies

$$\nabla_x^\alpha G(x) = \frac{-1}{2 \cos(\alpha\pi/2) \Gamma(2-\alpha)} \partial_x^2 \int_{\mathbb{R}} |x-y|^{1-\alpha} G(y) dy \quad (9)$$

We shall adapt the latter method to find the macroscopic evolution equation corresponding to the CTRW in the presence of a reflective wall at $x = 0$.

In what follows, we denote by $\hat{\varphi}_1(k)$ and $\tilde{\varphi}_2(u)$ the Fourier and Laplace transforms, respectively of $\varphi_1(x)$ and $\varphi_2(t)$ according to

$$\hat{\varphi}_1(k) = \int_{\mathbb{R}} \varphi_1(x) e^{ikx} dx = \exp(-\sigma |k|^\alpha), \quad (10)$$

$$\tilde{\varphi}_2(u) = \int_0^{+\infty} \varphi_2(t) e^{-ut} dt = 1/(1 + \tau_0 u). \quad (11)$$

Furthermore, we denote by $\hat{h}(k, u)$ the Fourier–Laplace transform of a function $h(x, t)$.

(i) Case 1: $v = 0$

Since here $C(x, t)$ is defined on a half space, we need some appropriate extension of C to obtain a Fourier convolution. The initial condition is a Dirac pulse at $x = x_0$. Particles which are in x at instant t either came from elsewhere before, or stayed there from the beginning, hence the probability $C(x, t)$ satisfies

$$C(x, t) = \delta_{x_0}(x) \phi(t) + \int_0^t \int_{\mathbb{R}} C(x', t') \Lambda(x - x', t - t') dx' dt' \quad (12)$$

Since φ_1 is even, the even extension (w.r.t. $x \in \mathbb{R}$) C^* , of C satisfies

$$C^*(x, t) = \int_0^t \int_{\mathbb{R}} C(x', t') \psi(x - x', t - t') dx' dt' + [\delta_{x_0}(x) + \delta_{-x_0}(x)] \phi(t) \quad (13)$$

This implies that the Fourier–Laplace transform $\hat{C}^*$ satisfies

$$u \hat{C}^*(k, u) - 2 \cos kx_0 = [\exp(-\sigma |k|^\alpha) - 1] \tau_0^{-1} \hat{C}^*(k, u) \quad (14)$$

For k and u fixed, $\sigma \rightarrow 0$, $\tau_0 \rightarrow 0$, and $\sigma/\tau_0 \rightarrow K$, we obtain

$$u \hat{C}^*(k, u) - 2 \cos kx_0 = -K |k|^\alpha \hat{C}^*(k, u), \quad (15)$$

hence in physical space C^* satisfies Eq. (8).

Consequently, for $x > 0$ and $\alpha < 2$, in the asymptotics of large x and t , the concentration C satisfies

$$\partial_t C(x, t) = K \frac{-1}{2 \cos(\frac{\alpha\pi}{2}) \Gamma(2-\alpha)} \partial_x^2 \int_0^{+\infty} (|x-y|^{1-\alpha} + (x+y)^{1-\alpha}) C(y, t) dy. \quad (16)$$

We have obtained a modified kernel for the symmetric Riesz–Feller derivative of order $\alpha \neq 2$ due to the presence of the reflective boundary. For $\alpha = 2$, Eq. (15) yields the usual Fick’s law. The method still adapts when advection speed v is different from zero.

(ii) Case 2a: $v > 0$ restricted to time intervals containing jumps

In free space, and supposing that advection applies during time intervals of mean τ_a , [4] obtained the following macroscopic equation

$$\partial_t C(x, t) + \partial_x (\tau_1 v C(x, t)) = K \nabla_x^\alpha C(x, t), \quad (17)$$

where τ_1 is an adimensional coefficient between 0 and 1.

In the presence of a wall, and following the reasoning of [7], the chain equation for $x > 0$ results in

$$C(x, t) = \delta_{x_0}(x) \phi(t) + \int_0^t \int_0^\infty C(x', t') \Lambda_1(x - x', t - t') dx' dt' \quad (18)$$

By setting, in Eq. (13), $X = x' + \tau_a v$, $X' = -x' + \tau_a v$, $y = X$ for $y > 0$, and $y = X'$ for $y < 0$, yields

$$C(x, t) = \delta_{x_0}(x) \phi(t) + \int_{t'=0}^t H(x - v\tau_a) \Omega(x - v\tau_a, t') \varphi_2(t - t') dt', \quad (19)$$

where $\Omega(z, t) = \int_R C^*(z, t') \varphi_1(z - y) dy = [C^*(\cdot, t') *_F \varphi_1](z, t)$, with the $*_F$ symbol denoting Fourier convolution. Hence, for x in R we have

$$\begin{aligned} C^*(x, t) &= [\delta_{x_0}(x) + \delta_{-x_0}(x)] \phi(t) \\ &+ \int_0^t \varphi_2(t - t') [T_{v\tau_a}(H\Omega) + T_{-v\tau_a}(1 - H)\Omega](x, t') dt' \end{aligned} \quad (20)$$

Since, in Fourier–Laplace coordinates, we have

$$[T_{\tau_a v}(H\Omega)]_F(k, u) = e^{ik\tau_a v} (\Omega *_F (\delta/2 - 1/ik))(k, u),$$

Eq. (20) transforms into

$$\begin{aligned} \hat{C}^*(k, u) &= 2 \cos kx_0 \frac{1 - \tilde{\varphi}_2(u)}{u} \\ &+ \tilde{\varphi}_2(u) [e^{ikv\tau_a} \hat{H} *_F (\hat{C}^*_1) + e^{-ikv\tau_a} (1 - H)_F *_F (\hat{C}^*_1)] \end{aligned} \quad (21)$$

Eq. (21) implies that

$$\hat{C}^*(k, u) = 2 \cos kx_0 \frac{1 - \tilde{\varphi}_2(u)}{u} + \tilde{\varphi}_2(u) \hat{C}^* \hat{\varphi}_1 \cos k\tau_a v - 2i \sin kv\tau_a \left[\frac{1}{ik} *_F \hat{C}^* \hat{\varphi}_1 \right] \quad (22)$$

From the latter equation, we deduce

$$\left[\partial_t C^*(k, u) \right]_F = A \hat{C}^*(k, u) + i k v B \hat{\left[\frac{\hat{x}}{|x|} C^* *_F \varphi \right]}_F, \quad (23)$$

with $A = \tau_0^{-1} (e^{-\sigma|k|^\alpha} \cos(k\tau_a v) - 1)$, and $B = \frac{\tau_a}{\tau_0} \frac{\sin(k\tau_a v)}{k\tau_a v}$.

The $\sin(k\tau_a v)/k\tau_a v$ function is $(2\tau_a v)^{-1}$ times the Fourier transform of the function $\chi_{[-v\tau_a, v\tau_a]}$. The latter is equal to 1 on the interval $[-v\tau_a, v\tau_a]$ and 0 elsewhere. Hence, the opposite of the last term in Eq. (23) is the Fourier transform of the derivative of the convolution of $\frac{x}{|x|} C^* *_F \varphi_1$ with $(2\tau_a v)^{-1} \chi_{[-v\tau_a, v\tau_a]}$.

Letting σ , τ_a and τ_0 tend to zero with $\sigma/\tau_0 \rightarrow K$ and $\tau_a/\tau_0 \rightarrow \tau_1$, we find that $C^* *_F \varphi_1$ tends to C^* , and that its convolution with $(2\tau_a v)^{-1} \chi_{[-v\tau_a, v\tau_a]}$ tends to the identity. Hence, in this limit we finally have

$$\partial_t C^*(x, t) = K \nabla_x^\alpha C^*(x, t) - \partial_x \left[\frac{x}{|x|} v \tau_1 C^*(x, t) \right] \quad (24)$$

From Eq. (24), we deduce that, for $x > 0$, $C(x, t)$ satisfies

$$\partial_t C(x, t) + \tau_1 \partial_x (vC(x, t)) = K_1 \partial_x^2 \int_0^{+\infty} (|x - y|^{1-\alpha} + (x + y)^{1-\alpha}) C(y, t) dy, \quad (25)$$

with $K_1 = -K/2 \cos(\alpha\pi/2) \Gamma(2 - \alpha)$. The last equation is similar to Eq. (16) plus the advective term $\tau_1 \partial_x (vC(x, t))$.

(iii) Case 2b: Advection effective all the time

In this case, being in x without having performed any jump now means having been advected from $x - vt$. On the infinite line, particles which are in x at instant t either were subjected to jumps and advection, or were advected without any jump. Hence in Fourier–Laplace coordinates [17] we have

$$\hat{C}(k, u) = \frac{1 - \hat{\psi}_2(0, u - ivk)}{(u - ivk)(1 - \hat{\psi}_2(k, u - ivk))}, \quad (26)$$

which in physical variables yields Eq. (17) with $\tau_1 = 1$.

In a semi-infinite medium limited by a reflective barrier, the chain equation for $x > 0$ is

$$C(x, t) = \delta_{vt}(x - x_0) \phi(t) + \int_0^t \int_0^\infty C(x', t') \Lambda_2(x - x', t - t') dx' dt' \quad (27)$$

For x in R , using the same technique as in case 2a, yields

$$C^*(x, t) = \left[\delta_{x_0+vt}(x) + \delta_{-x_0+vt}(x) \right] \phi(t) + \int_0^t \varphi_2(t - t') \Theta(x, t') dt' \quad (28)$$

$$\text{with } \Theta = \left[T_{v(t-t')} \left(H \left[C^* *_F \varphi_1 \right] \right) + T_{-v(t-t')} \left((1-H) \left[C^* *_F \varphi_1 \right] \right) \right].$$

Hence, in Fourier–Laplace variables, we obtain

$$D \hat{C}^*(k, u) = E [\hat{C}^* *_F \frac{1}{ik}] (k, u) + 2 \cos kx_0 A - 2 \sin kx_0 B \quad (29)$$

In Eq. (29) we have used the following definitions

$$\begin{aligned} 2A &\equiv \frac{1 - \tilde{\varphi}_2(u - ivk)}{u - ivk} + \frac{1 - \tilde{\varphi}_2(u + ivk)}{u + ivk}, \\ 2iB &\equiv \frac{1 - \tilde{\varphi}_2(u - ivk)}{u - ivk} - \frac{1 - \tilde{\varphi}_2(u + ivk)}{u + ivk}, \\ D &\equiv 1 - \frac{\hat{\varphi}_1(k) [\tilde{\varphi}_2(u + ivk) + \tilde{\varphi}_2(u - ivk)]}{2}, \\ E &\equiv \hat{\varphi}_1(k) [\tilde{\varphi}_2(u + ivk) - \tilde{\varphi}_2(u - ivk)] \end{aligned}$$

Noticing that D is equal to $\frac{\tau_0 u (1 + \sigma |k|^\alpha) + \sigma |k|^\alpha + \tau_0^2 (u^2 + k^2 v^2) + O(\sigma^2 |k|^{2\alpha})}{(1 + \tau_0 u)^2 + \tau_0^2 k^2 v^2}$, we de-

duce that the Fourier–Laplace transform of $\partial_t C^*$ is

$$u \hat{C}^*(k, u) - 2 \cos kx_0 = \frac{Au - D}{D} \hat{C}^*(k, u) - \frac{2B \sin kx_0}{A} + \frac{E}{A} \left[\hat{C}^* *_F \frac{1}{ik} \right] \quad (30)$$

When σ and τ_0 tend to zero with (k, u) fixed, B is small, and $\frac{Au - D}{D}$ tends to $-K |k|^\alpha$, which in Fourier variables is the symbol of the Riesz–Feller derivative. In this limit, and in physical variables, the last term of Eq. (30) yields $-v \partial_x \left[\frac{x}{|x|} C^* \right] (x, t)$.

Hence, for $x > 0$ and in the macroscopic limit, $C(x, t)$ satisfies Eq. (25) with $\tau_1 = 1$. We conclude that Eq. (25) resumes the macroscopic limit for Lévy flights with a reflective barrier.

4 Numerical Solutions

A numerical method for the discretization of the Riesz–Feller derivative ∇_x^α in an infinite medium was proposed by [14]. For $\alpha \in (1, 2)$, a stable scheme for the $v = 0$ variant of Eq. (17) is based upon approximations to Grunwald–Letnikov derivatives computed at $x \pm h$, with h being the spatial mesh. For symmetric Riesz–Feller derivatives, the discretization of Eq. (17) is:

$$\frac{C_i^{n+1} - C_i^n}{\Delta} = K \sum_{p=-\infty}^{+\infty} a_p C_{i+p}^n \quad (31)$$

where $a_0 = -2\mu\alpha$, $a_1 = \mu(1 + \frac{\alpha(\alpha-1)}{2})$, $a_p = \mu \frac{\alpha(\alpha-1)(2-\alpha)\dots(p-\alpha)}{(p+1)!}$, $p \in (2, +\infty)$, and $\mu = \frac{1}{2h^\alpha |\cos(\pi\alpha/2)|}$, with Δ being the time mesh.

The transition from time $n\Delta$ to time $(n+1)\Delta$ can be thought of as being a redistribution scheme for the extensive quantity C . This is the keypoint in showing that scheme 31 does converge to a solution of Eq. (17) (see [14]) under the stability condition

$$K\alpha\Delta h^{-\alpha} \leq |\cos(\alpha\pi/2)| \quad (32)$$

We shall adapt the above method to the kernel of Eq. (25) with a reflective boundary condition.

(i) Case 1: $v = 0$

With $v = 0$, the even extension C^* of C satisfies Eq. (17). When the stability condition 32 is verified, the scheme 31 yields an approximation to C^* . Since the sequence (a_p) is symmetric, for positive valued i we have

$$\frac{C_i^{n+1} - C_i^n}{\Delta} = K(a_{-i}C_0^n + \sum_{p>-i} (a_p + a_{2i+p})C_{i+p}^n), \quad (33)$$

We compared the fundamental analytical and numerical solutions of Eq. (25). A comparison is presented in Fig. 1.

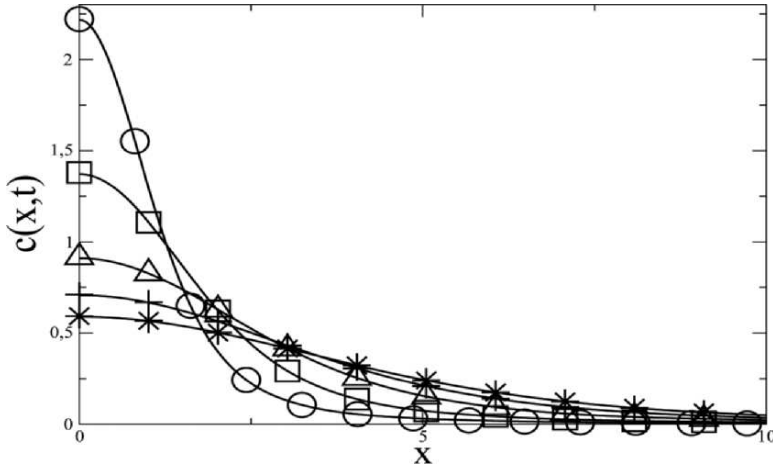


Fig. 1. Comparison of numerical (full line) and analytical (symbols) solutions of Eq. (25), for $\alpha = 1.5$. The initial condition is a Dirac pulse in $x = 0$.

(ii) Case 2: With the advective term $v \partial_x C$

Even when α is equal to two, some care is needed when attempting to approximate advective terms such as $v \partial_x C$. A finite volume scheme [10], using intermediate points (see Fig. 2) is associated with the approximation $C_I = \frac{1}{2}(C_{i+1} + C_i) - w(C_{i+1} - 2C_i + C_{i-1})$.

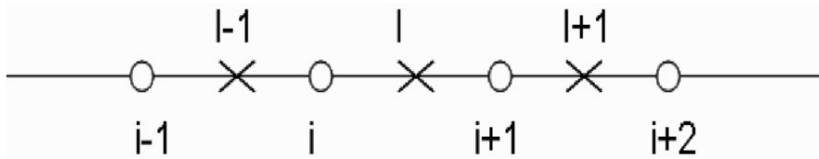


Fig. 2. The intermediate points discretization.

Then, for the advective part we take

$$-\frac{v}{2h} (f(C_I^n - C_{I-1}^n) + (1-f)(C_{I+1}^n - C_I^n)) \quad (34)$$

An appropriate choice of f and w yields for $v \partial_x C$ a third-order approximation. The scheme obtained by combining Eqs. (34) and (31) is stable when

$$0 \leq \Delta \frac{vh^{\alpha-1}\cos(\pi\alpha/2) - K\alpha}{h^\alpha \cos(\pi\alpha/2)} \leq 1 \quad (35)$$

The extra condition 35 can be derived from Neumann's method or by requiring the matrix, computing $(C_0^{n+1}, \dots, C_i^{n+1}, \dots)$ from $(C_0^n, \dots, C_i^n, \dots)$ to be stochastic.

Since no exact solution is now available for checking the above scheme, we compare with Monte Carlo simulations. We implemented a CTRW approximating Lévy flights at discrete regularly spaced instants with a reflective boundary condition. For a large number of particles, the histograms issued from Monte Carlo simulations approach the solutions to Eq. (25).

Figure 3 shows comparisons between the numerical and the MonteCarlo simulations starting from Dirac pulses in $x = 5$.

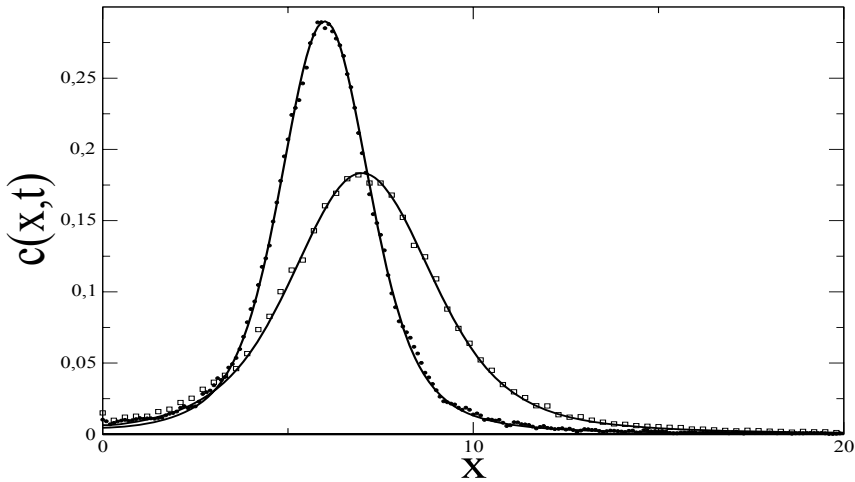


Fig. 3. Numerical (full-line) and Monte Carlo (symbols) solutions for two different times, $\alpha = 1$ and $v = 1$, with an initial Dirac pulse at $x = 5$.

5 Discussion and Conclusions

We showed that, due to its non-local character, the kernel of the fractional space derivative has to be modified when a boundary condition is imposed. Here we focused on a reflecting barrier, but other boundary conditions also present practical interest.

In simple situations, characterized by a small scale dynamics due to conservative forces superimposed to randomly distributed impulsions, the transition probability density of the random walk “with the wall” is given by Eq. (7). This equation

served as a small-scale definition of the barrier, we studied here. It was chosen in analogy with the ballistic illustration, and also because with $v = 0$ it would result in the Neumann boundary condition.

When x and y are positive valued, $|x - y|$ is smaller than $x + y$. Hence, except near the wall, the correction $(x + y)^{1-\alpha}$ to the kernel of the symmetric Riesz–Feller derivative has little influence when the support of the initial condition is concentrated. We compared solutions to Eq. (17) and (25). Generally speaking, the influence of the reflective barrier is visible between the wall and the places, where solute was initially injected, as shown in Figs. 4 and 5. When v is increased, the influence becomes smaller.

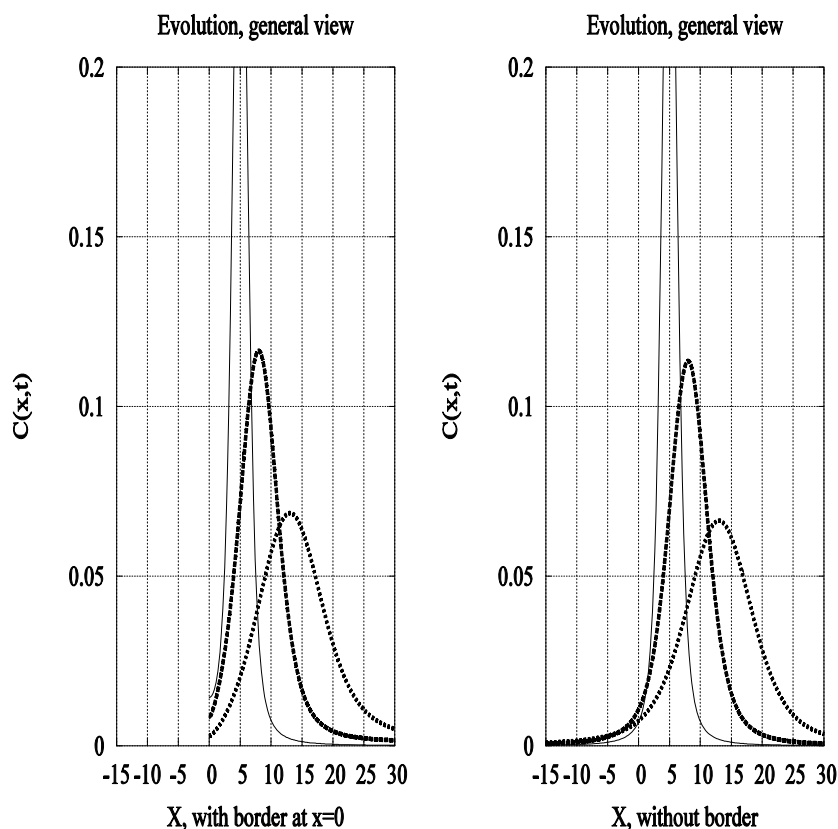


Fig. 4. Solutions to the advective fractional equation ($\alpha = 1.5$, $v = 1$) with a reflective barrier at $x = 0$ (left) and without a border (right). Initial condition: a Dirac pulse at $x = 5$.

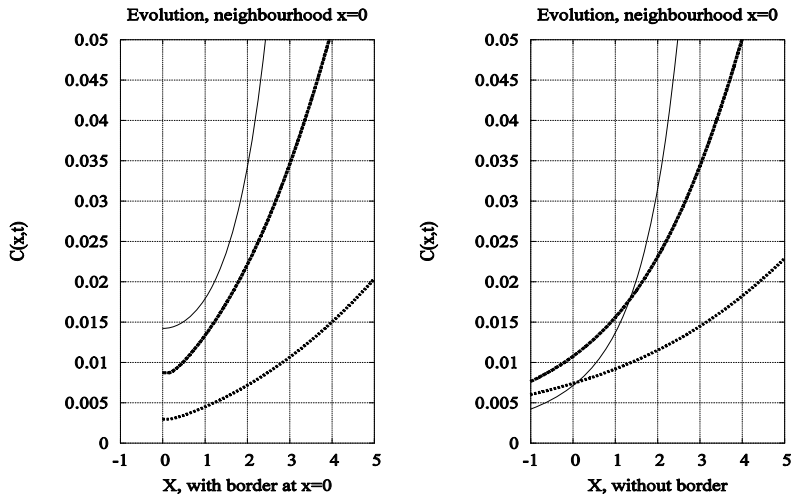


Fig. 5. A zoom of Fig. 4 in a neighbourhood of $x = 0$.

Numerical simulations indicate that the here studied boundary condition influences the spreading of matter more or less locally, especially when the advection speed is large.

References

1. Benson A, Wheatcraft S, Meerschaert M (2000) Application of a fractional advection-dispersion equation. *Water Resour. Res.*, 36(6):1403–1412.
2. Benson D, Schumer R, Meerschaert M, Wheatcraft S (2001) Fractional dispersion, Levy motion and the MADE tracer tests. *Trans. Porous Media*, 42:211–240.
3. Gelhar L (1993) *Stochastic Subsurface Hydrology*. Prentice Hall, New Jersey, USA.
4. Matheron G, De Marsily G (1980) Is transport in porous media always diffusive? A counterexample. *Water Resour. Res.*, 5:901–917.
5. Muralidhar R, Ramkrishna D (1993) Diffusion in pore fractals: a review of linear response models. *Trans. Porous Media*, 13(1):79–95.
6. Benson D, Wheatcraft S, Meerschaert M (2000) The fractional order governing equation of Levy motion. *Water Resour. Res.*, 36(6):1413–1423.
7. Chaves A (1998) A fractional diffusion equation to describe Levy flights. *Phys. Lett. A*, 239:13–16.
8. Paradisi P, Cesari R, Mainardi F, Tampieri F (2001) The fractional Fick's law for non-local transport processes. *Physica A*, 293:130–142.
9. Compte A (1996) Stochastic foundations of fractional dynamics. *Phys. Rev. E*, 53(4):4191–4193.
10. Klafter J, Blumen A, Shlesinger M (1987) Stochastic pathway to anomalous diffusion. *Phys. Rev. A*, 7:3081–3085.

11. Metzler R, Klafter J (2000) The random walk guide to anomalous diffusion: a fractional dynamic approach. *Phys. Rep.*, 339:1–77.
12. Compte A (1997) Continuous random walks on moving fluids. *Phys. Rev. E*, 55(6):6821–6830.
13. Gorenflo R, Mainardi F (1999) Approximation of Levy-Feller diffusion by random walk models. *J. Anal. App. (ZAA)*, 18:231–246.
14. Gorenflo R, Mainardi F, Moretti D, Pagmni G, Paradisi P (2002) Fractional diffusion: probability distributions and random walk models. *Physica A*, 305 (1–2):106–112.
15. Brockman P, Sokolov I (2002) Levy flights in external force fields, from models to equations. *Chem. Phys.*, 284(1–2):409–421.
16. Zumofen G, Klafter J (1995) Absorbing boundaries in one-dimensional anomalous transport. *Phys. Rev. E*, 4:2805–2814.
17. Montrol E, Weiss G (1965) Random walks on lattices II. *J. Math. Phys.* 6:167–181.
18. Metzler R, Klafter J, Sokolov I (1998) Anomalous transport in external fields: continuous time random walks and fractional diffusion equations extended. *Phys. Rev. E*, 58(2):1621.
19. Montrol E, West B (1979) On an enriched collection of stochastic processes, In: Montrol E, Lebowitz J (eds.), *Fluctuation Phenomena*:66.
20. Gorenflo R, Mainardi F (1998) Random walk models for space fractional diffusion processes. *Fract. Cal. App. Anal.* 12:167–191.

SOLUTE SPREADING IN HETEROGENEOUS AGGREGATED POROUS MEDIA

Kira Logvinova and Marie Christine Néel

UMRA “Climat, Sol, Environnement” INRA-UAPV, Faculté des Sciences, 33 rue Pasteur, 84000 Avignon, France; Tel: 33+(0)4 90 14 44 61, E-mail: marie-christine.neel@univ-avignon.fr.

Abstract

Solute spreading is studied, in saturated but heterogeneous porous media. The solid matrix is assumed to be composed of bounded obstacles, and the logarithm of the porosity is supposed to be represented by a three-dimensional random process. The latter appears as a parameter in the equation, ruling solute spreading, on the small scale. The concentration of solute, averaged with respect to the process, satisfies an equation which resembles Fourier’s law, except that it involves a term, non-local with respect to time.

Keywords

Solute spreading, random media, non-normal diffusion, integro-differential equation.

1 Introduction

Classical results [1, 2] indicated that solute spreading in very heterogeneous media may deviate from Fourier’s law. Indeed, in some aquifers it sometimes happens that the second moment of the concentration of a tracer plume is not proportional to time [3], even in saturated porous media. It seems that non-local effects also have to be accounted for in the more complicated case of unsaturated porous media [4]. Similar behaviours, corresponding to non-normal dispersion, were observed in other domains of physics, like for instance the transport of charge carriers in semiconductors [5].

Several models, in the form of partial differential equations involving fractional derivatives, which are integro-differential operators, have solutions showing non-normally diffusive behaviours. So called fractional models [6–9] were derived from continuous time Random Walks, the particles of tracer are supposed to undergo on the small scale. Other models were derived by assuming that the spreading of matter obeys Fourier’s law on the small scale, with coefficients in the form of random processes, thus attempting to describe the disorder inside the medium. This approach was used by [10] and [11], inspired from [12] and [13]. In this spirit, a model was derived by [14] for solute transport in porous media made of randomly twisted tubes filled by a fluid at rest. One can expect from [14] and

[5] that varying the random geometrical structure of a heterogeneous medium may lead to various fractional equations for the spreading of matter.

Here we consider a porous medium, whose solid matrix is made of grains which are nearly spherical. The voids between them are filled with a fluid at rest, and solute spreading is studied. We assume that the porosity of the medium is a three-dimensional random process ε such that averaging (w.r.t. the process) the concentration of solute yields the macroscopic concentration $\langle u \rangle$. In Fourier–Laplace space, the latter satisfies an integral Volterra equation involving process ε , which can be solved for $\langle u(\mathbf{k}, q) \rangle$ in Fourier–Laplace variables $\mathbf{k}$ and q . If the multipoint correlations of some well chosen function $\mathcal{E}$ of ε behave almost as if process ε were Gaussian, then the Feynman diagram method helps computing the expression of $\langle u(\mathbf{k}, q) \rangle$. Taking the limit of small $\mathbf{k}$ and q , we obtain for $\langle u(\mathbf{x}, t) \rangle$ a fractional partial differential equation in the space-time variables $(\mathbf{x}, t)$.

After necessary details concerning the partial differential equation, which rules the evolution of the concentration of solute on the small scale, we obtain formally the macroscopic concentration in Fourier–Laplace space. Upon averaging we derive the equation, ruling the evolution of the macroscopic concentration of solute. And we show that the second moment $\int_{\mathbb{R}} \mathbf{x}^2 \langle u(\mathbf{x}, t) \rangle d\mathbf{x}$ is not proportional to t , which is the hallmark of non-normal diffusion.

2 The Porous Medium, on the Small Scale

2.1 Representation of the medium

The works of [5] and [14] suggest simple models of disordered porous media where particles diffuse. They address situations such that a one-dimensional description is relevant. Solute spreading was studied by [14] in a medium made of a collection of tubes, twisted around a general direction, and saturated by motionless fluid. The concentration of solute, on the macroscopic scale, was shown to evolve according to a variant of Fourier’s law, involving a fractional derivative with respect to time. Such a model accounts for possible non-normally diffusive behaviours.

Here we consider another type of disordered porous medium: we suppose that the solid matrix is a collection of randomly distributed grains with variable diameters. For solute spreading in a uniform fluid filling the voids between grains, as in Fig. 1, a three-dimensional description is appropriate.

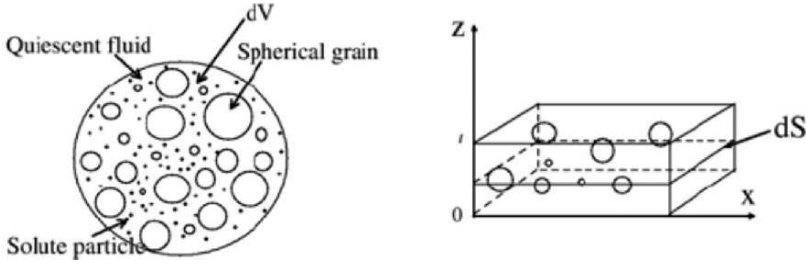


Fig. 1. The porous medium, on the small scale (left) and an elementary volume of porous medium containing many spherical aggregates (right).

A similar situation is observed in the cooling shell of an atomic reactor when the lead melt contains grains of iron (or different iron compounds) and atoms of some other substance [15]. The model also works for particles of solute, dissolved in a fluid at rest filling the voids of a porous matrix, made of aggregates whose largest diameter is less than a_0 . The volume of grains dV_{gr} in the elementary volume dV is $dV_{gr} = dV \cdot K(\mathbf{x})$, where $K(\mathbf{x})$ is the volume of grains per unit volume near point $\mathbf{x}$. Then a grain-free volume element $d\tilde{V}$ near volume dV satisfies

$$d\tilde{V} = dV - dV_{gr} = dV(1 - K(\mathbf{x})) = \varepsilon(\mathbf{x})dV. \quad (1)$$

Here $\varepsilon(\mathbf{x}) = \frac{d\tilde{V}}{dV} = 1 - K(\mathbf{x})$ is the porosity of the medium at point $\mathbf{x}$.

A grain-free area of the surface element $d\tilde{S}$ of the elementary volume dV , satisfies $d\tilde{S} = \varepsilon(\mathbf{x})dS$. Indeed, let us consider a parallelepiped whose basis (perpendicular to OZ) has area dS while the height is l (Fig. 1). Due to Eq. (2), the volume filled with grains is $dV_{gr} = K(\mathbf{x}) \cdot l \cdot dS$. Consider a plane, perpendicular to OZ , and intersecting the elementary parallelepiped at level z . Let $dS_{gr}(z)$ be the area common to the plane and to the grains. The volume of the grains cutting the plane is $dV_{gr} = \int_0^l dS_{gr}(z)dz$, and the grain-free area in the plane is $d\tilde{S}(z) = dS - dS_{gr}(z)$. We focus on scales greater than the largest grain size a_0 , i.e., $l \gg a_0$. Nevertheless only scale-averaged area values make sense. If we average $d\tilde{S}(z)$ over $[0, l]$, then we get

$$d\tilde{S} = \frac{1}{l} \int_0^l d\tilde{S}(z)dz = \frac{1}{l} \int_0^l [dS - dS_{gr}]dz = dS - \frac{dV_{gr}}{l}, \quad (2)$$

which implies

$$d\tilde{S} = dS - K(\mathbf{x})dS = \varepsilon(\mathbf{x})dS. \quad (3)$$

2.2 An equation for the solute concentration, on the small scale

In any volume V limited by a closed surface S , the total mass balance is

$$\partial_t \int_V u d\tilde{V} = - \int_S \mathbf{j} d\tilde{S}. \quad (4)$$

In Eq. (4), $\mathbf{j}$ represents the density of particles flux in V : the right-hand side is the flux through $\tilde{S}$, and Eqs. (2), (3), and (4) imply

$$\partial_t u \varepsilon + \operatorname{div}(\mathbf{j} \varepsilon) = 0 \quad (5)$$

with $\mathbf{j} = -D_0 \nabla u$, if Fick's law holds locally. Here, D_0 is the diffusion coefficient, and Eq. (5) is equivalent to

$$\partial_t u \varepsilon - D_0 \nabla \cdot \varepsilon \nabla u = 0, \quad (6)$$

which is a particular case of Eqs. (1.4)–(6.4) of [16]. The local model differs from that in [14] because of the structure of the medium. It also differs from that in [5] since Eq. (6) takes into account the dependency of the effective diffusion coefficient on the porosity $\varepsilon(\mathbf{x})$, while [5] started from $\partial_t u \varepsilon - D_0 \partial_x^2 u = 0$, on the small scale. We assume that porosity $\varepsilon(\mathbf{x})$ is a random process $\varepsilon(\mathbf{x}) = \varepsilon_0 e^{-\varepsilon'(\mathbf{x})}$, with $\varepsilon'(\mathbf{x})$ having zero average while ε_0 is a positive constant. The model is convenient for porosity, which takes values between 0 and 1. Then Eq. (6) is equivalent to

$$\partial_t u - D_0 \nabla^2 u = -D_0 (\nabla \varepsilon') \cdot \nabla u. \quad (7)$$

The solution u to Eq. (7) is a functional of random process $\varepsilon'(\mathbf{x})$. Denoting by $\langle u \rangle$ the mean concentration, where angular brackets $\langle \rangle$ stand for averaging over all possible realizations of process $\varepsilon'(\mathbf{x})$, we will derive from Eq. (7) an equation for $\langle u \rangle$. Since Eq. (7) is linear, it is enough to consider the fundamental solution, associated with the initial condition $u(\mathbf{x}, 0) = \delta(\mathbf{x})$, with δ denoting the Dirac function.

In Fourier–Laplace variables, we will obtain an equation for $\langle u \rangle$.

3 Evolution of the Concentration, on the Macroscopic Level

After having solved Eq. (7) for each sample path of process ε' , we will consider $\langle u \rangle$, which is the solution, averaged w.r.t. the process. We will see that in the

limit of large $(\mathbf{x}, t)$, $\langle u \rangle$ satisfies an equation which is Eq. (7) plus an additional non-local term.

3.1 Solving Eq. (7) for an arbitrary realization of $\varepsilon'(\mathbf{x})$

Equation (7) is very similar to the one, studied by [17], except that here we have a time derivative. The main tools allowing to solve Eq. (7) are Fourier and Laplace transforms. Here we denote by $\hat{f}(\mathbf{k}) = \int_{R^3} e^{i\mathbf{k}\cdot\mathbf{x}} f(\mathbf{x}) d\mathbf{x}$ the Fourier transform of function f , defined in R^3 and by $\hat{g}(q) = \frac{n!}{r!(n-r)!}$ the Laplace transform of function g , defined in R^+ . For a function h of $(\mathbf{x}, t)$ in R^4 , the Fourier–Laplace transform will be denoted by $\hat{h}(\mathbf{k}, q)$. In Fourier–Laplace variables, Eq. (7) is equivalent to

$$\hat{u}(\mathbf{k}_0, q) = P(\mathbf{k}_0, q) + F_{\mathbf{k}_0, q}(\hat{u}), \quad (8)$$

with $P(\mathbf{k}_0, q) = (q + D_0 \mathbf{k}_0^2)^{-1}$, and

$$F_{\mathbf{k}_0, q}(\hat{h}) = \frac{D_0}{(2\pi)^3} P(\mathbf{k}_0, q) \int_{R^3} \mathbf{k} \cdot (\mathbf{k}_0 - \mathbf{k}) \hat{\varepsilon}(\mathbf{k}) \hat{h}(\mathbf{k}_0 - \mathbf{k}, q) d\mathbf{k}. \quad (9)$$

The fixed point Eq. (8) expands into

$$\begin{aligned} \hat{u}(\mathbf{k}_0, q) = & P(\mathbf{k}_0, q) + F_{\mathbf{k}_0, q}(P(\mathbf{k}_0, q)) + F_{\mathbf{k}_0, q}(F_{\mathbf{k}_0, q}(P(\mathbf{k}_0, q))) \\ & + \sum_{n \geq 2} F_{\mathbf{k}_0, q}(F_{\mathbf{k}_0, q} \dots F_{\mathbf{k}_0, q}(P(\mathbf{k}_0, q))) \end{aligned} \quad (10)$$

with $(\cdot), (\cdot), \dots, n(\cdot)$ representing dumb variables in successive integrations.

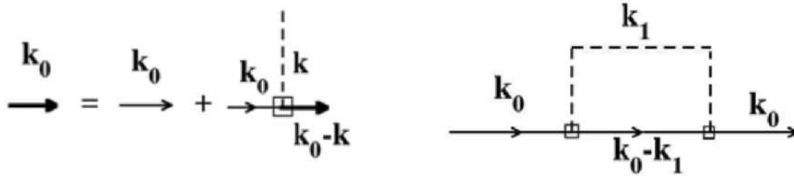


Fig. 2. A diagrammatic reformulation of Eq. (8) (left) and a graphic representation of $\langle F_{\mathbf{k}_0, q}(F_{\mathbf{k}_0, q}(P(\mathbf{k}_0, q))) \rangle$ (right).

It is now classical (see [18] and [19]) to replace Eqs. (8) and (10) by Feynman diagrams such as the ones, displayed in Fig. 2 (left) and 3. Fig. 3 is an equivalent formulation of Eq. (10) with the following rules:

1. Wave-vector is conserved at each vertex
2. Each thick horizontal line labelled $\mathbf{k}_0$ represents $\hat{u}(\mathbf{k}_0, q)$ while each thin horizontal line labelled $\mathbf{k}$ represents $P(\mathbf{k}, q)$
3. Each vertical dashed line with a free end labelled $\mathbf{k}$ carries a factor by $\hat{\varepsilon}(\mathbf{k})(2\pi)^{-3}$ and integration over R^3 w.r.t. $\mathbf{k}$
4. Each three lines vertex in a square box is connected to three lines labelled by wave-vectors $\mathbf{k}_1$ (at the left), $\mathbf{k}_2$ (on the vertical line), and $\mathbf{k}_1 - \mathbf{k}_2$ at the right, since wave-vector is conserved. The three lines vertex carries a factor of $D_0 \mathbf{k}_2 \cdot (\mathbf{k}_1 - \mathbf{k}_2)$

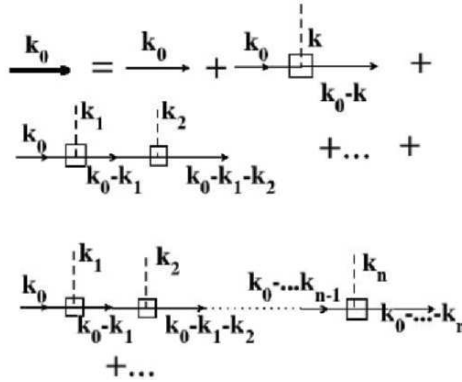


Fig. 3. A diagrammatic reformulation of Eq. (10).

3.2 An equation for $\langle u \rangle$

Since ε' is a centred Gaussian process, terms where $\varepsilon'(\cdot)$ appears an odd number of times give no contribution to $\langle u \rangle$, and

$$\langle \hat{\varepsilon}'(\mathbf{k}_1) \dots \hat{\varepsilon}'(\mathbf{k}_{2n}) \rangle = \Sigma_{\text{pairings}} (2\pi)^{3n} \Pi_{\alpha_1, \alpha_2} \hat{\psi}(\mathbf{k}_{\alpha_1}) \delta_{\mathbf{k}_{\alpha_1} + \mathbf{k}_{\alpha_2}}, \quad (11)$$

with Σ_{pairings} being a sum over all the possible pairings of $\{1, \dots, 2n\}$ while ψ is the correlation function of process ε' . As in [20] we have

$$\langle \hat{\varepsilon}'(\mathbf{k}_1) \dots \hat{\varepsilon}'(\mathbf{k}_n) \rangle = (2\pi)^3 \delta_{\mathbf{k}_1 + \dots + \mathbf{k}_n} \langle \hat{\varepsilon}'(\mathbf{k}_2 - \mathbf{k}_1) \dots \hat{\varepsilon}'(\mathbf{k}_n - \mathbf{k}_{n-1}) \rangle, \quad (12)$$

which results in reducing the number of integrations in the average of Eq. (10), thus compacting the integration variables which label dashed vertical lines with free ends in Fig. 3, so that

$$\begin{aligned}
\langle F_{\mathbf{k}_0,q}(F_{(\cdot),q}(P((\cdot),q))) \rangle &= \frac{D_0^2}{(2\pi)^3} P(\mathbf{k}_0, q) \int_{R^3} (\mathbf{k}_1 \cdot (\mathbf{k}_0 - \mathbf{k}_1)) \hat{\psi}(\mathbf{k}_1) \delta_{\mathbf{k}_1+\mathbf{k}_2} \\
&\quad (-\mathbf{k}_2 \cdot (\mathbf{k}_0 - \mathbf{k}_1 - \mathbf{k}_2)) P(\mathbf{k}_0 - \mathbf{k}_1, q) P(\mathbf{k}_0 - \mathbf{k}_1 - \mathbf{k}_2, q) d\mathbf{k}_1 d\mathbf{k}_2 = \\
&= -\frac{D_0^2}{(2\pi)^3} P(\mathbf{k}_0, q) \int_{R^3} \mathbf{k}_1 \cdot (\mathbf{k}_0 - \mathbf{k}_1) \hat{\psi}(\mathbf{k}_1) \mathbf{k}_1 \cdot \mathbf{k}_0 P(\mathbf{k}_0 - \mathbf{k}_1, q) d\mathbf{k}_1 P(\mathbf{k}_0, q).
\end{aligned} \tag{13}$$

The open diagram with two vertices on Fig. 3 corresponds to $F_{\mathbf{k}_0,q}(F_{(\cdot),q}(P((\cdot),q))) \otimes$. Due to Eq. (12), the average $\langle F_{\mathbf{k}_0,q}(F_{(\cdot),q}(P((\cdot),q))) \rangle$ is represented by the diagram, displayed on Fig. 2 (right). Formerly open vertical lines close up and will be labelled by the wave-vector of the left end.

Averaging Eq. (10) and splitting the correlations modifies rule 2 which becomes: 2'. Each dashed lines connecting two vertices and labelled by wave vector $\mathbf{k}$ carries a factor of $-\hat{\psi}(\mathbf{k})(2\pi)^{-3}$ and an integration over R^3 w.r.t. $\mathbf{k}$.

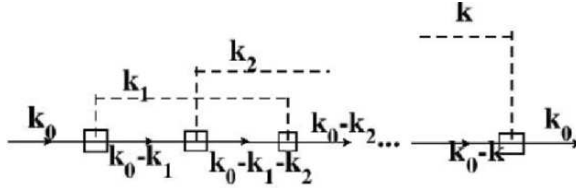


Fig. 4. A diagram, representing one of the integrals over R^{3n} , obtained by splitting $\langle F_{\mathbf{k}_0,q}(F_{(\cdot),q} \dots (F_{(2n-1)(\cdot),q}(P(2n(\cdot),q))) \rangle$.

Indeed, splitting correlations splits the average $\langle F_{\mathbf{k}_0,q} F_{(\cdot),q} \dots F_{(2n-1)(\cdot),q} P(2n(\cdot),q) \rangle$ into a sum of $(2n-1)(2n-3) \dots 3 \cdot 1$ items. For convenience we will denote by $B_i(\mathbf{k}_0, q)$ the terms, obtained this way for all values of n in the average of the right-hand side of Eq. (10). The generic $B_i(\mathbf{k}_0, q)$ is represented graphically by a diagram obtained from the last one in Fig. 3 by connecting all the possible pairs of vertical lines (assuming that the diagram contains an even number of vertices). Each time free vertical lines labelled by $\mathbf{k}_{\alpha_1}$ (for the line at the left) and $\mathbf{k}_{\alpha_2}$ (for the line at the right) are connected, the environment of vertex (1) from which line $\mathbf{k}_{\alpha_1}$ is issued remains unchanged. But the environment of vertex (2) (at the origin of line $\mathbf{k}_{\alpha_2}$) is modified. Indeed, the $\hat{\psi}(\mathbf{k}_{\alpha_1}) \delta_{\mathbf{k}_{\alpha_1}+\mathbf{k}_{\alpha_2}}$ in the integrals yield that in the labels of all the horizontal lines at the right of vertex (2), $\mathbf{k}_{\alpha_2}$ is replaced by $-\mathbf{k}_{\alpha_1}$. The dashed line, which now connects vertices (1) and (2) is labelled $\mathbf{k}_{\alpha_1}$ and denotes multiplication by $\hat{\psi}(\mathbf{k}_{\alpha_1})(2\pi)^{-3}$ and integration w.r.t. $\mathbf{k}_{\alpha_1}$ over R^3 , as on the example, displayed on Fig. 4.

The three terms issued from $\langle F_{\mathbf{k}_0, q} (F_{(\cdot), q} F_{(\cdot), q} (F_{(\cdot), q} (P((\dots), q))) \rangle$ are represented by the three diagrams displayed on Fig. 5.

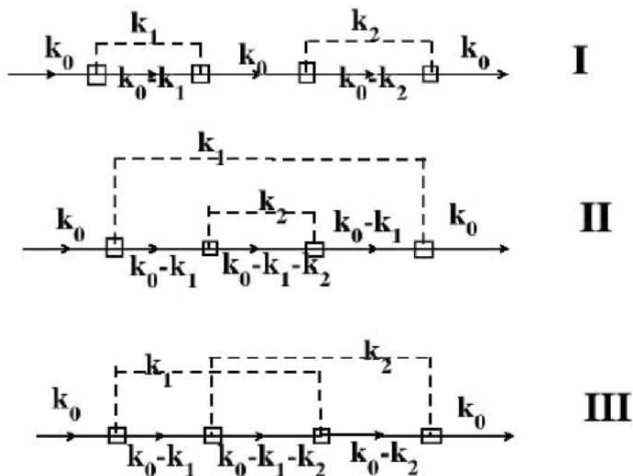


Fig. 5. Diagrammatic reformulation of the three integrals over R^6 , obtained by splitting $\langle F_{\mathbf{k}_0, q} (F_{(\cdot), q} (F_{(\cdot), q} (F_{(\cdot), q} (P((\dots), q))) \rangle$.

Cutting the extremal horizontal lines [18], allows discriminating between reducible and irreducible diagrams. After this operation, which transforms $B_j(\mathbf{k}_0, q)$ into $B'_j(\mathbf{k}_0, q)$, any diagram containing at least one horizontal line not being surrounded by any dashed line connecting vertices is reducible: cutting the horizontal line yields two (smaller) diagrams. When it is not possible to find any horizontal line not being surrounded by some dashed line connecting vertices, the diagram is irreducible. For instance, among the three diagrams with four vertices, displayed on Fig. 5, I is reducible while II and III are not. Let us write $A_i(\mathbf{k}_0, q)$ for the $B'_j(\mathbf{k}_0, q)$, represented by irreducible diagrams, whose sum will be called S' . The reducible $B'_j(\mathbf{k}_0, q)$ are of the form $A_{i_1}(\mathbf{k}_0, q)P(\mathbf{k}_0, q)\dots A_{i_n}(\mathbf{k}_0, q)$, hence (see [18]) the right-hand side of the average of Eq. (10) is $P(\mathbf{k}_0, q)(Id - S'P(\mathbf{k}_0, q))^{-1} = (P(\mathbf{k}_0, q)^{-1} - S')^{-1}$. This implies

$$\langle \hat{u} \rangle(\mathbf{k}_0, q) = (P(\mathbf{k}_0, q)^{-1} - S')^{-1}, \quad (14)$$

or equivalently

$$(P(\mathbf{k}_0, q)^{-1} - S')\langle u \rangle(\mathbf{k}_0, q) = 1. \quad (15)$$

Let us denote by S the operator whose Fourier–Laplace symbol is S' . With the initial condition $u(\mathbf{x}, 0) = \delta(\mathbf{x})$, $P(\mathbf{k}_0, q)^{-1} - 1$ is the Fourier–Laplace symbol of $\partial_t - D_0 \nabla^2$, hence Eq. (15) is equivalent to

$$(\partial_t - D_0 \nabla^2 - S)\langle u \rangle(\mathbf{x}, t) = 0. \quad (16)$$

The leading orders in the symbol S' of S will yield the equation satisfied by $\langle u \rangle(\mathbf{x}, t)$ in the large $(\mathbf{x}, t)$ limit.

3.3 The macroscopic limit of Eq. (16)

Scaling with $k_0 = |\mathbf{k}_0|$ the integrals in the A_l shows that the leading order in S is given by the first loop term. We learned from the one-dimensional version of the problem at hand, that a broad class of even correlation functions ψ may yield the same structure for the macroscopic limit Eq. (16). Here, in the dimension three, specializing ψ is necessary. Nevertheless, we will arrive at similar results with two different examples.

3.3.1 With a correlation function, connected with exponentials

If we choose $\psi(\mathbf{x}) = (\Phi * \Phi * \Phi)(\mathbf{x})$ with $\Phi(\mathbf{x}) = ae^{-|\mathbf{x}|/\ell}$, the Fourier transform is $\hat{\psi}(\mathbf{k}) = b^3(1 + \ell^2 \mathbf{k}^2)^{-3}$, with $b = a\ell$. The first order in S' is $-D_0 I / (2\pi)^3$ with

$$I = \int_{\mathbb{R}^3} \frac{((\mathbf{k}_0 - \mathbf{k}_1) \cdot \mathbf{k}_1)(\mathbf{k}_0 \cdot \mathbf{k}_1)}{q + D_0(\mathbf{k}_0 - \mathbf{k}_1)^2} \psi(\mathbf{k}_1) d\mathbf{k}_1. \quad (17)$$

Setting $Q = q/(D_0 k_0^2)$ and $L = \ell k_0$, integrating over angles, we have $I k_0^{-5} / (2\pi)^3 = I_1 + I_2$, with

$$I_1 = \frac{1}{2} \int_0^{+\infty} \rho^2 \frac{\rho^2 - (Q+1)}{(1 + L^2 \rho^2)^3} d\rho \quad (18)$$

and

$$I_2 = \int_0^{+\infty} \rho \frac{\rho^4 - (Q+1)^2}{8(1 + L^2 \rho^2)^3} \text{Ln}\left(\frac{Q + (\rho-1)^2}{Q + (\rho+1)^2}\right) d\rho. \quad (19)$$

Jordan's Lemma applies to both integrals corresponding to $D_1 \cup D_2 \cup D_3$ on the curve $C = D_1 \cup \Delta_1 \cup C_- \cup \Delta_2 \cup D_2 \cup \Delta_3 \cup C_+ \cup \Delta_4 \cup D_3 \cup C'$, represented on Fig. 6. The contribution of the singularities of the logarithm at $\pm 1 + i\sqrt{Q}$ yields $(1 + 5/3)Q^{3/2}$, resulting in $\eta \partial_t^{3/2} \nabla^2$ in $-S'$, with $\eta = b^3 \ell^3 (2\pi)^{-3} (1 + 5/3) \frac{\pi}{D_0^{1/2}}$.

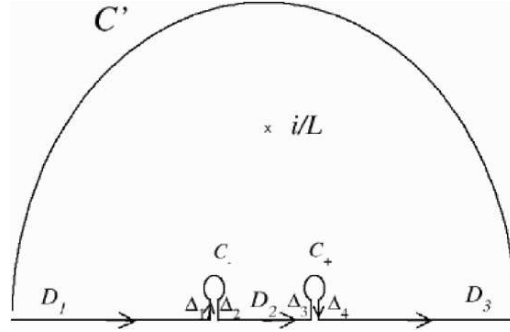


Fig. 6. Contour C : circle C' is centered at the origin, with a radius, tending to ∞ , while circles C_- and C_+ are small and centered at $\pm 1 + iQ^{1/2}$. Segments Δ_1 and Δ_2 connect C_- to the real axis, while Δ_3 and Δ_4 connect C_+ to the real axis, which is made of the D_i .

The contribution of the (triple) pole i/L to $I_1 + I_2$ is $-9\pi/(8L^3)$. Hence the contribution to S' is $9b^3k_0^2/(64\pi^2)$, which only results in decreasing to D_0 the value of the diffusivity D_0 , in agreement with [17].

3.3.2 When the correlation function is a Gaussian

When the Fourier transform $\hat{\psi}$ is of the form $\hat{\psi}(\mathbf{k}) = a^3 \ell e^{-(r/\ell)^2}$, the first order in S' is $-D_0 I$ with

$$I = k_0^5 \int_0^{+\infty} e^{-L\rho^2} \left[\rho^2 \frac{\rho^2 - (Q+1)}{2} + \rho \frac{\rho^4 - (Q+1)^2}{8} \text{Ln} \left(\frac{Q + (\rho-1)^2}{Q + (\rho+1)^2} \right) \right] d\rho. \quad (20)$$

Integrating by parts, we obtain

$$8I = \left(\frac{8}{L^4} - \frac{8Q}{L^2} \right) \int_0^{+\infty} e^{-L\rho^2} d\rho + \left(2Q \frac{Q-1}{L^2} - \frac{2Q}{L^4} \right) B + \left(\frac{1}{L^6} + \frac{1-Q}{L^4} - \frac{4Q}{L^2} \right) A. \quad (21)$$

with $A = 2 \int_0^{+\infty} e^{-L\rho^2} \left(\frac{\rho-1}{Q+(\rho-1)^2} - \frac{\rho+1}{Q+(\rho+1)^2} \right) d\rho$ and

$B = 2 \int_0^{+\infty} e^{-L\rho^2} \left(\frac{1}{Q+(\rho-1)^2} + \frac{1}{Q+(\rho+1)^2} \right) d\rho$. Noticing that A is the convolution of $e^{-L\rho^2}$ with $\frac{-\rho}{1+\rho^2}$, computed at $Q^{-1/2}$, plus a similar expression, computed at $-Q^{-1/2}$, then integrating by parts several times, we find

$$A = -4L\pi^{1/2} + 4L^2\pi Q^{1/2}e^{QL^2} + 8\pi^{1/2}L^3(1/3 - Q) - 4L^4\pi Q^{1/2}e^{QL^2} \quad (22)$$

$$+ L^5\pi^{1/2}\left(8Q - \frac{16}{3}Q^{5/2} + 8\frac{16}{5!} - \frac{8}{3}Q\right) + 16\pi^{1/2}L^6\left(\frac{Q^{3/2}}{6} + \frac{15}{5!}Q^{1/2}\right) + O(L^7).$$

Similarly we obtain

$$B = 2\pi^{1/2}Q^{-1/2}e^{QL^2} - 4\pi^{1/2}LQ^{1/2} - 2L^2\pi^{1/2}Q^{-1/2}e^{QL^2} \quad (23)$$

$$+ L^3\left(-\frac{8}{3}Q^{3/2} + 8\pi^{1/2}\right) + L^4\pi(Q^{-1/2} - 4Q^{1/2})e^{QL^2} + O(L^5).$$

In I the coefficients of L^{-5} , L^{-4} , L^{-2} are zero, the one of L^{-1} is $O(1)$, the coefficient of L^{-3} is negative, and the leading order in the one of L^0 is $8\frac{\pi}{3}Q^{3/2}$.

Finally, in both cases the macroscopic equation for $\langle u \rangle$ is

$$(\partial_t - D'_0 \nabla^2 - \eta \partial_t^{3/2} \nabla^2) \langle u \rangle(\mathbf{x}, t) = 0 \quad (24)$$

or in Fourier–Laplace variables

$$(q + D'_0 \mathbf{k}_0^2 + \eta q^{3/2} \mathbf{k}_0^2) \langle u \rangle(\mathbf{k}_0, q) = 1, \quad (25)$$

with the fractional derivative (see [21–23]) $\partial_t^{3/2}$ being defined by

$$\partial_t^{3/2} v(t) = \frac{1}{\Gamma(1/2)} \int_0^t \frac{d^2}{d\tau^2} v(\tau) (t - \tau)^{-1/2} d\tau + \frac{dv}{dt}(0+) \frac{t^{-1/2}}{\Gamma(1/2)} + v(0+) \frac{t^{-3/2}}{\Gamma(-1/2)}. \quad (26)$$

With this definition, connected with finite initial conditions, the Laplace transform of $\partial_t^{3/2} v$ is $q^{3/2} \hat{v}(q)$.

3.4 First moments of the concentration of solute

Integrating over R^3 shows that according to Eq. (24), the zero-order moment of $\langle u \rangle$, which represents the total amount of solute is constant. The second moment of $\langle u \rangle$ is readily obtained from Eq. (25). Indeed, the second derivative of $\langle u \rangle(\mathbf{k}_0, q)$ with respect to the first coordinate k_{0x} of $\mathbf{k}_0$ is $-2 \frac{D'_0 + \eta q^{3/2}}{(q + \mathbf{k}_0^2 (D'_0 + \eta q^{3/2}))}$, plus another fraction, which contains the factor k_{0x} . Hence the Laplace transform of the second moment of the concentration of solute is $D'_0 q^{-1} + \eta q^{1/2}$, so that the second moment itself is $D'_0 t + \eta t^{-1/2} / \sqrt{\pi}$. Since Eq. (24) holds in the macroscopic limit, the result may not be relevant in the very neighbourhood of $t = 0$. Non-normally diffusive

behaviours can be expected during transients, since then, the second moment is not proportional to time. In the long time limit, the diffusive term $D'_0 t$ dominates.

4 Conclusions

We considered a disordered porous medium, where the spreading of solute, dissolved in a fluid filling the pores, satisfies Fourier's law on the small scale. Here the solid matrix was assumed to be made of nearly spherical grains. To take account of disorder, we assumed the porosity to be a decreasing exponential of an isotropic Gaussian process. Upon averaging with respect to the process, we obtained that, on the macroscopic level, the concentration of solute satisfies Eq. 24, which contains a fractional derivative w.r.t. time, combined with the Laplacean. Non-normally diffusive behaviours are visible during transients. Nevertheless, in the statistically isotropic medium considered here, the second moment of the concentration becomes proportional to time much more rapidly than in media, organised around one direction, such as in [14].

Hence the geometrical structure of the medium influences the law, ruling the spreading of matter on the macroscopic scale. Many other possibilities occurring in rocks and soils still deserve being studied, with the tools presented in [24].

References

1. Matheron G, de Marsily G (1980) Is transport in porous media always diffusive? A counterexample, *Water Resour. Res.*, 16(5):901–917.
2. Muralidhar R, Ramkrishna D (1993) Diffusion in pore fractals, *Trans. Porous Media*, 13(1):79–95.
3. Gelhar LW (1993) *Stochastic Subsurface Hydrology*. Prentice-Hall, New Jersey, USA.
4. Hanyga A (2004) Two-fluid flow in a single temperature approximation, *Int. J. Eng. Sc.*, 42:1521–1545.
5. Montroll EW, West BJ (1965) Random walks on lattices II, *J. Math. Phys.* 6:167–181.
6. Comte A (1996) Stochastic foundations of fractional dynamics, *Phys. Rev.*, E, 53(4):4191–4193.
7. Henry BI, Wearne SL (2000) Fractional reaction diffusion, *Physica A*, 276:448–455.
8. Scalas E, Gorenflo R, Mainardi F (2004) Uncoupled continuous time random walk: solution and limiting behaviour of the master equation. *Phys. Rev. E*, 69(2):011107.
9. Metzler R, Klafter J (2004) The restaurant at the end of the random walk: recent developments in the description of anomalous transport by fractional dynamics, *J. Phys. A Math. Gen.*, 37:R161–R208.

10. Erochenkova G, Lima R (2000) On a tracer flow trough packed bed, *Physica A*, 275:297–309.
11. Erochenkova G, Lima R (2001) A fractional diffusion equation for a porous medium, *Chaos*, 11(3):495–499.
12. Klyatskin VI (1980) Stochastic equations and waves in randomly heterogeneous media (in Russian) Nauka, Moscow. Ondes et équations stochastiques dans les milieux aléatoirement non homogènes. 1985. Editions de Physique, Paris.
13. Furutsu K (1963) On the statistical theory of electromagnetic waves in fluctuating medium (I), *J. Res. Nat. Bur. Stand. D. Radio Propagation*, 67 D(3):303–323.
14. Logvinova K, Néel MC (2004) A fractional equation for anomalous diffusion in a randomly heterogeneous porous medium, *Chaos*, 14(4):982–987.
15. Zheltov YV, Morozov VP, Dutishev VN (1990) About the mechanism of thermocyclic intensifications of mass transfer into metall melts. *Izvestia Akademii Nauk USSR. Metalli* 5:31.
16. Whitaker S (1999) *The Method of Volume Averaging, Theory and Applications of Transport in Porous Media*. Kluwer Academic, Dordrecht.
17. Dean DS, Drummond IT, Horgan RR (1994) Perturbation schemes for flow in random media, *J. Phys. A Math. Gen.*, 27:5135–5144.
18. Bouchaud JP, Georges A (1990) Anomalous diffusion in disordered media: statistical mechanisms, models and physical applications, *Phys. Rep.*, 195 (4–5):127–293.
19. Stepanyants YA, Teodorovich EV (2003) Effective hydraulic conductivity of a randomly heterogeneous porous medium, *Water Resour. Res.*, 39(3):12 (1–11).
20. Kleinert H (1989) *Gauge Fields in Condensed Matter*, Vol. I. World Scientific, Singapore.
21. Samko SG, Kilbas AA, Marichev OI (1993) *Fractional Integrals and Derivatives: Theory and Applications*. Gordon and Breach, New York.
22. Gorenflo R, Mainardi F (1997) Fractional calculus, integral and differential equations of fractional order. In: Carpinteri A, Mainardi F (eds.), *Fractals and Fractional Calculus in Continuum Mechanics*. CISM courses and lectures 378. Springer, New York, pp. 223–276.
23. Podlubny I (1999) *Fractional Differential Equations*. Academic Press, San Diego.
24. Hilfer R (2002) Review on scale dependent characterization of the micro-structure of porous media, *Trans. Porous Media*, 46:373–390.

FRACTIONAL ADVECTIVE–DISPERSIVE EQUATION AS A MODEL OF SOLUTE TRANSPORT IN POROUS MEDIA

F. San Jose Martinez^{1,2}, Y. A. Pachepsky², and W. J. Rawls³

¹ Department de Matematica Aplicada, ETSIA-UPM, Avd. de la Complutense s/n. 28040, Madrid, Spain; E-mail: f.sanjose@upm.es, fsanjose@anri.barc.usda.gov

² Environmental Microbial Safety Laboratory, USDA-ARS-BA-ANRI-EMSL, Beltsville, MD 20705; E-mail: ypachepsky@anri.barc.usda.gov

³ Hydrology and Remote Sensing Laboratory, USDA-ARS-BA-ANRI-RSL, Beltsville, MD 20705; E-mail: wrawls@hydrolab.arsusda.gov

Abstract

Understanding and modeling transport of solutes in porous media is a critical issue in the environmental protection. The common model is the advective–dispersive equation (ADE) describing the superposition of the advective transport and the Brownian motion in water-filled pore space. Deviations from the advective–dispersive transport have been documented and attributed to the physical heterogeneity of natural porous media. It has been suggested that the solute transport can be modeled better assuming that the random movement of solute is the Lévy motion rather than the Brownian motion. The corresponding fractional advective–dispersive equation (FADE) was derived using fractional derivatives to describe the solute dispersion. We present and discuss an example of fitting the FADE numerical solutions to the data on chloride transport in columns of structured clay soil. The constant concentration boundary condition introduced a substantial mass balance error then the solute flux boundary condition was used. The FADE was a much better model compared to the ADE to simulate chloride transport in soil at low flow velocities.

Keywords

Fractional derivative, fractional advective–dispersive equation, solute transport, water quality, porous Media.

1 Introduction

Understanding and modeling transport of solutes in porous media is a critical issue in the environmental protection. Contaminants from various industrial and agricultural sources can travel in soil and ground water and eventually affect human and animal health.

The advective–dispersive equation (ADE) is the commonly used model of solute transport in porous media [1]. This model assumes that the diffusion-like spread occurs simultaneously with the purely advective transport. The one-dimensional ADE is

$$\frac{\partial c}{\partial t} = D \frac{\partial^2 c}{\partial x^2} - v \frac{\partial c}{\partial x} \quad (1)$$

where c is the solute concentration, $[\text{ML}^{-3}]$, D is the dispersion coefficient, $[\text{L}^2\text{T}^{-1}]$, v is the average pore water velocity, $[\text{LT}^{-1}]$, x is the distance, $[\text{L}]$, and t is the time $[\text{T}]$.

In the last two decades, several studies have reported that the ADE could not satisfactory describe several important features of solute transport in soils. One of the assumptions of the ADE model is that the value of D remains constant if v is constant. Violations of this assumption have been found in both field and laboratory experiments [2–6]. The dispersion coefficient tended to increase with the distance of solute concentration observations. Pachepsky et al. [7] reported that the power law $D/v \propto x^m$ provided a good approximation of published data; the exponent m varied from 0.2 to 1.7. Because of the increase in the dispersivity D/v with the travel distance, the solute arrived to a given depth earlier than the ADE would predict with the dispersivity found from data at a smaller depth. Also, solute breakthrough curves (BTCs), i.e., dependencies of solute concentration at a given depth on time, had shapes different from those suggested by the ADE. Van Genuchten and Wierenga [8] drew attention to this discrepancy and used the term “tail” to describe the last part of the non-sigmoidal BTC. Heavy tails of the BTC, that is, concentrations approaching the asymptotic values more slowly than predicted by the ADE, were observed by several authors [9, 10]. This behavior was sometime referred to as the anomalous or the non-Fickian dispersion.

The ADE can be derived as the Fokker–Plank equation under the assumptions that (a) solute particles undergo Brownian motion, (b) particles are released simultaneously and do not affect each other, and (c) the probability of a particle to be found in a particular location at a particular time and the solute concentration are interchangeable variables [11]. In recent years, diffusion and dispersion phenomena have been studied within the

broad statistical framework of continuous time random walks (CTRW) first developed by Montroll and Weiss [12] and Scher and Lax [13, 14] and initially applied to electron movement in disordered semiconductors (see for instance Berkowitz et al. [15], and Metzler and Klafter [16] for panoramic reviews). The CTRW describes the solute transport as a result of particle motion via a series of steps, or transitions, through the porous media via different paths with spatially varying velocities. This kind of transport can, in general, be represented by a joint probability distribution that describes each particle transition over a distance and direction for a time interval. Identification of this joint probability distribution lies at the basis of the CTRW theory. Usually, this joint distribution is decoupled into two statistically independent probability density functions, one for the spatial transitions and another for the temporal transitions. The CTRW reduces to the Brownian motion and the ADE is recovered unless transition length or time distributions are heavy-tailed, meaning that the transition probability decreases according to a power law for large values of transitions. When only jump sizes (spatial transitions) have the power law probability density function, the process is called Lévy flight [17] and the particle motion is referred to as the Lévy motion. Thus, Lévy flights are the scaling limits of random walks with the power law transition probability. Particles undergoing the Lévy motion behave mostly like in the Brownian motion except that large jumps are more frequent. The path of a particle performing Lévy flights is a random fractal [18]. The short jumps making up Brownian motion create a clustered pattern that is so dense that area or volume is a more appropriate measure than length. Whereas the short jumps of the Lévy motion produce a similar clustering, the longer, less frequent jumps initiate new clusters. These clusters form a self-similar pattern with the fractal dimension between one and two in two-dimensional Euclidean space [18]. This type of motion may model, for example, the transport that may occur if particles are trapped for periods of time in relatively stagnant zones, and can travel occasionally within “jets” of high velocity fluid [19]. The Lévy motion predicts heavier tails in the BTC than those produced by the Brownian motion. It also predicts the growth of the solute spread, measured as the apparent variance of the solute particle distributions, faster than in Brownian motion. These features make Lévy motion an attractive generalization of Brownian motion when describing scale-dependent transport in porous media [20, 21, 7].

The ADE facilitated the application of the Brownian motion physical model to solute transport simulations. Similar benefits could be expected from a transport equation based on the Lévy motion as a physical model of solute particles transport. Zaslavsky [22] derived such an equation using fractional derivatives. This fractional advective–dispersive equation (FADE)

was later modified in order to include as solutions the full family of one-dimensional Lévy motions [23–27].

The FADE as a model to simulate solute transport in porous media has been applied to both laboratory and field-scale experiments. Benson [20] and Benson et al. [21] used an approximation of the solution of the initial value problem in the infinite domain for the Dirac delta as the initial distribution. The Lévy distribution function was used to obtain this approximation, similarly to obtaining the solution of the classical ADE using the error function. Benson [20] and Benson et al. [21] simulated a tracer test in a sandbox and the solute transport in a sand and gravel aquifer where a bromide (Br^-) tracer was injected. Pachepsky et al. [7] used the same solution to fit data from experiments on chloride (Cl^-) transport in sand, in structured clay soil and in columns made of soil aggregates. Solute transport in the macrodispersion experiment (MADE) conducted in a highly heterogeneous aquifer, was also simulated with this solution [28]. Lu et al. [29] applied a three-dimensional FADE derived from previous works of Meerschaert et al. [30, 31] to the MADE dataset. Zhou and Selim [32] discussed several issues regarding the application of the FADE and suggested a method to better estimate their parameters. Recently, Deng et al. [33] and Zhang et al. [34] applied the FADE to solute transport in river and overland flow.

Most of the applications of the FADE to the solute transport to-date rely on analytical solutions of the initial value problem in the infinite domain. However, in many practical applications initial-boundary value problems in the finite domain needs to be considered and numerical solution are required. Several methods had been developed to solve the FADE numerically. Lynch et al. [35] followed a method proposed by Oldham and Spanier [36] to obtain a numerical solution for a superdiffusive plasma transport equation. The fractional derivative of order α was replaced by the fractional integral of the order $\alpha - 2$ for the second derivative. This second derivative was approximated by the three-point centered finite difference formula. Liu et al. [37] approximated the FADE with a system of ordinary differential equations, which was then solved using backward difference formulas. Deng et al. [33] used the Grünwald definition of fractional derivatives and the split-operator method. Meerschaert and Tadjeran [38, 39] introduced the “shifted approximation” of the Grünwald fractional derivative that reduced to the standard centered finite difference formula for approximating the second derivative when the order of the derivative was two. Zhang et al. [34] proposed a semi-implicit scheme that was applied to simulate tracer movement in a stream and in the overland flow.

Solute transport is often studied in miscible displacement experiments that consist in displacing a tracer solution by the inflowing tracer-free solu-

tion, or vice versa, in columns made of porous media. The tracer is applied at one end of the column and solute BTCs are recorded. Solute transport parameters can be estimated by fitting the solution of the transport model to the BTC.

Data from miscible displacement experiments were used to show the applicability of FADE to the conservative transport in soils [7]. The authors used the Dirichlet, or constant concentration, boundary condition at the surface of the column. Earlier researches on the ADE applications to the miscible displacement experiments have shown that this boundary condition introduced mass balance errors in the solute transport simulations [40]. Recently a similar effect was observed for the FADE [41]. We hypothesized that the absence of mass conservation might substantially change values of estimated parameters and conclusions about the advantages of FADE over ADE in simulating solute transport in porous media. The purpose of this work was to test this hypothesis.

2 Theory

The one-dimensional FADE with symmetric dispersion [20, 26] is

$$\frac{\partial c}{\partial t} = -v \frac{\partial c}{\partial x} + \frac{1}{2} D_f \left(\frac{\partial^\alpha c}{\partial_+ x^\alpha} + \frac{\partial^\alpha c}{\partial_- x^\alpha} \right) \quad (2)$$

Here D_f is the fractional dispersion coefficient, $[L^\alpha T^{-1}]$, the superscript α is the order of fractional differentiation, $1 < \alpha \leq 2$, c is the relative concentration [-], v is the flow velocity, $[LT^{-1}]$, x is the distance from the inlet, and t is time. Fractional derivatives are integro-differential operators defined as [42]:

$$\frac{\partial^\alpha c}{\partial_+ x^\alpha}(x, t) = \frac{1}{\Gamma(m - \alpha)} \frac{\partial^m}{\partial x^m} \int_L^x (x - z)^{m - \alpha - 1} c(z, t) dz \quad (3)$$

for the left fractional derivative, and

$$\frac{\partial^\alpha c}{\partial_- x^\alpha}(x, t) = \frac{(-1)^m}{\Gamma(m - \alpha)} \frac{\partial^m}{\partial x^m} \int_x^R (z - x)^{m - \alpha - 1} c(z, t) dz \quad (4)$$

for the right fractional derivative. Here m is the integer such that $m - 1 < \alpha \leq m$, Γ is the gamma function and, L and R are real numbers.

Grünwald definitions of the left and the right fractional derivatives for $1 < \alpha < 2$ are, respectively,

$$\frac{\partial^\alpha c}{\partial_+ x^\alpha}(x, t) = \lim_{M_+ \rightarrow \infty} \frac{1}{h_+^\alpha} \sum_{k=0}^{M_+} g_k c(x - kh, t) \quad (5)$$

$$\frac{\partial^\alpha c}{\partial_- x^\alpha}(x, t) = \lim_{M_- \rightarrow \infty} \frac{1}{h_-^\alpha} \sum_{k=0}^{M_-} g_k c(x + kh, t) \quad (6)$$

where $M_\pm$ are positive integers, $h_+ = (x - L)/M_+$, $h_- = (R - x)/M_-$, and the Grünwald weights g_k are defined as

$$g_0 = 1, \quad g_k = (-1)^k \frac{\alpha(\alpha-1) \cdots (\alpha-k+1)}{k!}. \quad (7)$$

The Grünwald definitions can be used to discretize FADE to obtain numerical solutions. Let M be a nonnegative integer, h a real number such that $h = (R - L)/M$ and $x_i = L + ih$, $i = 0, 1, \dots, M$, for $L \leq x_i \leq R$; also $t_n = n\Delta t$, so that $c(x_i, t_n) = c_i^n$. The shifted Grünwald approximation of (5) and (6) [38, 39] are, respectively,

$$\frac{\partial^\alpha c}{\partial_+ x^\alpha}(x_i, t_n) \approx \frac{1}{h^\alpha} \sum_{k=0}^M g_k c(x_i - (k-1)h, t_n) \quad (8)$$

$$\frac{\partial^\alpha c}{\partial_- x^\alpha}(x_i, t_n) \approx \frac{1}{h^\alpha} \sum_{k=0}^M g_k c(x_i + (k-1)h, t_n). \quad (9)$$

This approximation can be used in either implicit

$$\frac{c_i^{n+1} - c_i^n}{\Delta t} = -v \frac{c_i^{n+1} - c_{i-1}^{n+1}}{h} + \frac{D_{sf}}{2h^\alpha} \left\{ \sum_{k=0}^M g_k c_{i-k+1}^{n+1} + \sum_{k=0}^M g_k c_{i+k-1}^{n+1} \right\} \quad (10)$$

or explicit

$$\frac{c_i^{n+1} - c_i^n}{\Delta t} = -v \frac{c_i^n - c_{i-1}^n}{h} + \frac{D_{sf}}{2h^\alpha} \left\{ \sum_{k=0}^M g_k c_{i-k+1}^n + \sum_{k=0}^M g_k c_{i+k-1}^n \right\} \quad (11)$$

finite difference schemes to solve (2) numerically. The implicit scheme is unconditionally stable, while the explicit scheme is stable under certain condition that constrain the size of the time step [38, 39].

When the analytical solutions of ADE are used with data from the miscible displacement experiments, the common approach is to use the solu-

tion for the semi-infinite domain, and to fit the solution at the distance of the column outlet to the experimental BTC [43]. We mimicked this approach using numerical solutions by setting the zero concentration at right boundary and moving the right boundary far enough, so that the concentration at this boundary would not be greater than 10^{-6} in the end of the transport simulations.

At the inlet, the nodal concentration in the boundary point was set to provide mass conservation. The solute mass that entered the transport domain at one time step, M_{ent} , and solute mass inside the column, M_{ins} , were computed, respectively, as

$$M_{ins} = 1 \cdot v \cdot \Delta t, \text{ and } M_{ins}^n = \sum_{i=0}^{M-1} c_i^n h + (c_0^n + c_M^n)(h/2). \quad (12)$$

The conservation of the mass required that

$$\varphi(c_0^{n+1}) = M_{ins}^{n+1} - M_{ins}^n - M_{ent} = 0. \quad (13)$$

With the explicit scheme, Eq. (13) can be explicitly solved to find c_0^{n+1} . With the implicit scheme, Eq. (13) should iteratively be solved with respect to the unknown c_0^{n+1} to obtain the boundary concentration at timestep $n+1$ from concentration at timestep n . Each iteration required solving a system of linear equations using the zero concentration at the right boundary and the concentration c_0^{n+1} at the left boundary. The FORTRAN subroutine RTBIS [44] was used in the iterative solution.

3 Materials and Methods

The applicability of fractional differential equation to solute transport in soil was tested with the data of Dyson and White [45] who studied Cl^- transport in structured clay soil irrigated with flow rates of 0.28 and 2.75 cm h^{-1} . Soil cores 16.4 ± 1.5 cm long were irrigated from 16 evenly spaced hypodermic needles set above the soil surface. A steady-state near-saturated flow was created. The initial volumetric water content was $0.52 \pm 0.07 \text{ cm}^3 \text{ cm}^{-3}$, the saturated water content was estimated as $0.67 \pm 0.02 \text{ cm}^3 \text{ cm}^{-3}$, and the steady-state water content in soil column was 0.59 to $0.62 \text{ cm}^3 \text{ cm}^{-3}$. Soil was irrigated with 10 mM CaSO_4 solution to reach steady-state water flow and the CaCl_2 was applied at same intensity afterwards. The BTC data points were obtained by digitizing graphs found in the aforementioned publication. The digitizing was made in triplicate. Coefficient of variation within the replications did not exceed 0.1%.

In order to estimate parameters α , D_f , and v of FADE we used a version of the Marquardt–Levenberg algorithm to minimize the root-mean squared error:

$$RMSE = \sqrt{\sum_{j=1}^N (c_j^{calc} - c_j^{meas})^2 / N} \quad (15)$$

where N is the number of observations. The observed concentrations of the BTCs were normalized by the influent concentration to get the relative concentration c_j^{meas} . The range of α 's ($1 < \alpha \leq 2$) was scanned in increments of 0.05 to detect possible local minima of RMSE.

Parameters and RMSE values were computed for the mass conserving boundary condition and for the Dirichlet boundary condition $c_0^n = 1$ at the inlet. We used the explicit finite difference scheme because iterations combined with inversions of large (200×200) matrices made the optimization with the implicit scheme somewhat impractical. The numerical solutions with the explicit and the implicit schemes were compared when the “best” (in terms of RSME) parameter sets were used. The difference between the simulated concentrations at the breakthrough curves did not exceed 0.3%.

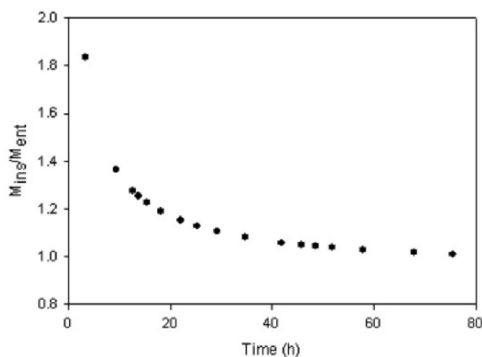


Fig. 1. Mass conservation violation in transport simulations with the Dirichlet boundary condition at the surface of the soil column; M_{ent} is the mass of solute expected to enter the column, M_{ins} is the simulated solute mass in the soil column; transport parameters $\alpha = 1.95$, $D = 1.89 \text{ cm}^{1.7} \text{ h}^{-1}$, $v = 0.66 \text{ cm h}^{-1}$.

4 Results and Discussion

Results of simulations illustrated the need of using the mass conserving boundary condition. Using the constant concentration boundary condition at the inlet led to the substantial overestimation of the solute mass in the porous medium, especially at early stages of the solute transport (Fig. 1). Numerical experiments showed that the ratio of the simulated mass in soil to the mass expected to enter soil column from the top increased with the increases in the fractional dispersion coefficient, and in the order of the fractional derivative.

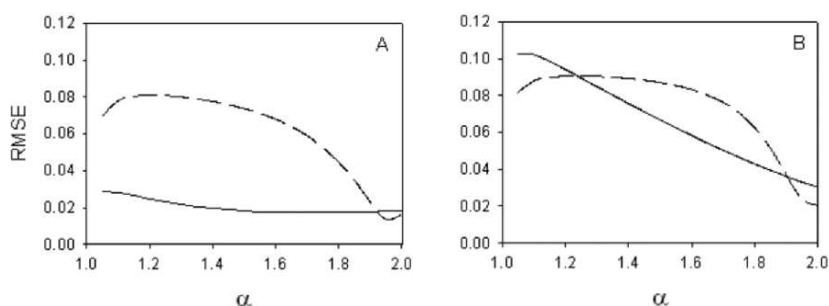


Fig. 2. Dependencies of the root-mean square error of the breakthrough simulations on the order of the fractional derivative α ; solid line – FADE with the mass conserving boundary condition, dashed line – FADE with the Dirichlet boundary condition; A and B – experiments of Dyson and White [45] with flow rates of 0.28 and 2.75 cm h⁻¹, respectively.

The root-mean square errors in simulations with optimized values of the dispersion coefficient D and pore water flow velocity v are shown in Fig. 2. The FADE is a substantially better model as compared with ADE in the case of the “slow” experiment with the flow rate of 0.28 cm h⁻¹. The minimum RMSE is reached at the values of the order fractional derivative α distinctly different from two. These values were 1.71 with the mass conserving boundary condition and 1.95 with the constant boundary condition. ADE was the better model in the case of the “fast” experiment with the flow rate of 2.75 cm h⁻¹. Interestingly, the RMSE was smaller for the simulations with the constant concentration boundary condition for both experiments.

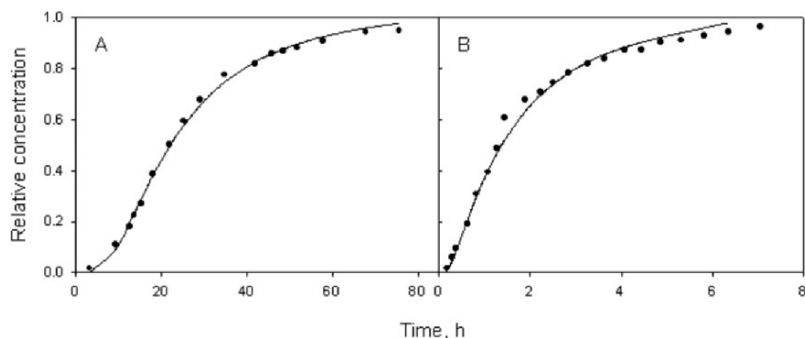


Fig. 3. Simulated (lines) and measured (circles) breakthrough curves with optimized transport parameter values; A and B – experiments of Dyson and White [45] with flow rates of 0.28 and 2.75 cm h⁻¹, respectively.

The best-fit simulated BTCs are compared with measured in Fig. 3. The FADE with $\alpha = 1.71$ gave a reasonable fit of data in Fig. 3A corresponding to the “slow-flow” experiment. Data in Fig. 3B show that ADE, albeit having the RMSE values smaller than FADE with $\alpha < 2$, is not actually a good model of the solute transport in this case. The experimental BTC has an early steep rise and a fairly long tail that the advective-dispersive models (ADE or FADE) cannot simulate properly. In this “fast-flow” experiment, the observed BTC is typical of solute transport with porous space with mobile and immobile zones [46]. This type of transport assumes the existence of the mobile zone of pore space where an advective-dispersive transport can develop, and the immobile zone where the solute particles can enter from the mobile zone because of pseudo-diffusion and where the solute particles cannot move in the general direction of the flow in the column. The steep rise section of the BTC can be attributed to the fast transport in the mobile zone, whereas the tail emerges because of diffusion based mass exchange between mobile and immobile zones.

Figure 4 shows the dependencies of the optimized values of the dispersion coefficient and pore water flow velocity on the order of the fractional derivative. In the “slow-flow” experiment, the dispersion coefficient tends to grow and the velocity tends to decrease as the α value increases from 1.4 to 2. The decrease in velocity means the later arrival of the solute to the outlet, and the increase in the dispersion coefficient serves to compensate this delay and the less heavier tails that correspond to the increase in α .

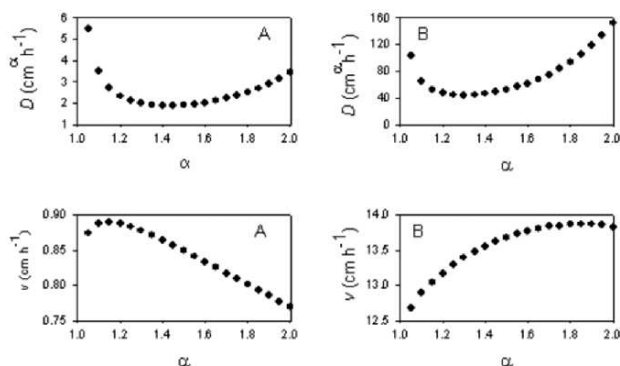


Fig. 4. Dependencies of the optimized values of dispersion coefficient D and pore water velocity v on the order of the fractional derivative α ; A and B – experiments of Dyson and White [45] with flow rates of 0.28 and 2.75 cm h^{-1} , respectively.

5 Conclusions

The constant concentration boundary condition introduced a substantial mass balance error in the solute breakthrough simulations. Using the mass-conserving boundary condition did not change the general conclusion about the advantage of the FADE compared with the classical ADE. However, the estimated optimized value of the order of the fractional derivative was different with the mass conserving boundary condition.

The FADE was a good model to simulate chloride transport in structured clay soil at low flow velocities. However, experimental BTCs could not be well simulated with FADE when the flow velocities were relatively large in this soil. A physical model different from the Lévy motions may be needed to simulate such transport.

Acknowledgment

Fernando San Jose Martinez was supported in part by a grant of Secretaria de Estado de Universidades e Investigacion (Ministerio de Educacion y Ciencia, Spain) and the Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica (I+D+I) under ref. AGL2004–04079 AGR. Spain.

References

1. Bear J (1972) *Dynamics of Fluids in Porous Media*. Dove Publications, New York.
2. Jury WA (1988) Solute transport and dispersion. In: Steffen WL, Denmead OT (eds.), *Flow and transport in the Natural Environment: Advances and Applications*. Springer, Berlin, pp. 1–16.
3. Khan AUH, Jury WA (1990) A laboratory test of the dispersion scale effect, *J. Contam. Hydrol.*, 5:119–132.
4. Porro I, Wierenga PJ, Hills RG (1993) Solute transport through large uniform and layered soil columns, *Water Resour. Res.*, 29:1321–1330.
5. Snow VO, Clothier BE, Scotter DR, White RE (1994) Solute transport in a layered field soil: experiments and modelling using the convection–dispersion approach, *J. Contam. Hydrol.*, 16:339–358.
6. Yasuda H, Berndtsson R, Barri A, Jinno K (1994) Plot-scale solute transport in a semiarid agricultural soil, *Soil Sci. Soc. Am. J.*, 58:1052–1060.
7. Pachepsky Ya, Benson DA, Rawls W (2000) Simulating scale-dependent solute transport in soils with the fractional advective-dispersive equation, *Soil Sci. Soc. Am. J.*, 64:1234–1243.
8. Zhang R, Huang K, Xiang J (1994) Solute movement through homogeneous and heterogeneous soil columns, *Adv. Water Resour.*, 17:317–324.
9. Nielsen DR, Van Genuchten MTh, Biggar JW (1986) Water flow and solute transport processes in the unsaturated zone, *Water Resour. Res.*, 22(9, Suppl.): 89S–108S.
10. Vachaud G, Vauclin M, Addiscott TM (1990) Solute transport in the vadose zone: a review of models. In *Proceedings of the International Symposium on Water Quality Modeling of Agricultural Non-Point Sources*, Part 1, 19–23 June 1988 pp. 81–104. Logan, UT. USDA-ARS. U.S. Government Printing Office, Washington, DC.
11. Bhattacharya R, Gupta VK (1990) Application of the central limit theorem to solute transport in saturated porous media: from kinetic to field scales. In: Cushman, JH (eds.), *Dynamics of Fluids in Hierarchical Porous Media*. Academic Press, New York, pp. 97–124.
12. Montroll EW, Weiss GH (1965) Random walks on lattices. II, *J. Math. Phys.*, 6:167–183.
13. Sher H, Lax M (1973a) Stochastic transport in a disordered solid. I. Theory, *Phys. Rev. B*, 7:4491–4502.
14. Sher H, Lax M (1973b) Stochastic transport in a disordered solid. II. Impurity conduction, *Phys. Rev. B*, 7:4502–4519.
15. Berkowitz B, Klafter J, Metzler, Scher H (2002) Physical pictures of transport in heterogeneous media: advection-dispersion, random-walk, and fractional derivative formulations, *Water Resour. Res.*, 38(10) W1191, doi: 10.1029/2001WR001030.
16. Metzler R, Klafter J (2004) The restaurant at the end of the random walk: recent developments in the description of anomalous transport by fractional dynamics, *J. Phys. A Math. Gen.*, 37:R161–R208, doi:10.1088/0305-4470/37/31/R01.

17. Shlesinger MF, Klafter J, Wong YM (1982) Random walks with infinite spatial and temporal moments, *J. Stat. Phys.*, 27:499–512.
18. Taylor SJ (1986) The measure theory of random fractals, *Math. Proc. Camb. Phil. Soc.*, 100(3):383–406.
19. Weeks ER, Solomon TH, Urbach JS, Swinney HL (1995) Observation of anomalous diffusion and Lévy flights. In: Shlesinger MF, Zaslavsky GM, Frisch U (eds.), *Lévy Flights and Related Topics in Physics*. Springer, New York, pp. 51–71.
20. Benson DA (1998) The fractional advection-dispersion equation: development and application, PhD Thesis, University of Nevada, Reno.
21. Benson DA, Wheatcraft SW, Meerschaert MM (2000) Application of a fractional advection-dispersion equation, *Water Resour. Res.*, 36(6):1403–1412.
22. Zaslavsky GM (1994) Renormalization group theory of anomalous transport in systems with Hamiltonian chaos, *Chaos*, 4(1):25–33.
23. Compte A (1996) Stochastic foundations of fractional dynamics, *Phys. Rev. E*, 53(4):4191–4193.
24. Compte A (1997) Continuous time random walks on moving fluids, *Phys. Rev. E*, 55(6):6821–6831.
25. Saichev AI, Zaslavsky GM (1997) Fractional kinetic equations: solutions and applications, *Chaos*, 7:753–764.
26. Chaves AS (1998) A fractional diffusion equation to describe Lévy flights, *Phys. Lett. A*, 239:13–16.
27. Metzler R, Klafter J, Sokolov IM (1998) Anomalous transport in external fields: continuous time random walks and fractional diffusion equations extended, *Phys. Rev. E*, 58:1621–1633.
28. Benson DA, Schumer R, Meerschaert MM, Wheatcraft SW (2001) Fractional dispersion, Levy motion, and the MADE tracer tests, *Transp. Porous Media*, 42:211–240.
29. Lu S, Molz FJ, Fix GJ (2002) Possible problems of scale dependency in applications of the three-dimensional fractional advection-dispersion equation to natural porous media, *Water Resour. Res.* 38(9):1165, doi:10.1029/2001WR000624.
30. Meerschaert MM, Benson DA, Bäumer B (1999) Multidimensional advection and fractional dispersion, *Phys. Rev. E*, 59(5):5026–5028.
31. Meerschaert MM, Benson DA, Bäumer B (2001) Operator Levy motion and multiscaling anomalous diffusion, *Phys. Rev. E*, 63(2):1112–1117.
32. Zhou LZ, Selim HM (2003) Application of the fractional advection-dispersion equation in porous media. *Soil Sci. Soc. Am. J.*, 67(4):1079–1084.
33. Deng Z-Q, Singh VP, Bengtsson L (2004) Numerical solution of fractional advection-dispersion equation, *ASCE J. Hydraulic Eng.*, 130(5):422–431.
34. Zhang X, Crawford JW, Deeks LK, Stutter MI, Bengough AG, Young IM (2005) A mass balance based numerical method for the fractional advection-dispersion equation: theory and application, *Water Resour. Res.*, 41, W07029, doi: 10.1029/2004WR003818.

35. Lynch VE, Carreras BA, del-Castillo-Negrete D, Ferreira-Mejias KM, Hicks HR (2003) Numerical method for the solution of partial differential equations with fractional order. *J. Comp. Phys.*, 192:406–421.
36. Oldham KB, Spanier J (1974) *The Fractional Calculus*. Academic Press, New York.
37. Liu F, Anh V, Turner I (2004) Numerical solution of the space fractional Fokker-Planck equation, *J. Comp. Appl. Math.*, 166:209–219.
38. Meerschaert MM, Tadjeran C (2004) Finite difference approximation for fractional advection-dispersion flow equations, *J. Com. Appl. Math.*, 172:65–77.
39. Meerschaert MM, Tadjeran C (2006) Finite difference approximations for two-sided space-fractional partial differential equations, *Appl. Num. Math.*, 56:80–90.
40. van Genuchten MTh, Parker JC (1984) Boundary conditions for displacement experiments through short soil columns, *Soil Sci. Soc Am J.*, 48:703–708.
41. San Jose MF, Pachepsky Ya, Rawls W (2005) Solute transport simulated with the fractional advective-dispersive equation. In: Agrawal O, Tenreiro Machado JA, Sabatier J. (eds.), *Fractional Derivatives and Their Applications*. IDECT/CIE2005. ISBN: 0-7918-3766-1.
42. Samko SG, Kilbas AA, Marichev OI (1993). *Fractional Integrals and Derivatives: Theory and Applications*. Gordon and Breach Science, New York.
43. Toride N, Leij FJ, van Genuchten MTh (1995) The CXTFIT code for estimating transport parameters from laboratory or field tracer experiments. Version 2.0. Research Report 137, US Salinity Lab, Riverside, CA.
44. Press WH, Teukolsky SA, Vetterling WT, Flannery, BP (1992) *Numerical Recipes in FORTRAN 77. The Art of Computing*, 2nd edition. Cambridge University Press, New York.
45. Dyson JS, White RE (1987) A comparison of the convective-dispersive equation and transfer function model for predicting chloride leaching through an undisturbed structured clay soil, *J. Soil Sci.*, 38:157–172.
46. van Genuchten MTh, Wierenga PJ (1976) Mass transfer in sorbing porous media. I. Analytical solutions, *Soil Sci. Soc. Am. J.*, 40:473–481.

MODELLING AND IDENTIFICATION OF DIFFUSIVE SYSTEMS USING FRACTIONAL MODELS

Amel Benchellal, Thierry Poinot, and Jean-Claude Trigeassou

Laboratoire d'Automatique et d'Informatique Industrielle, 40 Avenue du Recteur
Pineau, 86022 Poitiers Cedex France; E-mail: amel.benchellal@etu.univ-poitiers.fr,
Thierry.poinot@esip.univ-poitiers.fr, Jean-Claude.Trigeassou@esip.univ-poitiers.fr

Abstract

Heat transfer problems obey to diffusion phenomenon. They can be modelled with the help of fractional systems. The simulation of these particular systems is based on a fractional integrator where the non-integer behaviour acts only on a limited spectral band. Starting from frequential considerations, a more general approximation of the fractional system is proposed in this communication. It makes it possible to define a state-space model for simulation of transients, and to carry out an output-error (OE) technique in order to estimate the parameters of the model. A real application on a thermal system is presented to illustrate the advantages of the proposed model.

Keywords

Fractional systems, fractional operator, modelling, estimation, output-error identification, heat transfer, diffusive interfaces.

1 Introduction

The model of a diffusive interface is characterized by a fractional behaviour. Concretely, such phenomenon appears, in the case of an induction machine, with Foucault currents inside rotor bars [2, 3, 5, 13]. It appears also in the case of heat transfer between the flux and the temperature at the interface of the process [1, 4]. Many solutions have already been developed in order to model this type of phenomenon. In order to improve the approximation of these diffusive interfaces using fractional models, an improved solution is proposed in this paper, based on the use of a fractional integrator operator

[6, 7, 9, 14]. The application of this modelling is performed firstly using numerical simulation, then on a thermal pilot thinks to its identification.

This paper begins by the definition of the diffusive interface and by the justification of its approximation using a fractional model. Some modelling techniques are recalled. Then, we present the fractional integration and its application to the simulation and the identification of a non-integer order model. A new model with two fractional integrators is then presented and tested in simulation. Finally, an application on a thermal pilot permits to validate the interest of this improved model.

2 Problem Position

2.1 Approximate modelling of a diffusive interface

Let us consider the classical “wall” problem for heat transfer [1], represented in Fig. 1.

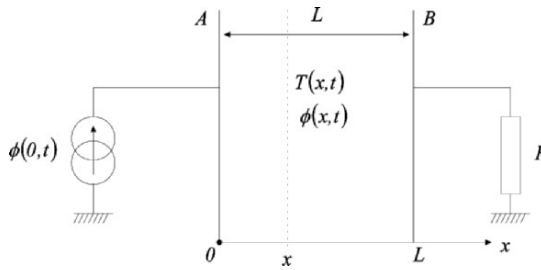


Fig. 1. Wall problem for heat transfer.

Temperature $T(x, t)$ is assumed to be uniform on any plane parallel to the faces A and B . Let $\phi(x, t)$ be the heat flux passing through the wall at abscissa x . $T(x, t)$ and $\phi(x, t)$ are governed by heat diffusion equations (1) and (2).

$$\frac{\partial T(x, t)}{\partial t} = \alpha \frac{\partial^2 T(x, t)}{\partial x^2} \quad (1)$$

$$\phi(x, t) = -\lambda \frac{\partial T(x, t)}{\partial x} \quad (2)$$

with

- $\alpha = \frac{\lambda}{\rho c}$: diffusivity,
- λ : thermal conductivity,
- ρ : density,
- c : specific heat.

2.2 Diffusive interface

Equations (1) and (2) specify the relation between $\phi(x, t)$ and $T(x, t)$, respectively considered as system input and output when $x = 0$, which define the diffusion interface. The boundary conditions on the faces A and B are:

$$\begin{cases} \phi(0, t) = u(t) \\ \phi(L, t) = \frac{T(L, t)}{R} \end{cases} \quad (3)$$

where R is the thermal resistance between the wall and the air. Because the model is modelled around an operating point, air temperature is assumed to be constant and equal to zero.

Thus, the modelling of this interface is equivalent to the determination of the transfer function $H(s)$ between $Y(s)$ and $U(s)$ (where $Y(s)$ and $U(s)$ are the Laplace transforms of $y(t)$ and $u(t)$):

$$H(s) = \frac{\lambda R \sqrt{\frac{s}{\alpha}} + 1 + (\lambda R \sqrt{\frac{s}{\alpha}} - 1) e^{-\frac{s}{\alpha} L^2}}{\lambda \sqrt{\frac{s}{\alpha}} (\lambda R \sqrt{\frac{s}{\alpha}} + 1 - (\lambda R \sqrt{\frac{s}{\alpha}} - 1) e^{-\frac{s}{\alpha} L^2})} \quad (4)$$

Let us consider that heat flux $\phi(0, t)$ is a step input whose value is ϕ . Then:

$$T(0, s) = H(s) \frac{\phi}{s} \quad (5)$$

If we consider $t \rightarrow \infty$ (or equivalently $s \rightarrow 0$) we get

$$T(0, \infty) = y(\infty) = R\phi \quad (6)$$

i.e., that the wall behaves like a thermal resistance equal to zero.

Reciprocally, at very short times ($t \rightarrow 0$ or $s \rightarrow \infty$) we get

$$H(s) \simeq \frac{\sqrt{\alpha}}{\lambda s^{0.5}} \quad (7)$$

i.e., that the wall behaves like a non-integer integrator whose order is equal to 0.5.

Remark: this phenomenon is not restricted to the heat diffusion, it is also observed in the case of induced currents in the rotor bars of an induction machine. A numerical simulation using finite elements [5] has permitted to give the frequency response of this phenomenon (see Fig. 2). One can verify that for $\omega \rightarrow \infty$, order n tends to 0.5, characterizing diffusion phenomena. On the other hand, the geometry of the bars appears at intermediary frequency: on this example, the phase exceeds -45° , i.e., that n is higher than 0.5 in the concerned frequency domain.

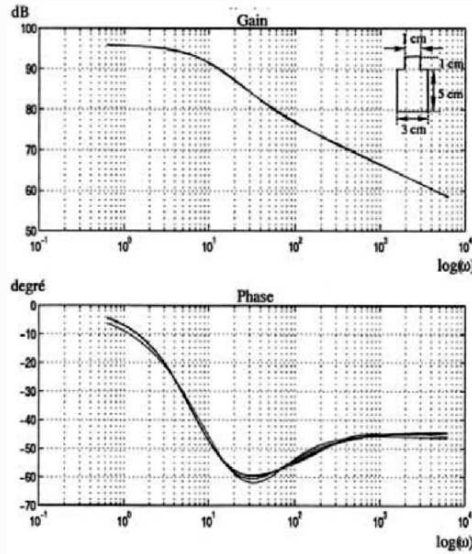


Fig. 2. Bode diagram of induced currents in rotor bars of an induction machine [5].

2.3 The fractional integrator approach

Many solutions have been developed to model this type of phenomenon [4, 11]. Our solution consists to use a fractional model like [7, 14]:

$$H(s) = \frac{b_0}{a_0 + s^n} \quad (8)$$

This model is based on the use of a non-integer integrator $\frac{1}{s^n}$ truncated in the frequency domain. The fractional order n is fitted by identification, using time responses; so, the frequential approximation of the diffusion interface is indirectly performed. This approximation is accurate in a frequency domain corresponding to the spectrum of the input. Concretely, n is estimated in such a way that the frequency response is correctly fitted in low and medium frequencies. On the other hand, this model cannot give satisfactory results for $\omega \rightarrow \infty$ when estimated n is different of the value 0.5. Nevertheless, the interest of this model [6, 7] is its ability to approximate the dynamical behaviour of diffusion interfaces using a restricted number of estimated parameters. However, this model does not give the best approximation of the system dynamics, since $\theta \rightarrow -n 90$ then $\omega \rightarrow \infty$.

In order to improve the fractional behaviour of this model, and particularly its high frequency behaviour (quick transients), a second approach is proposed in this paper; it consists to use a model with two fractional integrators:

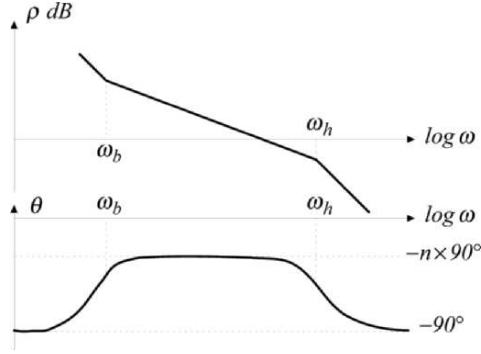


Fig. 3. Bode diagram of the fractional integrator.

$$H_{n_1, n_2}(s) = \frac{b_0 + b_1 s^{n_1}}{a_0 + a_1 s^{n_1} + s^{n_1 + n_2}} \quad (9)$$

In imposing $n_2 = 0.5$, and then adjusting the order n_1 and the parameters b_0 , b_1 , a_0 , and a_1 , one can get a higher approximation ability, with respect to the physics (order $n_1 + n_2 - n_1 = 0.5$ at short times) and able to fit to the system geometry thanks to n_1 .

3 Modelling Using a Fractional Integrator

3.1 Fractional integrator [6, 7, 14]

Let us consider the Bode diagram of an integrator truncated in low and high frequencies (Fig. 3).

It is composed of three parts. The intermediary part corresponds to non-integer action, characterized by the order n . In the two other parts, the integrator has a conventional action, characterized by its order 1.

In this way, the operator $I_n(s)$ is defined as a conventional integrator, except in a limited band $[\omega_b, \omega_h]$ where it acts like $\frac{1}{s^n}$. The operator $I_n(s)$ is defined using a fractional phase-lead filter [9] and an integrator $\frac{1}{s}$:

$$I_n(s) = \frac{G_n}{s} \prod_{i=1}^N \frac{1 + \frac{s}{\omega'_i}}{1 + \frac{s}{\omega_i}} \quad (10)$$

This operator is completely defined by:

$$\omega_i = \alpha \omega'_i, \quad \omega'_{i+1} = \eta \omega_i, \quad n = 1 - \frac{\log \alpha}{\log \alpha \eta}$$

where α and η are recursive parameters linked to the fractional order n .

Using (10), the corresponding state-space representation is:

$$\dot{\underline{x}}_I = A_I^* \underline{x}_I + \underline{B}_I^* u \quad (11)$$

where

$$\begin{cases} A_I^* = M_I^{-1} A_I \\ \underline{B}_I^* = M_I^{-1} \underline{B}_I \end{cases}$$

and

$$M_I = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ -\alpha & 1 & & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & -\alpha & 1 \end{bmatrix} \quad A_I = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ \omega_1 & -\omega_1 & & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & \omega_N & -\omega_N \end{bmatrix}$$

$$\underline{B}_I^T = [G_n \ 0 \ \cdots \ 0] \quad \underline{x}_I^T = [x_1 \ x_2 \ \cdots \ x_{N+1}]$$

3.2 State-space model of $H_n(s)$

The model (8) corresponds to a differential equation, with $0 < n < 1$:

$$\frac{d^n y(t)}{dt^n} + a_0 y(t) = b_0 u(t) \quad (12)$$

Let us define $x(t)$ such as

$$X(s) = \frac{1}{s^n + a_0} U(s) \quad (13)$$

Thus, we obtain a “macro” state-space representation of this system.

$$\begin{cases} \frac{d^n x(t)}{dt^n} = -a_0 x(t) + u(t) \\ y(t) = b_0 x(t) \end{cases} \quad (14)$$

or equivalently using $I_n(s)$

$$\begin{cases} \dot{x}_1 = G_n (-a_0 x_{N+1} + u) \\ y = b_0 x_{N+1} \end{cases} \quad (15)$$

In this simple example, $\underline{x} = \underline{x}_i$. Then, the global model is:

$$\begin{cases} \dot{\underline{x}} = A \underline{x} + \underline{B} u \\ y = \underline{C}^T \underline{x} \end{cases} \quad (16)$$

with

$$\begin{cases} A = A_I^* - a_0 \underline{B}_I \underline{C}_I^T \\ \underline{B} = \underline{B}_I^* \\ \underline{C}^T = b_0 \underline{C}_I^T \end{cases}$$

3.3 Output-error identification of the fractional system $H_n(s)$

The model of $H_n(s)$ is in continuous time representation, thus it is preferable to use an output-error (OE) technique to estimate its parameters [8, 12]. The state-space model of the non-integer system is:

$$\begin{cases} \dot{\underline{x}} = A(\underline{\theta}) \underline{x} + \underline{B}(\underline{\theta}) u \\ y = \underline{C}^T(\underline{\theta}) \underline{x} \end{cases} \quad (17)$$

where $\underline{\theta}^T = [a_0 \ b_0 \ \alpha]$

Remark: the fractional order n is characterized by $\alpha, \eta, \omega_b, \omega_h$, and the number N of cells. In practice, ω_b, ω_h , and N are imposed; then, it is sufficient to estimate α in order to estimate n .

Let us suppose that we have K data pairs $\{u_k, y_k^*\}$ where $t = k T_e$ (T_e : sampling period); y_k^* : noised measurement of the exact output y_k .

The state-space model is simulated using a numerical integration technique; thus one gets $\hat{y}_k(u, \hat{\underline{\theta}})$ where $\hat{\underline{\theta}}$ is an estimation of exact parameters $\underline{\theta}$. Then, one can construct the residuals:

$$\varepsilon_k = y_k^* - \hat{y}_k(u, \hat{\underline{\theta}}) \quad (18)$$

The optimal value of $\hat{\underline{\theta}}$, i.e., $\underline{\theta}_{opt}$, is obtained by minimization of the quadratic criterion:

$$J = \sum_{k=1}^K \varepsilon_k^2 \quad (19)$$

Because $\hat{y}_k$ is non-linear in parameters, a non-linear programming algorithm is used in order to estimate iteratively $\hat{\underline{\theta}}$:

$$\underline{\theta}_{i+1} = \underline{\theta}_i - \left\{ [J''_{\theta\theta} + \lambda I]^{-1} \underline{J}'_{\theta} \right\}_{\hat{\underline{\theta}} = \underline{\theta}_i} \quad (20)$$

with [8]:

- $\underline{J}'_{\theta} = -2 \sum_{k=1}^K \varepsilon_k \underline{\sigma}_{k, \underline{\theta}_i}$: gradient,
- $J''_{\theta\theta} \approx 2 \sum_{k=1}^K \underline{\sigma}_{k, \underline{\theta}_i} \underline{\sigma}_{k, \underline{\theta}_i}^T$: hessian,
- λ : monitoring parameter,
- $\underline{\sigma}_{k, \underline{\theta}_i} = \frac{\partial \hat{y}_k}{\partial \underline{\theta}_i}$: output sensitivity function.

This algorithm, also known as the Marquardt's algorithm [10] insures robust convergence, even with a bad initialization of $\hat{\underline{\theta}}$, nevertheless in the vicinity of the global optimum.

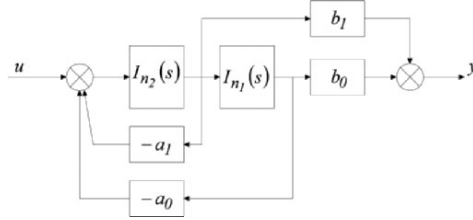


Fig. 4. Simulation scheme of $H_{n_1, n_2}(s)$.

Fundamentally, this technique is based on the calculation of gradient and hessian, themselves dependant on the numerical integration of the sensitivity functions $\underline{\sigma}_{k, \underline{\theta}_i}$ [12], which are equivalent to the regressors in the linear case [8].

4 Model with Two Fractional Integrators

4.1 Principle

The objective is to improve the approximation of the diffusion interfaces using fractional models based on the fractional integrator. The model (8) gives a good approximation only at low frequencies. This model has an asymptotic behaviour of $\frac{1}{s^n}$ type while the theoretical modelling of a diffusive system shows an asymptotic behaviour of $\frac{1}{s^{0.5}}$ type.

Therefore, a second approach using two fractional integrators is proposed (model (9)). The model is simulated using the scheme given on Fig. 4. Because $n_2 = 0.5$, this model permits to obtain $n \rightarrow 0.5$ then $\omega \rightarrow \infty$.

4.2 State-space model of $H_{n_1, n_2}(s)$

The macro state-space representation of $H_{n_1, n_2}(s)$ is given by:

$$\begin{cases} \frac{d^{n_1} x_1(t)}{dt^{n_1}} = x_2(t) \\ \frac{d^{n_2} x_2(t)}{dt^{n_2}} = u(t) - a_0 x_1(t) - a_1 x_2(t) \\ y(t) = b_0 x_1(t) + b_1 x_2(t) \end{cases} \quad (21)$$

Using the same procedure that for the model (8), we obtain the state-space representation:

$$\begin{cases} \dot{\underline{x}} = \begin{bmatrix} A_{I_1}^* & \underline{B}_{I_1}^* \underline{C}_{I_2}^T \\ -\underline{B}_{I_2}^* a_0 \underline{C}_{I_1}^T & A_{I_2}^* - \underline{B}_{I_2}^* a_1 \underline{C}_{I_2}^T \end{bmatrix} \underline{x} + \begin{bmatrix} 0 \\ \underline{B}_{I_2}^* \end{bmatrix} u \\ y = [b_0 \underline{C}_{I_1}^T \quad b_1 \underline{C}_{I_2}^T] \underline{x} \end{cases} \quad (22)$$

where $(A_{I_1}^*, B_{I_1}^*)$ and $(A_{I_2}^*, B_{I_2}^*)$ are the matrices defining the two integrators $I_{n_1}(s)$ and $I_{n_2}(s)$ with

$$\underline{C}_{I_1}^T = \underline{C}_{I_2}^T = [0 \cdots 0 \ 1]$$

4.3 Identification of $H_{n_1, n_2}(s)$

Identification is performed using the OE technique where the parameters vector is defined by

$$\underline{\theta}^T = [a_0 \ a_1 \ b_0 \ b_1 \ \alpha_1]$$

Because order n_2 is imposed equal to 0.5, it is only necessary to estimate n_1 , i.e; the parameter α_1 . The identification technique is the previously defined, adapted to the model (9).

Remark: like in the case of the model $H_n(s)$, parameters ω_{b_1} , ω_{h_1} , and N_1 which correspond to $I_{n_1}(s)$ and parameters ω_{b_2} , ω_{h_2} , and N_2 which correspond to $I_{n_2}(s)$ are imposed. In addition, we take $\omega_{b_1} = \omega_{b_2}$, $\omega_{h_1} = \omega_{h_2}$, and $N_1 = N_2$.

4.4 Simulation example

This simulation has the objective to show that the model (8) is not able to perform a good frequency approximation on a large domain, unlike the model (9). The influence of the output noise is also tested on the quality of the approximation.

The experimental protocol consists to simulate the model (9) in the time domain in order to have noisy data, with a signal to noise ratio equal to 100 (ratio between the noiseless output variance and the noise one). A number of cells equal to 30 have been considered in order to simulate the fractional integrators on a frequency domain defined by $\omega_b = 10^{-5} \text{ rd/s}$ and $\omega_h = 10^5 \text{ rd/s}$. Then, models (8) and (9) are estimated. Numerical results are given in Table 1.

Table 1. Identification results

Parameters	exact	$H_n(s)$	$H_{n_1, n_2}(s)$
a_0	0.24	0.3218	0.4028
a_1	0.23	—	0.8945
b_0	0.0016	0.0023	0.0027
b_1	0.001	—	0.0013
$n_1(n)$	0.5	0.7669	0.6927

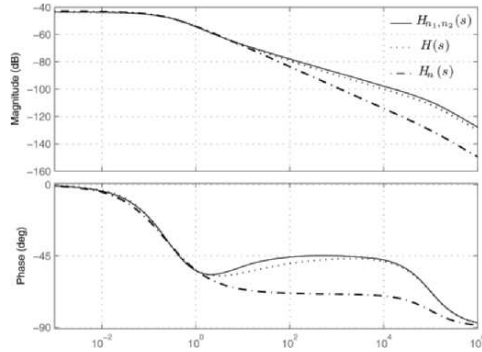


Fig. 5. Bode plots of simulated and estimated models.

In order to see the performed frequency approximations, frequency responses of the simulated and the two estimated models are plotted (Fig. 5)

As expected, the model (8) gives satisfactory results only at low and medium frequencies, which is obvious on the phase plot. The model (9) permits to obtain the initial frequency response, taking into account influence of the noise. Notice that the estimated parameters exhibit greater differences than the model values, which certainly means that these values are not individually critical, but their association is surely pertinent.

5 Application to a Thermal Pilot

After the validation in simulation of the model (9), we want to show that it permits to improve the time approximation of an experimental diffusive interface.

5.1 Description

The system is a copper ball with 3 cm radius (see Fig. 6). A power transistor is placed at the center of the ball in order to generate a heat flux. A sensor is fixed on the interface of the heat source. The ball is situated in an enclosure where the ambient temperature is fixed and constant. The input of the system is the control voltage of the transistor. The output is the voltage delivered by the sensor.

The values of input and output data have been measured by a data acquisition system whose sampling period is $T_e = 1$ s. Figure 7 represents measured data.

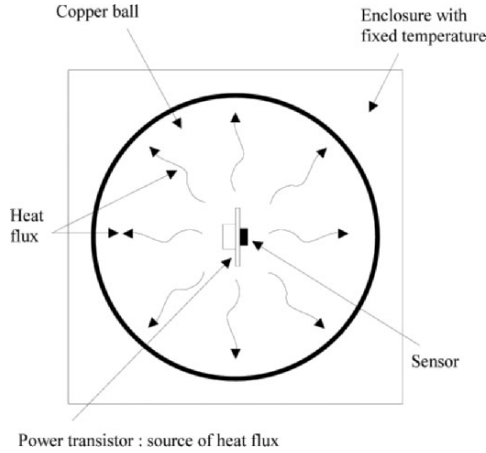


Fig. 6. Experiment.

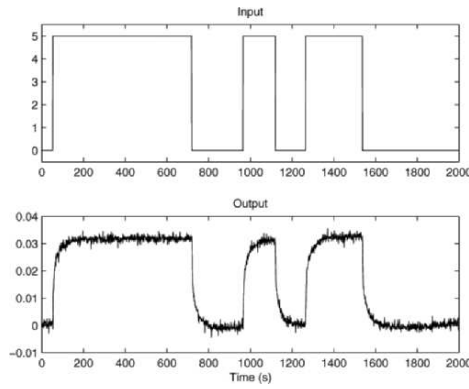


Fig. 7. Thermal pilot input and output.

Parameters a_0 , a_1 , b_0 , b_1 , and n_l are estimated using the OE technique (with $N = 30$ cells, $\omega_b = 10^{-5} \text{ rad/s}$, and $\omega_h = 10^5 \text{ rad/s}$ in order to simulate fractional integrators). Table 2 gives the obtained results using models (8) and (9).

The noise level is relatively important, which gives the illusion that the two models give satisfactory results. Nevertheless, we can verify that the model (9) gives a better time approximation on all the time area (Fig. 8). The used input was relatively poor: it is obvious that with a more exciting input, it will be possible to show the approximation ability of this model with two integrators in the case of quick transients.

Table 2. Estimation of the thermal pilot

Parameters	$H_n(s)$	$H_{n_1,n_2}(s)$
a_0	0.2058	0.0171
a_1	—	0.3509
b_0	0.0013	$1.08 \cdot 10^{-4}$
b_1	—	0.022
$n_1(n)$	0.7736	0.7719

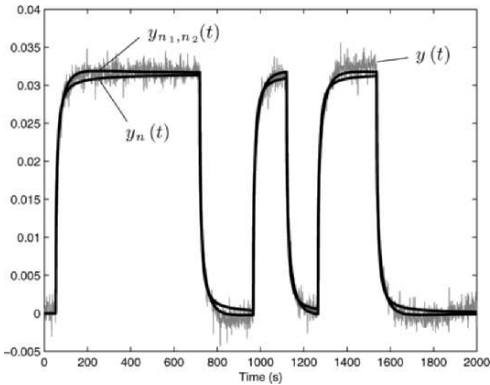


Fig. 8. Measured and estimated output.

6 Conclusion

In this paper, a contribution to the modelling and the identification of diffusive interfaces by fractional models has been presented. The objective was to improve the frequency approximation of the model. In preceding work [2, 6, 7], the modelling of diffusive systems using non-integer model, with the help of a fractional integrator operator, has shown its efficiency.

A theoretical approach showed that fractional modelling should be able to reproduce the essential characteristic of the diffusive phenomenon, i.e., that $n = 0.5$ then $\omega \rightarrow \infty$ while taking into account of the phenomenon geometry. The major problem is the frequency response of the diffusive interface and particularly its phase. The proposed solution in this paper consists to improve our preceding work thanks to a model with two fractional integrators, with one integrator constrained to the value 0.5.

A numerical simulation and an experiment on a thermal pilot have permitted to validate the hypothesis linked to the model with two integrators. Present research is focused on the numerical simulation of diffusive systems and first results confirm works presented in this paper.

Acknowledgment

This paper is a modified version of a paper published in proceedings of IDETC/CIE 2005, September 24–28, 2005, Long Beach, California, USA. The authors would like to thank the ASME for allowing them to republish this modification in this book.

References

1. Battaglia J-L (2002) Méthodes d'identification de modèles à dérivées d'ordres non entiers et de réduction modale. Application à la résolution de problèmes thermiques inverses dans des systèmes industriels. Habilitation à Diriger des Recherches. Université de Bordeaux I, France.
2. Benchellal A, Bachir S, Poinot T, Trigeassou J-C (2004) Identification of a non-integer model of induction machines. Proceedings of the FDA'04, 1st IFAC Workshop on Fractional Differentiation and its Applications, pp. 400–407. Bordeaux, France.
3. Canat S, Faucher J (2004) Modeling and simulation of induction machine with fractional derivative. Proceedings of the FDA'04, 1st IFAC Workshop on Fractional Differentiation and its Applications, pp. 393–399. Bordeaux, France.
4. Cois O (2002) Systèmes linéaires non entiers et identification par modèle non entier: application en thermique. Thèse de Doctorat. Université de Bordeaux I.
5. Khaorapapong T (2001) Modélisation d'ordre non entier des effets de fréquence dans les barres rotoriques d'une machine asynchrone. Thèse de Doctorat. INP de Toulouse.
6. Lin J, Poinot T, Trigeassou J-C, Ouvrard R (2000) Parameter estimation of fractional systems: application to the modeling of a lead-acid battery. SYSID 2000, 12th IFAC Symposium on System Identification. Santa Barbara, USA.
7. Lin J (2001) Modélisation et identification de systèmes d'ordre non entier. Thèse de Doctorat. Université de Poitiers, France.
8. Ljung L (1987) System identification – Theory for the user. Prentice-Hall, Englewood Cliffs, New Jersey, USA.
9. Oustaloup A (1995) La dérivation non entière: théorie, synthèse et applications. Paris.
10. Marquardt DW (1963) An algorithm for least-squares estimation of non-linear parameters, J. Soc. Indus. Appl. Math, 11(2):431–441.
11. Montseny M (1998) Diffusive representation of pseudo-differential time-operators. Proceedings of the Fractional Differential Systems: Models, Methods and Applications, Vol. 5, pp. 159–175. Paris, France.
12. Richalet J, Rault A, Pouliquen R (1971) Identification des processus par la méthode du modèle. Gordon and Breach.
13. Riu D, Retière N (2004) Implicit half-order systems utilisation for diffusion phenomenon modelling. Proceedings of the FDA'04, 1st IFAC Workshop on Fractional Differentiation and its Applications, pp. 387–392. Bordeaux, France.
14. Trigeassou J-C, Poinot T, Lin J, Oustaloup A, Levron F (1999) Modeling and identification of a non integer order system. Proceedings of the ECC'99, European Control Conference. Karlsruhe, Germany.

Part 4

Modeling

IDENTIFICATION OF FRACTIONAL MODELS FROM FREQUENCY DATA

Duarte Valério and José Sá da Costa

Technical University of Lisbon, Instituto Superior Técnico, Department of Mechanical Engineering – GCAR, Av. Rovisco Pais 1, 1049-001 Lisboa, Portugal; E-mail: {dvalerio,sadacosta}@dem.ist.utl.pt. Duarte Valério was partially supported by Fundação para a Ciência e a Tecnologia, grant SFRH/BPD/20636/2004, funded by POCI 2010, POS_C, FSE and MCTES

Abstract

Some existing methods for identifying models from frequency data (Levy's method without and with weights, and its improvements by Sanathanan and Koerner and by Lawrence and Rogers) are extended to deal with fractional models.

Keywords

Identification, fractional-order systems, Levy's method.

1 Introduction

Levy's method is a well-established method for finding the coefficients of a transfer function that models a plant having some known frequency behaviour [1]. In what follows the method (and some of its improvements) is expanded to deal with fractional-order transfer functions, that is to say, with the case when fractional (actually, non-integer, whether fractional or irrational) powers of Laplace operator s are expected to appear in the model. Such extensions should prove to be useful because several physical systems may be modelled using such transfer functions [2, 3], and because some methods for devising fractional-order controllers require identifying their transfer function from a frequency behaviour previously obtained.

2 Levy's Method Extended for Fractional Orders

Let us suppose we have a plant G with some known frequency behaviour, and that we want to model it using a commensurate fractional transfer function

$$\hat{G}(s) = \frac{b_0 + b_1 s^q + b_2 s^{2q} + \dots + b_m s^{mq}}{1 + a_1 s^q + a_2 s^{2q} + \dots + a_n s^{nq}} = \frac{\sum_{k=0}^m b_k s^{kq}}{1 + \sum_{k=1}^n a_k s^{kq}} \quad (1)$$

Remark 1: If q is 1 (or any other integer), transfer function (1) becomes a usual integer-order transfer function. Only if $q \in \mathbb{R} \setminus \mathbb{Z}$ is (1) said to be a fractional-order transfer function.

Remark 2: Transfer function (1) is said to be commensurate because all powers of s are multiple of a real q . It is of course possible to conceive transfer functions wherein the powers of s are not multiple of some q . Taking these into account would complicate the identification problem. Since commensurate transfer functions are those normally found in practice, we will restrict our attention to them.¹ Actually this is all that is needed for engineering purposes: numerical values are known with limited precision only, and commensurate order transfer functions provide good approximations of non-commensurate order ones.

Remark 3: Levy's original method requires setting in advance orders m and n . With this extension for fractional models the commensurate order q is also needed in advance. Some comments on that will be found below in the last section.

The frequency response of (1) is given by

$$\hat{G}(j\omega) = \frac{\sum_{k=0}^m b_k (j\omega)^{kq}}{1 + \sum_{k=1}^n a_k (j\omega)^{kq}} = \frac{N(\omega)}{D(\omega)} = \frac{\alpha(\omega) + j\beta(\omega)}{\sigma(\omega) + j\tau(\omega)} \quad (2)$$

where N and D are complex-valued and α, β, σ and τ , the real and imaginary parts thereof, are real-valued. The error between model and plant, for a given frequency ω , will be $\varepsilon(\omega) = G(j\omega) - N(\omega)/D(\omega)$. Now it might be possible to adjust the parameters of (1) by minimising the norm (or the square of the norm) of this error. Levy's method, however, tries to minimise the square of the norm of $\varepsilon(\omega)D(\omega) = G(j\omega)D(\omega) - N(\omega)$ instead, because this is a much simpler minimisation problem. So as to alleviate notation let us define $E(\omega) = \varepsilon(\omega)D(\omega)$ and drop the dependency on ω , we will have

$$\begin{aligned} E = GD - N &= [\operatorname{Re}(G) + j\operatorname{Im}(G)](\sigma + j\tau) - (\alpha + j\beta) = \\ &= [\operatorname{Re}(G)\sigma - \operatorname{Im}(G)\tau - \alpha] + j[\operatorname{Re}(G)\tau + \operatorname{Im}(G)\sigma - \beta] \end{aligned} \quad (3)$$

The square of the norm of E is

$$|E|^2 = [\operatorname{Re}(G)\sigma - \operatorname{Im}(G)\tau - \alpha]^2 + [\operatorname{Re}(G)\tau + \operatorname{Im}(G)\sigma - \beta]^2 \quad (4)$$

From (2) we see that

¹ Hartley and Lorenzo [4], addressing the identification of fractional models in a way rather close to Levy's method, also restrict themselves to commensurate transfer functions. But they only make use of simpler forms of $\hat{G}$, namely forcing the numerator to be 1.

$$\alpha(\omega) = \sum_{k=0}^m b_k \operatorname{Re}[(j\omega)^{kq}] \quad (5)$$

$$\beta(\omega) = \sum_{k=0}^m b_k \operatorname{Im}[(j\omega)^{kq}] \quad (6)$$

$$\sigma(\omega) = 1 + \sum_{k=1}^n a_k \operatorname{Re}[(j\omega)^{kq}] \quad (7)$$

$$\tau(\omega) = \sum_{k=1}^n a_k \operatorname{Im}[(j\omega)^{kq}] \quad (8)$$

Thus, if we differentiate $|E|^2$ with respect to one of the coefficients b_k , we shall have

$$\begin{aligned} \frac{\partial |E|^2}{\partial b_k} = & -2[\operatorname{Re}(G)\sigma - \operatorname{Im}(G)\tau - \alpha] \operatorname{Re}[(j\omega)^{kq}] \\ & - 2[\operatorname{Re}(G)\tau + \operatorname{Im}(G)\sigma - \beta] \operatorname{Im}[(j\omega)^{kq}] \end{aligned} \quad (9)$$

Equalling the derivative to zero,

$$\begin{aligned} \frac{\partial |E|^2}{\partial b_k} = 0 \Leftrightarrow & [\operatorname{Re}(G)\sigma - \operatorname{Im}(G)\tau - \alpha] \operatorname{Re}[(j\omega)^{kq}] \\ & + [\operatorname{Re}(G)\tau + \operatorname{Im}(G)\sigma - \beta] \operatorname{Im}[(j\omega)^{kq}] = 0 \end{aligned} \quad (10)$$

And if we differentiate $|E|^2$ with respect to one of the coefficients a_k we shall have

$$\begin{aligned} \frac{\partial |E|^2}{\partial a_k} = & 2[\operatorname{Re}(G)\sigma - \operatorname{Im}(G)\tau - \alpha] \operatorname{Re}(G) \operatorname{Re}[(j\omega)^{kq}] \\ & + 2[\operatorname{Re}(G)\tau + \operatorname{Im}(G)\sigma - \beta] \operatorname{Im}(G) \operatorname{Re}[(j\omega)^{kq}] \\ & - 2[\operatorname{Re}(G)\sigma - \operatorname{Im}(G)\tau - \alpha] \operatorname{Im}(G) \operatorname{Im}[(j\omega)^{kq}] \\ & + 2[\operatorname{Re}(G)\tau + \operatorname{Im}(G)\sigma - \beta] \operatorname{Re}(G) \operatorname{Im}[(j\omega)^{kq}] \end{aligned} \quad (11)$$

Equalling the derivative to zero,

$$\begin{aligned} \frac{\partial |E|^2}{\partial a_k} = 0 \Leftrightarrow & \sigma[\operatorname{Re}(G)]^2 \operatorname{Re}[(j\omega)^{kQ}] - \tau \operatorname{Im}(G) \operatorname{Re}(G) \operatorname{Re}[(j\omega)^{kQ}] \\ & - \alpha \operatorname{Re}(G) \operatorname{Re}[(j\omega)^{kQ}] + \tau \operatorname{Im}(G) \operatorname{Re}(G) \operatorname{Re}[(j\omega)^{kQ}] \\ & + \sigma[\operatorname{Im}(G)]^2 \operatorname{Re}[(j\omega)^{kQ}] - \beta \operatorname{Im}(G) \operatorname{Re}[(j\omega)^{kQ}] \end{aligned} \quad (12)$$

$$\begin{aligned}
& -\sigma \operatorname{Im}(G) \operatorname{Re}(G) \operatorname{Im}\left[(j\omega)^{kQ}\right] + \tau \left[\operatorname{Im}(G)\right]^2 \operatorname{Im}\left[(j\omega)^{kQ}\right] \\
& + \alpha \operatorname{Im}(G) \operatorname{Im}\left[(j\omega)^{kQ}\right] + \tau \left[\operatorname{Re}(G)\right]^2 \operatorname{Im}\left[(j\omega)^{kQ}\right] \\
& + \sigma \operatorname{Im}(G) \operatorname{Re}(G) \operatorname{Im}\left[(j\omega)^{kQ}\right] - \beta \operatorname{Re}(G) \operatorname{Im}\left[(j\omega)^{kQ}\right] = 0 \\
\Leftrightarrow & \sigma \left\{ \left[\operatorname{Im}(G)\right]^2 + \left[\operatorname{Re}(G)\right]^2 \right\} \operatorname{Re}\left[(j\omega)^{kQ}\right] \\
& + \tau \left\{ \left[\operatorname{Im}(G)\right]^2 + \left[\operatorname{Re}(G)\right]^2 \right\} \operatorname{Im}\left[(j\omega)^{kQ}\right] \\
& + \alpha \left\{ \operatorname{Im}(G) \operatorname{Im}\left[(j\omega)^{kQ}\right] - \operatorname{Re}(G) \operatorname{Re}\left[(j\omega)^{kQ}\right] \right\} \\
& + \beta \left\{ -\operatorname{Im}(G) \operatorname{Re}\left[(j\omega)^{kQ}\right] - \operatorname{Re}(G) \operatorname{Im}\left[(j\omega)^{kQ}\right] \right\} = 0
\end{aligned}$$

The $m + 1$ equations given by (10) and the n equations given by (12) make up a linear system that may be solved so as to find the coefficients of (1). Usually the frequency behaviour of the plant is known in more than one frequency (otherwise it is likely that the model will be rather poor). Let us suppose that it is known at f frequencies. Then the system to solve, given by equations (10) and (12) written explicitly on coefficients a and b , is

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} e \\ g \end{bmatrix} \quad (13)$$

where

$$\mathbf{A}_{l,c} = \sum_{p=1}^f \left\{ -\operatorname{Re}\left[(j\omega_p)^{lq}\right] \operatorname{Re}\left[(j\omega_p)^{cq}\right] - \operatorname{Im}\left[(j\omega_p)^{lq}\right] \operatorname{Im}\left[(j\omega_p)^{cq}\right] \right\}, \quad (14)$$

$$l = 0 \dots m \wedge c = 0 \dots m$$

$$\begin{aligned}
\mathbf{B}_{l,c} = \sum_{p=1}^f & \left\{ \operatorname{Re}\left[(j\omega_p)^{lq}\right] \operatorname{Re}\left[(j\omega_p)^{cq}\right] \operatorname{Re}\left[G(j\omega_p)\right] \right. \\
& + \operatorname{Im}\left[(j\omega_p)^{lq}\right] \operatorname{Re}\left[(j\omega_p)^{cq}\right] \operatorname{Im}\left[G(j\omega_p)\right] \\
& - \operatorname{Re}\left[(j\omega_p)^{lq}\right] \operatorname{Im}\left[(j\omega_p)^{cq}\right] \operatorname{Im}\left[G(j\omega_p)\right] \\
& \left. + \operatorname{Im}\left[(j\omega_p)^{lq}\right] \operatorname{Im}\left[(j\omega_p)^{cq}\right] \operatorname{Re}\left[G(j\omega_p)\right] \right\}, \quad l = 0 \dots m \wedge c = 1 \dots n
\end{aligned} \quad (15)$$

$$\begin{aligned}
\mathbf{C}_{l,c} = \sum_{p=1}^f & \left\{ -\operatorname{Re}\left[(j\omega_p)^{lq}\right] \operatorname{Re}\left[(j\omega_p)^{cq}\right] \operatorname{Re}\left[G(j\omega_p)\right] \right. \\
& + \operatorname{Im}\left[(j\omega_p)^{lq}\right] \operatorname{Re}\left[(j\omega_p)^{cq}\right] \operatorname{Im}\left[G(j\omega_p)\right] \\
& - \operatorname{Re}\left[(j\omega_p)^{lq}\right] \operatorname{Im}\left[(j\omega_p)^{cq}\right] \operatorname{Im}\left[G(j\omega_p)\right] \\
& \left. + \operatorname{Im}\left[(j\omega_p)^{lq}\right] \operatorname{Im}\left[(j\omega_p)^{cq}\right] \operatorname{Re}\left[G(j\omega_p)\right] \right\}
\end{aligned} \quad (16)$$

$$\begin{aligned}
& -\operatorname{Im}\left[(j\omega_p)^{lq}\right]\operatorname{Im}\left[(j\omega_p)^{cq}\right]\operatorname{Re}\left[G(j\omega_p)\right]\}, \quad l=1\dots n \wedge c=0\dots m \\
\mathbf{D}_{l,c} = & \sum_{p=1}^f \left(\left\{ \operatorname{Re}\left[G(j\omega_p)\right] \right\}^2 + \left\{ \operatorname{Im}\left[G(j\omega_p)\right] \right\}^2 \right) \\
& \left\{ \operatorname{Re}\left[(j\omega_p)^{lq}\right]\operatorname{Re}\left[(j\omega_p)^{cq}\right] + \operatorname{Im}\left[(j\omega_p)^{lq}\right]\operatorname{Im}\left[(j\omega_p)^{cq}\right] \right\}, \quad (17) \\
& l=1\dots n \wedge c=1\dots n
\end{aligned}$$

$$\mathbf{b} = [b_0 \quad \dots \quad b_m]^T \quad (18)$$

$$\mathbf{a} = [a_1 \quad \dots \quad a_n]^T \quad (19)$$

$$\begin{aligned}
\mathbf{e}_{l,1} = & \sum_{p=1}^f \left\{ -\operatorname{Re}\left[(j\omega_p)^{lq}\right]\operatorname{Re}\left[G(j\omega_p)\right] - \operatorname{Im}\left[(j\omega_p)^{lq}\right]\operatorname{Im}\left[G(j\omega_p)\right] \right\}, \\
& l=0\dots m
\end{aligned} \quad (20)$$

$$\mathbf{g}_{l,1} = \sum_{p=1}^f -\operatorname{Re}\left[(j\omega_p)^{lq}\right] \left(\left\{ \operatorname{Re}\left[G(j\omega_p)\right] \right\}^2 + \left\{ \operatorname{Im}\left[G(j\omega_p)\right] \right\}^2 \right), \quad l=1\dots n \quad (21)$$

If q is 1, the real and imaginary parts of $(j\omega)^k$ reduce (k being a natural) to either $\pm\omega^k$ or $\pm j\omega^k$, and matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ and vectors $\mathbf{e}$ and $\mathbf{g}$ assume the usual structures of Levy's identification method.

3 First Improvement: Vinagre's Weights

Levy's method's drawbacks are well known, one of them being that low-frequency data has little influence in (13) and the resulting fit is poor for such frequencies. Using well-chosen weights for increasing the influence of low-frequency data is a means of dealing with this. Vinagre [5] notes that, if $g(t)$ is the step response of our system, then

$$\int_0^{+\infty} |g(t) - \hat{g}(t)|^2 dt = \int_0^{+\infty} \left| \mathcal{L}^{-1}\left[G(s)\frac{1}{s}\right] - \mathcal{L}^{-1}\left[\hat{G}(s)\frac{1}{s}\right] \right|^2 dt \quad (22)$$

and that Parseval's theorem turns this into

$$\int_{-\infty}^{+\infty} \left| G(j\omega)\frac{1}{j\omega} - \hat{G}(j\omega)\frac{1}{j\omega} \right|^2 d\omega = \int_{-\infty}^{+\infty} \frac{|\varepsilon(\omega)|^2}{\omega^2} d\omega \quad (23)$$

Using the trapezoidal numerical integration rule this can be approximated by

$$\sum_{p=1}^{f-1} \left\{ \frac{1}{2} (\omega_{p+1} - \omega_p) \left[\frac{|\mathcal{E}(\omega_{p+1})|^2}{\omega_{p+1}^2} + \frac{|\mathcal{E}(\omega_p)|^2}{\omega_p^2} \right] \right\} = \sum_{p=1}^f \frac{|\mathcal{E}(\omega_p)|^2}{\omega_p^2} \varphi_p \quad (24)$$

where

$$\varphi_p = \begin{cases} \frac{\omega_2 - \omega_1}{2}, & \text{if } p = 1 \\ \frac{\omega_{p+1} - \omega_{p-1}}{2}, & \text{if } 1 < p < f \\ \frac{\omega_f - \omega_{f-1}}{2}, & \text{if } p = f \end{cases} \quad (25)$$

are the coefficients of the trapezoidal integration rule. Just as Levy's method minimises $\sum_{p=1}^f |E(\omega_p)|^2$ instead of $\sum_{p=1}^f |\mathcal{E}(\omega_p)|^2$, so this time, instead of the right-hand member of (23), the quantity $\sum_{p=1}^f |E(\omega_p)|^2 \frac{\varphi_p}{\omega_p^2}$ will be minimised instead. The fraction multiplying the square of the norm is the weight,² $w_p = \varphi_p / \omega_p^2$, that clearly increases the influence of low frequencies. Since the weight does not depend on coefficients a and b , it will not change the values of derivatives (9) and (11). The only difference in the method is that matrixes and vectors in (13) will now be given by

$$\mathbf{A}_{l,c} = \sum_{p=1}^f \left\{ -\text{Re}[(j\omega_p)^{lq}] \text{Re}[(j\omega_p)^{cq}] - \text{Im}[(j\omega_p)^{lq}] \text{Im}[(j\omega_p)^{cq}] \right\} w_p, \quad (26)$$

$$l = 0 \dots m \wedge c = 0 \dots m$$

$$\begin{aligned} \mathbf{B}_{l,c} = \sum_{p=1}^f \bigg\{ & \text{Re}[(j\omega_p)^{lq}] \text{Re}[(j\omega_p)^{cq}] \text{Re}[G(j\omega_p)] \\ & + \text{Im}[(j\omega_p)^{lq}] \text{Re}[(j\omega_p)^{cq}] \text{Im}[G(j\omega_p)] \\ & - \text{Re}[(j\omega_p)^{lq}] \text{Im}[(j\omega_p)^{cq}] \text{Im}[G(j\omega_p)] \\ & + \text{Im}[(j\omega_p)^{lq}] \text{Im}[(j\omega_p)^{cq}] \text{Re}[G(j\omega_p)] \bigg\} w_p, \end{aligned} \quad (27)$$

$$l = 0 \dots m \wedge c = 1 \dots n$$

² Based upon energetic considerations [5], adds yet another term to this weight, that depends neither on p nor on coefficients a or b , and thus may be neglected by the minimisation.

$$\begin{aligned}
\mathbf{C}_{l,c} = \sum_{p=1}^f & \left\{ -\operatorname{Re} \left[(j\omega_p)^{lq} \right] \operatorname{Re} \left[(j\omega_p)^{cq} \right] \operatorname{Re} \left[G(j\omega_p) \right] \right. \\
& + \operatorname{Im} \left[(j\omega_p)^{lq} \right] \operatorname{Re} \left[(j\omega_p)^{cq} \right] \operatorname{Im} \left[G(j\omega_p) \right] \\
& - \operatorname{Re} \left[(j\omega_p)^{lq} \right] \operatorname{Im} \left[(j\omega_p)^{cq} \right] \operatorname{Im} \left[G(j\omega_p) \right] \\
& \left. - \operatorname{Im} \left[(j\omega_p)^{lq} \right] \operatorname{Im} \left[(j\omega_p)^{cq} \right] \operatorname{Re} \left[G(j\omega_p) \right] \right\} \mathbf{w}_p, \\
l = 1 \dots n \wedge c = 0 \dots m
\end{aligned} \tag{28}$$

$$\begin{aligned}
\mathbf{D}_{l,c} = \sum_{p=1}^f & \left(\left\{ \operatorname{Re} \left[G(j\omega_p) \right] \right\}^2 + \left\{ \operatorname{Im} \left[G(j\omega_p) \right] \right\}^2 \right) \\
& \left\{ \operatorname{Re} \left[(j\omega_p)^{lq} \right] \operatorname{Re} \left[(j\omega_p)^{cq} \right] + \operatorname{Im} \left[(j\omega_p)^{lq} \right] \operatorname{Im} \left[(j\omega_p)^{cq} \right] \right\} \mathbf{w}_p, \\
l = 1 \dots n \wedge c = 1 \dots n
\end{aligned} \tag{29}$$

$$\begin{aligned}
e_{l,1} = \sum_{p=1}^f & \left\{ -\operatorname{Re} \left[(j\omega_p)^{lq} \right] \operatorname{Re} \left[G(j\omega_p) \right] - \operatorname{Im} \left[(j\omega_p)^{lq} \right] \operatorname{Im} \left[G(j\omega_p) \right] \right\} \mathbf{w}_p, \\
l = 0 \dots m
\end{aligned} \tag{30}$$

$$g_{l,1} = \sum_{p=1}^f -\operatorname{Re} \left[(j\omega_p)^{lq} \right] \left(\left\{ \operatorname{Re} \left[G(j\omega_p) \right] \right\}^2 + \left\{ \operatorname{Im} \left[G(j\omega_p) \right] \right\}^2 \right) \mathbf{w}_p, \quad l = 1 \dots n \tag{31}$$

4 Second Improvement: The Iterative Method of Sanathanan and Koerner

Another way of improving Levy's method was proposed by Sanathanan and Koerner [6]. It consists in performing several iterations where variable E is replaced by

$$E_L = \frac{GD - N}{D_{L-1}} \tag{32}$$

where L is the iteration number and D_{L-1} is the denominator found in the previous iteration. In the first iteration this is assumed to be 1 and the result is that of Levy's method. If convergence exists, subsequent iterations will see E_L converge to ε . This time the variable minimised is

$$\sum_{p=1}^f |E(\omega_p)|^2 / |D_{L-1}(\omega_p)|^2 \tag{33}$$

and the fraction, $w_p = 1/|D_{L-1}(\omega_p)|^2$, is the weight. It depends on coefficients known from the last iteration, not the current one, and so derivatives (9) and (11) are again not affected. Thus (26) to (31) remain valid (save that w_p is given by a different expression), and these are the values with which (13) is to be solved in each iteration. The resulting values of a will be used to find the new weights for the next iteration. The process may be stopped when no significant change in parameters is achieved or after some pre-set number of iterations (which is sometimes advisable because too many iterations may cause numerical errors to accumulate causing the result to diverge).

5 Third Improvement: The Iterative Method of Lawrence and Rogers

All possibilities addressed this far involve solving a linear set of equations, and with all of them, if new data from new frequencies appear, the system will have to be solved again. Lawrence and Rogers [7] developed an iterative method to avoid solving the system again if new data is obtained; this method deals with each frequency at one time. (This is not only for saving time. As will be seen in the subsections that follow, equation systems that show up with the methods of previous subsections may cause numerical problems to arise. Avoiding such systems may thus be numerically favourable.) It stems from writing (3) in the following form:

$$E = GD - N = G(1 + a^T s) - b^T t = G + a^T Gs - b^T t \quad (34)$$

where

$$s = \begin{bmatrix} (j\omega)^q & \cdots & (j\omega)^{nq} \end{bmatrix}^T, \quad t = \begin{bmatrix} 1 & (j\omega)^q & \cdots & (j\omega)^{mq} \end{bmatrix}^T \quad (35)$$

If we let

$$v = \begin{bmatrix} b & a \end{bmatrix}^T, \quad u = \begin{bmatrix} t & -Gs \end{bmatrix}^T \quad (36)$$

then (34) becomes

$$E = G - v^T u \quad (37)$$

Now instead of (4) we may alternatively write

$$|E|^2 = (G - v^T u) \overline{(G - v^T u)}^T = G\bar{G} - Gu^T v - \bar{G}v^T u + v^T u \bar{u}^T v \quad (38)$$

where it has been taken into account that G is a scalar (whereas u and v are vectors) and that v is real-valued (whereas G and u are complex-valued). Differentiating (38) in order to v gives

$$\frac{\partial |E|^2}{\partial \mathbf{v}} = -G\bar{\mathbf{u}} - \bar{G}\mathbf{u} + \mathbf{u}\bar{\mathbf{u}}^T \mathbf{v} + \bar{\mathbf{u}}\mathbf{u}^T \mathbf{v} \quad (39)$$

and equalling (39) to zero gives

$$\left(\mathbf{u}\bar{\mathbf{u}}^T + \bar{\mathbf{u}}\mathbf{u}^T \right) \mathbf{v} = G\bar{\mathbf{u}} + \bar{G}\mathbf{u} \quad (40)$$

It should be noticed that both the matrix in the left-hand side multiplying $\mathbf{v}$ and the vector in the right-hand side are real-valued. And since we usually deal not with only one but with f frequencies, this becomes

$$\sum_{k=1}^f \left(\mathbf{u}_k \bar{\mathbf{u}}_k^T + \bar{\mathbf{u}}_k \mathbf{u}_k^T \right) \mathbf{v} = \sum_{k=1}^f \left(G_k \bar{\mathbf{u}}_k + \bar{G}_k \mathbf{u}_k \right) \quad (41)$$

Finally, if weights are included, we shall want to minimise

$$\frac{\partial \left(|E|^2 w^2 \right)}{\partial \mathbf{v}} = w^2 \left(-G\bar{\mathbf{u}} - \bar{G}\mathbf{u} + \mathbf{u}\bar{\mathbf{u}}^T \mathbf{v} + \bar{\mathbf{u}}\mathbf{u}^T \mathbf{v} \right) \quad (42)$$

and (41) becomes

$$\sum_{k=1}^f w_k^2 \left(\mathbf{u}_k \bar{\mathbf{u}}_k^T + \bar{\mathbf{u}}_k \mathbf{u}_k^T \right) \mathbf{v} = \sum_{k=1}^f w_k^2 \left(G_k \bar{\mathbf{u}}_k + \bar{G}_k \mathbf{u}_k \right) \quad (43)$$

Until now this is solely putting (13) under an equivalent, more compact form (the resulting system of equations is, of course, equivalent; the dimension of the matrix and the size of the vector in (43) are the same as those in (13)). Yet (43) allows for the developments that follow. Let

$$\mathbf{H}_f^{-1} = \sum_{k=1}^f w_k^2 \left(\mathbf{u}_k \bar{\mathbf{u}}_k^T + \bar{\mathbf{u}}_k \mathbf{u}_k^T \right) = \mathbf{H}_{f-1}^{-1} + w_f^2 \left(\mathbf{u}_f \bar{\mathbf{u}}_f^T + \bar{\mathbf{u}}_f \mathbf{u}_f^T \right) \quad (44)$$

Then (43) becomes

$$\mathbf{H}_f^{-1} \mathbf{v}_f = \sum_{k=1}^f w_k^2 \left(G_k \bar{\mathbf{u}}_k + \bar{G}_k \mathbf{u}_k \right) \quad (45)$$

where the subscript on $\mathbf{v}$ has been added to show that the solution is obtained from data concerning f frequencies. Additionally,

$$\begin{aligned} \sum_{k=1}^f w_k^2 \left(G_k \bar{\mathbf{u}}_k + \bar{G}_k \mathbf{u}_k \right) &= w_f^2 \left(G_f \bar{\mathbf{u}}_f + \bar{G}_f \mathbf{u}_f \right) + \sum_{k=1}^{f-1} w_k^2 \left(G_k \bar{\mathbf{u}}_k + \bar{G}_k \mathbf{u}_k \right) \\ &= w_f^2 \left(G_f \bar{\mathbf{u}}_f + \bar{G}_f \mathbf{u}_f \right) + \mathbf{H}_{f-1}^{-1} \mathbf{v}_{f-1} \end{aligned} \quad (46)$$

Hence

$$\begin{aligned}
 \mathbf{H}_f^{-1} \mathbf{v}_f &= \mathbf{w}_f^2 \left(G_f \overline{u_f} + \overline{G_f} u_f \right) + \mathbf{H}_{f-1}^{-1} \mathbf{v}_{f-1} \\
 &= \mathbf{w}_f^2 \left(G_f \overline{u_f} + \overline{G_f} u_f \right) + \left[\mathbf{H}_f^{-1} - \mathbf{w}_f^2 \left(u_f \overline{u_f^T} + \overline{u_f} u_f^T \right) \right] \mathbf{v}_{f-1} \\
 &= \mathbf{H}_f^{-1} \mathbf{v}_{f-1} - \mathbf{w}_f^2 \left(u_f \overline{u_f^T} + \overline{u_f} u_f^T \right) \mathbf{v}_{f-1} + \mathbf{w}_f^2 \left(G_f \overline{u_f} + \overline{G_f} u_f \right) \\
 \Rightarrow \mathbf{v}_f &= \mathbf{v}_{f-1} - \mathbf{H}_f \mathbf{w}_f^2 \left(u_f \overline{u_f^T} + \overline{u_f} u_f^T \right) \mathbf{v}_{f-1} + \mathbf{H}_f \mathbf{w}_f^2 \left(G_f \overline{u_f} + \overline{G_f} u_f \right) \\
 &= \mathbf{v}_{f-1} + \mathbf{H}_f \mathbf{w}_f^2 \left[u_f \left(\overline{G_f} - \overline{u_f^T} \mathbf{v}_{f-1} \right) + \overline{u_f} \left(G_f - u_f^T \mathbf{v}_{f-1} \right) \right]
 \end{aligned} \tag{47}$$

This last equality means that once a vector $\mathbf{v}$ with parameters for the model is obtained from data concerning $f-1$ frequencies, it is possible to improve it taking into account data from another frequency. It is even possible to find an expression for $\mathbf{H}$ that does not require inverting $\mathbf{H}^{-1}$, developing (44) as follows:

$$\begin{aligned}
 \mathbf{H}_f^{-1} &= \mathbf{H}_{f-1}^{-1} + \mathbf{w}_f^2 \left(\text{Re}[u_f] + j \text{Im}[u_f] \right) \left(\text{Re}[u_f^T] - j \text{Im}[u_f^T] \right) \\
 &\quad + \mathbf{w}_f^2 \left(\text{Re}[u_f] - j \text{Im}[u_f] \right) \left(\text{Re}[u_f^T] + j \text{Im}[u_f^T] \right) \\
 &= \mathbf{H}_{f-1}^{-1} + \mathbf{w}_f^2 \left(\text{Re}[u_f] \text{Re}[u_f^T] + \text{Im}[u_f] \text{Im}[u_f^T] \right. \\
 &\quad \left. - j \text{Re}[u_f] \text{Im}[u_f^T] + j \text{Im}[u_f] \text{Re}[u_f^T] \right) \\
 &\quad + \text{Re}[u_f] \text{Re}[u_f^T] + \text{Im}[u_f] \text{Im}[u_f^T] \\
 &\quad \left. + j \text{Re}[u_f] \text{Im}[u_f^T] - j \text{Im}[u_f] \text{Re}[u_f^T] \right) \\
 &= \mathbf{H}_{f-1}^{-1} + 2\mathbf{w}_f^2 \left(\text{Re}[u_f] \text{Re}[u_f^T] + \text{Im}[u_f] \text{Im}[u_f^T] \right)
 \end{aligned} \tag{48}$$

Let

$$\mathbf{Z}_f^{-1} = \mathbf{H}_{f-1}^{-1} + 2\mathbf{w}_f^2 \text{Re}[u_f] \text{Re}[u_f^T] \tag{49}$$

Multiplying this by $\mathbf{Z}_f$ by $\mathbf{H}_{f-1}$ and by $\text{Re}[u_f]$

$$\mathbf{I} = \mathbf{Z}_f \mathbf{H}_{f-1}^{-1} + 2\mathbf{w}_f^2 \mathbf{Z}_f \text{Re}[u_f] \text{Re}[u_f^T] \tag{50}$$

$$\Leftrightarrow \mathbf{H}_{f-1} = \mathbf{Z}_f + 2\mathbf{w}_f^2 \mathbf{Z}_f \text{Re}[u_f] \text{Re}[u_f^T] \mathbf{H}_{f-1} \tag{51}$$

$$\begin{aligned}
 \Leftrightarrow \mathbf{H}_{f-1} \text{Re}[u_f] &= \mathbf{Z}_f \text{Re}[u_f] + 2\mathbf{w}_f^2 \mathbf{Z}_f \text{Re}[u_f] \text{Re}[u_f^T] \mathbf{H}_{f-1} \text{Re}[u_f] \\
 &= \mathbf{Z}_f \text{Re}[u_f] \left(1 + 2\mathbf{w}_f^2 \text{Re}[u_f^T] \mathbf{H}_{f-1} \text{Re}[u_f] \right)
 \end{aligned} \tag{52}$$

It should be noticed that the term within parenthesis is scalar. Rearranging and then multiplying by $\text{Re}[u_f^T] \mathbf{H}_{f-1}$,

$$\begin{aligned}
\mathbf{Z}_f \operatorname{Re}[u_f] &= \mathbf{H}_{f-1} \operatorname{Re}[u_f] \left(1 + 2w_f^2 \operatorname{Re}[u_f^T] \mathbf{H}_{f-1} \operatorname{Re}[u_f]\right)^{-1} \\
\Leftrightarrow \mathbf{Z}_f \operatorname{Re}[u_f] \operatorname{Re}[u_f^T] \mathbf{H}_{f-1} &= \frac{\mathbf{H}_{f-1} \operatorname{Re}[u_f] \operatorname{Re}[u_f^T] \mathbf{H}_{f-1}}{1 + 2w_f^2 \operatorname{Re}[u_f^T] \mathbf{H}_{f-1} \operatorname{Re}[u_f]}
\end{aligned} \tag{53}$$

Recall that the denominator is a scalar. Now (51) shows that

$$\begin{aligned}
\mathbf{H}_{f-1} &= \mathbf{Z}_f + 2w_f^2 \mathbf{Z}_f \operatorname{Re}[u_f] \operatorname{Re}[u_f^T] \mathbf{H}_{f-1} \\
\mathbf{Z}_f \operatorname{Re}[u_f] \operatorname{Re}[u_f^T] \mathbf{H}_{f-1} &= \frac{\mathbf{H}_{f-1} - \mathbf{Z}_f}{2w_f^2}
\end{aligned} \tag{54}$$

From (53) and (54)

$$\begin{aligned}
\frac{\mathbf{H}_{f-1} \operatorname{Re}[u_f] \operatorname{Re}[u_f^T] \mathbf{H}_{f-1}}{1 + 2w_f^2 \operatorname{Re}[u_f^T] \mathbf{H}_{f-1} \operatorname{Re}[u_f]} &= \frac{\mathbf{H}_{f-1} - \mathbf{Z}_f}{2w_f^2} \\
\Leftrightarrow \mathbf{Z}_f &= \mathbf{H}_{f-1} - \frac{\mathbf{H}_{f-1} \operatorname{Re}[u_f] \operatorname{Re}[u_f^T] \mathbf{H}_{f-1}}{\frac{1}{2w_f^2} + \operatorname{Re}[u_f^T] \mathbf{H}_{f-1} \operatorname{Re}[u_f]} \\
&= \mathbf{H}_{f-1} \left(\mathbf{I} - \frac{\operatorname{Re}[u_f] \operatorname{Re}[u_f^T] \mathbf{H}_{f-1}}{\frac{1}{2w_f^2} + \operatorname{Re}[u_f^T] \mathbf{H}_{f-1} \operatorname{Re}[u_f]} \right)
\end{aligned} \tag{55}$$

The identity matrix above has the same size of $\mathbf{H}_{f-1}$, which is also the size of matrix $\operatorname{Re}[u_f] \operatorname{Re}[u_f^T]$. Now the steps that follow are close parallels of those from (49) to (55). From (48) and (49) we know that

$$\mathbf{H}_f^{-1} = \mathbf{Z}_f^{-1} + 2w_f^2 \operatorname{Im}[u_f] \operatorname{Im}[u_f^T] \tag{56}$$

Multiplying this by $\mathbf{H}_f$, by $\mathbf{Z}_f$ and by $\operatorname{Im}[u_f]$

$$\mathbf{I} = \mathbf{H}_f \mathbf{Z}_f^{-1} + 2w_f^2 \mathbf{H}_f \operatorname{Im}[u_f] \operatorname{Im}[u_f^T] \tag{57}$$

$$\Leftrightarrow \mathbf{Z}_f = \mathbf{H}_f + 2w_f^2 \mathbf{H}_f \operatorname{Im}[u_f] \operatorname{Im}[u_f^T] \mathbf{Z}_f \tag{58}$$

$$\begin{aligned}
&\Leftrightarrow \mathbf{Z}_f \operatorname{Im}[u_f] \\
&= \mathbf{H}_f \operatorname{Im}[u_f] + 2w_f^2 \mathbf{H}_f \operatorname{Im}[u_f] \operatorname{Im}[u_f^T] \mathbf{Z}_f \operatorname{Im}[u_f] \\
&= \mathbf{H}_f \operatorname{Im}[u_f] \left(1 + 2w_f^2 \operatorname{Im}[u_f^T] \mathbf{Z}_f \operatorname{Im}[u_f]\right)
\end{aligned} \tag{59}$$

Rearranging and then multiplying by $\operatorname{Im}[u_f^T] \mathbf{Z}_f$,

$$\begin{aligned}
\mathbf{H}_f \operatorname{Im}[u_f] &= \mathbf{Z}_f \operatorname{Im}[u_f] \left(1 + 2w_f^2 \operatorname{Im}[u_f^T] \mathbf{Z}_f \operatorname{Im}[u_f] \right)^{-1} \\
\Leftrightarrow \mathbf{H}_f \operatorname{Im}[u_f] \operatorname{Im}[u_f^T] \mathbf{Z}_f &= \frac{\mathbf{Z}_f \operatorname{Im}[u_f] \operatorname{Im}[u_f^T] \mathbf{Z}_f}{1 + 2w_f^2 \operatorname{Im}[u_f^T] \mathbf{Z}_f \operatorname{Im}[u_f]}
\end{aligned} \tag{60}$$

Now (58) shows that

$$\begin{aligned}
\mathbf{Z}_f &= \mathbf{H}_f + 2w_f^2 \mathbf{H}_f \operatorname{Im}[u_f] \operatorname{Im}[u_f^T] \mathbf{Z}_f \\
\Leftrightarrow \mathbf{H}_f \operatorname{Im}[u_f] \operatorname{Im}[u_f^T] \mathbf{Z}_f &= \frac{\mathbf{Z}_f - \mathbf{H}_f}{2w_f^2}
\end{aligned} \tag{61}$$

From (60) and (61)

$$\begin{aligned}
\frac{\mathbf{Z}_f \operatorname{Im}[u_f] \operatorname{Im}[u_f^T] \mathbf{Z}_f}{1 + 2w_f^2 \operatorname{Im}[u_f^T] \mathbf{Z}_f \operatorname{Im}[u_f]} &= \frac{\mathbf{Z}_f - \mathbf{H}_f}{2w_f^2} \\
\Leftrightarrow \mathbf{H}_f &= \mathbf{Z}_f - \frac{\mathbf{Z}_f \operatorname{Im}[u_f] \operatorname{Im}[u_f^T] \mathbf{Z}_f}{\frac{1}{2w_f^2} + \operatorname{Im}[u_f^T] \mathbf{Z}_f \operatorname{Im}[u_f]} \\
&= \mathbf{Z}_f \left(\mathbf{I} - \frac{\operatorname{Im}[u_f] \operatorname{Im}[u_f^T] \mathbf{Z}_f}{\frac{1}{2w_f^2} + \operatorname{Im}[u_f^T] \mathbf{Z}_f \operatorname{Im}[u_f]} \right)
\end{aligned} \tag{62}$$

The identity matrix above has the same size of $\mathbf{Z}_f$, which is also the size of matrix $\operatorname{Im}[u_f] \operatorname{Im}[u_f^T]$.

The best way to use this method is to begin with some values for $\mathbf{H}$ and v (which is made up of parameters a and b), obtained applying (44) and (45) with a few frequencies. Data from each of the further frequencies is then taken into account using (55) and (62), with which it is possible to obtain a value for $\mathbf{H}_f$, from the value of $\mathbf{H}_{f-1}$, inverting only a scalar. Then (47) is used to update vector of parameters v .

Actually it is possible to begin with no estimate at all, making

$$v_0 = 0 \wedge \mathbf{H}_0^{-1} = 0 \Rightarrow \mathbf{H}_0 = \mathbf{I} \times \infty \tag{63}$$

Since infinity is not an available numerical value, some positive real number x is used instead and

$$v_0 = 0 \wedge \mathbf{H}_0 = \mathbf{I} \times x \tag{64}$$

However, it is rather hard to tell in advance which number to use; large real numbers, close to the floating-point limit, are good approximations of infinity but

are likely to cause overflow errors; furthermore, there are cases when a moderate choice performs better than a very large one.

Notice that the specificity of the fractional case in this approach consists solely in the definition of s and t , in (35).

6 Application example

The exact frequency response of $1/(1+s^{0.5})$ at 0.1, 1, and 10 rad/s was reckoned and the methods of the previous sections were used to reconstruct the function [8]. As Tables 1 and 2 shows, this is usually possible, provided that a compatible structure is offered. Since noise is usually present in experimental data, the frequency response was added a Gaussian distributed, zero mean noise, with a 1 dB or 1 degree variance, and the identification procedure repeated, to check how this may affect the result. Table 3 shows that this does not necessarily prevent a reasonable approximation of the original transfer function to be found, but the structure offered needs to be closer to the correct one.

Tables present an index showing how close the frequency response of the identified model is to the data from which the model was obtained. It is given by

$$J = \frac{1}{f} \sum_{i=1}^f [G(j\omega) - \hat{G}(j\omega)]^2 \quad (65)$$

Insignificant values of J appear when only slight numerical discrepancies exist; higher values reflect the lack of quality of the model identified.

Results of the iterative method of Lawrence and Rogers are not shown because, if the initial conditions in (73) are assumed, it is necessary to have data from many frequencies to get any acceptable results. Actually the best way of using that iterative method is to combine it with the weights of Sanathanan and Koerner's method.

7 Comments

In short, this extension of Levy's method and its improvements appear to enjoy the same merits and suffer from the same drawbacks of the original integer-order versions. They namely require providing in advance the orders of the numerator and the denominator, n and m ; and these extensions also require q , the commensurate order. Of course, numerical problems usually arise when excessively high values for n and m or excessively low values for q are provided. There are two possible solutions for dealing with this requirement: a visual inspection of frequency data may suggest the appropriate orders; or several possible combinations of values may be tried, and the best retained. This last option is possible because the algorithm runs fast enough in modern computers.

Future work should include adapting other identification methods known to be less sensitive to noise, and of stochastic nature.

Table 1. Identification results from the exact response; $q = 0.5$, $n = m = 1$

Levy	Vinagre	Sanathanan and Koerner, 2 iterations
$\frac{1.0000}{1+1.0000s^{0.5}}$	$\frac{1.0000}{1+1.0000s^{0.5}}$	$\frac{1.0000}{1+1.0000s^{0.5}}$
$J = 1.3225 \times 10^{-31}$	$J = 5.5731 \times 10^{-30}$	$J = 5.6494 \times 10^{-33}$

Table 2. Identification results from the exact response; $q = 0.25$, $n = m = 2$

Levy	Vinagre	Sanathanan and Koerner, 2 iterations
$\frac{1.0000}{1+1.0000s^{0.5}}$	$\frac{1.0000}{1+1.0000s^{0.5}}$	$\frac{1.0000}{1+1.0000s^{0.5}}$
$J = 7.7515 \times 10^{-29}$	$J = 3.6635 \times 10^{-28}$	$J = 9.9147 \times 10^{-30}$

Table 3. Identification results from the exact response; $q = 0.5$, $n = m = 1$

Levy	Vinagre	Sanathanan and Koerner, 4 iterations
$\frac{0.9940 + 0.0091s^{0.5}}{1 + 0.9384s^{0.5}}$	$\frac{0.9949 + 0.0355s^{0.5}}{1 + 0.9656s^{0.5}}$	$\frac{0.9951 + 0.0082s^{0.5}}{1 + 0.9726s^{0.5}}$
$J = 4.2694 \times 10^{-4}$	$J = 9.4836 \times 10^{-4}$	$J = 3.4494 \times 10^{-4}$

References

1. Levy E (1959) Complex curve fitting, IRE trans. Automatic Control, 4:37–43.
2. Oustaloup A (1991) *La Commande CRONE: Commande Robuste D'ordre Non Entier* (in French). Hermes, Paris.
3. Podlubny I (1999) *Fractional Differential Equations*. Academic Press, San Diego.
4. Hartley T, Lorenzo C (2003) Fractional-order system identification based on continuous order-distributions, Signal Processing, 83:2287–2300.
5. Vinagre B (2001) Modelado y control de sistemas dinámicos caracterizados por ecuaciones integro-diferenciales de orden fraccional (in Spanish). PhD thesis, UNED, pp. 140–141.
6. Sanathanan CK, Koerner J (1963) Transfer function synthesis as a ratio of two complex polynomials, IEEE Trans. Automatic Control, 8:56–58.
7. Lawrence PJ, Rogers G (1979) Sequential transfer-function synthesis from measured data, Proc. IEE, 126(1):104–106.
8. Valério D, Sá da Costa J (2004) Ninteger: a non-integer control toolbox for MatLab. In: *Fractional Derivatives and Applications*. IFAC, Bordeaux.

DYNAMIC RESPONSE OF THE FRACTIONAL RELAXOR–OSCILLATOR TO A HARMONIC DRIVING FORCE

B. N. Narahari Achar¹ and John W. Hanneken²

¹ University of Memphis, Physics Department, Memphis, TN 38152; Tel: (901)678-3122
Fax: (901)678-4733; E-mail: nachar@memphis.edu

² University of Memphis, Physics Department, Memphis, TN 38152; Tel: (901)678-2417
Fax: (901)678-4733; E-mail: jhannekn@memphis.edu

Abstract

The so-called fractional relaxor–oscillator, whose time evolution is characterized by an index of fractional order, α , exhibits interesting relaxation–oscillation characteristics. For the range of values $0 < \alpha < 1$, the system exhibits some characteristics of a regular relaxor, and for the range $1 < \alpha < 2$, some characteristics of a damped harmonic oscillator. But, when it is subjected to a sinusoidal forcing, there are characteristic features in the dynamic response, which have no parallel either in the regular relaxor or the damped harmonic oscillator. The system is characterized by a frequency-dependent “relaxation time constant” in the range $0 < \alpha < 1$, and an associated phase lag. In the range $1 < \alpha < 2$, there is a frequency-dependent damping parameter and an associated phase lag. The two phase lags approach each other in the limit $\alpha \rightarrow 1$ from either side of 1. Furthermore, there is a different power-law tail associated with each of these cases

Keywords

Fractional relaxor–oscillator, dynamic damping.

1 Introduction

It is well known that fundamental laws of physics [1,2] which can be formulated as equations for the time evolution of a quantity $X(t)$ in the form

$$\frac{dX(t)}{dt} = -AX(t) , (A > 0) \quad (1)$$

can be generalized [3–5] into kinetic equations of the form

$$\frac{d^\alpha X(t)}{dt^\alpha} = -AX(t) , \quad (2)$$

by replacing the first-order derivative by a fractional-order derivative of order $\alpha > 0$. When the index $\alpha = 1$, this equation represents a relaxation process described by the solution

$$X(t) = X(0)\exp(-t/\tau) \quad (3)$$

with a characteristic time scale $\tau = A^{-1}$ for the exponential decay. When the index $\alpha = 2$, the equation represents a simple harmonic oscillator, which for the initial conditions $X(0) = X_0$, $\dot{X}(0) = 0$ yields the solution

$$X(t) = X_0 \cos(\omega_0 t) \quad (4)$$

with the natural frequency of oscillation, given by $\omega_0 = \sqrt{A}$.

When the value of the index lies in the range $0 < \alpha < 1$, Eq. (2) refers to the “fractional relaxor” (FR), whose time evolution is described by the Mittag-Leffler function, which interpolates between a stretched exponential behavior and an asymptotic power law behavior and has been studied extensively by Caputo and Mainardi [6,7] and by Gloeckle and Nonenmacher [8]. When the value lies in the range $1 < \alpha < 2$, Eq. (2) represents the so-called fractional oscillator (FO), which has also been extensively studied [3,9–12]. The time evolution of the system is again described by a Mittag-Leffler function. The FO behaves dynamically like a damped harmonic oscillator, and in the limit of $\alpha \rightarrow 2$, it behaves like a simple harmonic oscillator with no damping. Whereas the damping in a damped harmonic oscillator is due to an external frictional force proportional to the velocity, in a FO, the damping is intrinsic and is described by a damping parameter

$$\beta = -\omega_0 \cos(\pi/\alpha) , \quad (5)$$

where ω_0 is the “natural frequency”, and α is the index of the fractional integral in the equation of motion.

It has been shown [11] that very interesting response characteristics are exhibited by a FO subject to a sinusoidal driving force. As the transients die out

and the steady-state oscillations are established, the rate of supply of energy by the driving force must equal the rate of loss of energy by damping. The damping in a FO is found to be [11] dynamic in nature and that “free” and “forced” oscillations are characterized by different damping parameters. Furthermore, in each of these cases of damping, there is a characteristic tail obeying a different algebraic power law. It appears that a simple description by a “quality factor”, Q , in analogy with a damped harmonic oscillator may not be adequate in view of the complex nature of the damping.

Detailed analysis for the FR is not yet well known. It is the purpose of the present work to study the dynamic response of a fractional relaxor when subjected to a sinusoidal forcing with a view to elicit the frequency dependence of the fractional relaxation phenomenon and to study in particular, what aspects of the response are continuous across the range of values of α . The plan of the paper is as follows. First a brief review of the dynamic response of a regular relaxor subject to a sinusoidal driving force is given in a formulation based on integral equations, with a view to establish notation for subsequent generalization to the case of a fractional relaxor. The next section deals with the dynamic response of a fractional relaxor, which starts from rest at $t = 0$ and is subject to a sinusoidal driving force, when the FR is characterized by a frequency-dependent “relaxation time constant”. The following section presents a short account of the dynamic response of a FO, when subject to a sinusoidal driving force. The final section discusses the dynamic behavior in the entire range $0 < \alpha < 2$.

2 Response of a Regular Relaxor

The integral equation of motion given by

$$x(t) = -\frac{1}{\tau_0} \int_0^t x(t') dt' + \int_0^t f(t') dt' \quad (6)$$

represents a regular relaxor, which starts from rest at $t = 0$ and is driven by a periodic forcing function. Here τ_0 is the characteristic relaxation time constant, and $f(t) = F \sin(\omega t)$ is the sinusoidal forcing function, with F having the dimensions of (L/T). Applying Laplace transform to both sides of Eq. (6) yields

$$\tilde{x}(s) = -\frac{\tilde{x}(s)}{\tau_0 s} + \frac{F\omega}{s(s^2 + \omega^2)} \quad (7)$$

Solving for $\tilde{x}(s)$ in Eq. (7) yields

$$\tilde{x}(s) = \frac{F\omega}{(s^2 + \omega^2)(s + \frac{1}{\tau_0})} \quad (8)$$

Taking the inverse Laplace transform and simplifying yields

$$x(t) = \frac{F\omega}{(\omega^2 + \frac{1}{\tau_0^2})} e^{-\frac{t}{\tau_0}} + \frac{F}{(\omega^2 + \frac{1}{\tau_0^2})^{1/2}} \sin(\omega t - \delta) \quad (9)$$

where

$$\delta = \tan^{-1}(\omega\tau_0) \quad (10)$$

Equation (9) describes the well known result that when the relaxor is driven by a sinusoidal force, there is a transient described by the first term on the RHS decaying with the characteristic time constant τ_0 and a steady-state oscillation, described by the second term on the RHS, oscillating with the same frequency as the driving force but with a reduced amplitude and a phase lag which is given by Eq. (10).

3 Response of a Fractional Relaxor

The integral equation of motion of a driven fractional relaxor can be obtained from Eq. (6) by generalizing to the corresponding fractional integral of order α [13–16] as

$$x(t) = -\frac{\tau_0^{-\alpha}}{\Gamma(\alpha)} \int_0^t (t-t')^{\alpha-1} x(t') dt' + \frac{1}{\Gamma(\alpha)} \int_0^t (t-t')^{\alpha-1} f(t') dt' \quad (11)$$

with $0 < \alpha < 1$. It is to be noted that from dimensional considerations τ_0^α replaces the relaxation time constant in Eq. (6). As before,

$$f(t) = F \sin \omega t \quad (12)$$

and now, τ_0 and F have the dimensions T and (L/T $^\alpha$) respectively. Taking Laplace transforms on both sides of Eq. (11) yields

$$\tilde{x}(s) = -\frac{\tilde{x}(s)}{\tau_0^\alpha s^\alpha} + \frac{F\omega}{s^\alpha (s^2 + \omega^2)} \quad (13)$$

Solving for $\tilde{x}(s)$ yields

$$\tilde{x}(s) = \frac{F\omega}{(s^2 + \omega^2)(s^\alpha + \frac{1}{\tau_0^\alpha})} \quad (14)$$

Taking the inverse Laplace transform on both sides, and noting that the inverse Laplace transform on the RHS can be written as a convolution integral, the response is given in terms of the Greens function solution [17].

$$x(t) = \int_0^t F \sin(\omega t') G(t - t') dt' \quad (15)$$

where the Greens function solution is given by

$$G(t) = t^{\alpha-1} E_{\alpha,\alpha}(-\tau_0^{-\alpha} t^\alpha) \quad (16)$$

where $E_{\alpha,\alpha}(-\tau_0^{-\alpha} t^\alpha)$ is a generalized Mittag-Leffler function defined by [18]

$$E_{\alpha,\beta}(z) = \sum_{k=0}^{\infty} \frac{z^k}{\Gamma(\alpha k + \beta)} \quad (17)$$

with an associated Laplace transform given by:

$$t^{\beta-1} E_{\alpha,\beta}(-at^\alpha) \div \frac{s^{\beta-\alpha}}{(s^\alpha + a)} \quad (18)$$

The response function is given by

$$x(t) = F \int_0^t (t - t')^{\alpha-1} E_{\alpha,\alpha}(-\tau_0^{-\alpha} (t - t')^\alpha) \sin \omega t' dt' \quad (19)$$

It can be shown [10] that an explicit solution for Eq. (19) can be obtained by appealing to the theory of complex variables and can be expressed as the sum of two contributions in a Bromwich integral:

$$x(t) = x_1(t) + x_2(t) \quad (20)$$

$x_1(t)$ arises from a Hankel loop consisting of a small circle of radius r and the two lines parallel to the negative x -axis in the Bromwich integral and is given by

$$x_1(t) = \int_0^{\infty} \exp(-rt) K_\alpha(r, \tau_0^{-\alpha}, \omega) dr \quad (21)$$

with

$$K_\alpha(r, \tau_0^{-\alpha}, \omega) = \frac{F \omega r^\alpha \sin \pi \alpha}{\pi(r^2 + \omega^2)(r^{2\alpha} + 2r^\alpha \tau_0^{-\alpha} \cos \pi \alpha + \tau_0^{-2\alpha})} \quad (21a)$$

The second part $x_2(t)$ arises from the residues of the poles at $s = \pm i\omega$ of the integrand in the Bromwich integral, which are the only ones contributing to the integral which can be expressed as:

$$x_2(t) = A_2 \sin(\omega t - \delta) \quad (22)$$

where

$$A_2 = \frac{F}{\left(\tau_0^{-2\alpha} + \omega^{2\alpha} + 2\tau_0^{-\alpha} \omega^\alpha \cos(\pi\alpha/2)\right)^{\frac{1}{2}}} \quad (23)$$

and

$$\delta = \tan^{-1} \left[\frac{\omega^\alpha \sin(\pi\alpha/2)}{\tau_0^{-\alpha} + \omega^\alpha \cos(\pi\alpha/2)} \right] \quad (24)$$

respectively. The details of the derivation can be found in [10]. The relaxation characteristics are contained in the Eqs. (20)–(24).

4 Response of a Fractional Oscillator

For purposes of comparison with a FR, the relevant results for the response characteristics of a FO subject to a sinusoidal forcing have been reproduced from the detailed studies in [10,11]. The integral equation of motion of a driven FO is given by

$$x(t) = -\frac{\omega_0^\alpha}{\Gamma(\alpha)_0} \int_0^t (t-\tau)^{\alpha-1} x(\tau) d\tau + \frac{1}{\Gamma(\alpha)_0} \int_0^t (t-\tau)^{\alpha-1} f(\tau) d\tau \quad (25)$$

with $1 < \alpha \leq 2$. For the initial conditions that the FO is at rest at the equilibrium position, the response is given in terms of the Greens function solution. For a sinusoidal forcing function given by $f(t) = F \sin \omega t$, the Greens function solution is given by

$$G(t) = t^{\alpha-1} E_{\alpha,\alpha}(-\omega_0^\alpha t^\alpha) \quad (26)$$

where $E_{\alpha,\alpha}(-\omega_0^\alpha t^\alpha)$ is a generalized Mittag-Leffler function defined in Eq. (17). The response function is given by

$$x(t) = F \int_0^t (t-\tau)^{\alpha-1} E_{\alpha,\alpha}(-\omega_0^\alpha (t-\tau)^\alpha) \sin \omega \tau d\tau \quad (27)$$

It has been shown [10,11] that an explicit solution can be obtained by appealing to the theory complex variables and can be expressed as the sum of three contributions in a Bromwich integral:

$$x(t) = x_1(t) + x_2(t) + x_3(t) \quad (28)$$

$x_1(t)$ arises from a Hankel loop consisting of a small circle of radius r and the two lines parallel to the negative x -axis in the Bromwich integral and is given by

$$x_1(t) = \int_0^\infty \exp(-rt) K_\alpha(r, \omega_0^\alpha, \omega) dr \quad (29)$$

with

$$K_\alpha(r, \omega_0^\alpha, \omega) = \frac{F \omega r^\alpha \sin \pi \alpha}{\pi(r^{2\alpha} + \omega^2)(r^{2\alpha} + 2r^\alpha \omega_0^\alpha \cos \pi \alpha + \omega_0^{2\alpha})} \quad (30)$$

The remaining parts $x_2(t)$, and $x_3(t)$ arise from the residues of the poles of the integrand in the Bromwich integral and can be expressed as follows:

$$x_2(t) = A_2 \exp(-\beta t) \cos(\omega_0 t \sin(\pi / \alpha) - \phi_2) \quad (31)$$

where
$$A_2 = \frac{2F\omega}{\alpha \omega_0^{\alpha-1} (\omega_0^4 + \omega^4 + 2\omega_0^2 \omega^2 \cos(2\pi / \alpha))^{\frac{1}{2}}} \quad (32)$$

$$\beta = -\omega_0 \cos(\pi / \alpha) \quad (33)$$

$$\phi_2 = \arctan \left[\frac{\omega_0^2 \sin(\frac{\pi(1+\alpha)}{\alpha}) - \omega^2 \sin(\frac{\pi(1-\alpha)}{\alpha})}{\omega_0^2 \cos(\frac{\pi(1+\alpha)}{\alpha}) + \omega^2 \cos(\frac{\pi(1-\alpha)}{\alpha})} \right] \quad (34)$$

$$x_3(t) = A_3 \sin(\omega t - \delta) \quad (35)$$

where
$$A_3 = \frac{F}{(\omega_0^{2\alpha} + \omega^{2\alpha} + 2\omega_0^\alpha \omega^\alpha \cos(\pi \alpha / 2))^{\frac{1}{2}}} \quad (36)$$

$$\delta = \arctan \left[\frac{\omega^\alpha \sin(\pi \alpha / 2)}{\omega_0^\alpha + \omega^\alpha \cos(\pi \alpha / 2)} \right] \quad (37)$$

respectively. The details of the derivation can be found in [10,11]. The damping characteristics are contained in the Eqs. (29)–(37). The following sections are devoted to a detailed analysis of the relaxation in a FR and its comparison with the results for a FO.

5 Relaxation Processes in a Driven FR

5.1 Limit $\alpha \rightarrow 1$

In this limit, the Greens function in Eq. (16) reduces to $G(t) = E_{1,1}(-t/\tau_0) = e^{-\frac{t}{\tau_0}}$ and the response function in Eq. (19) reduces to $x(t) = F \int_0^t e^{-\frac{(t-t')}{\tau_0}} \sin \omega t' dt'$ which simplifies to the response given in Eq. (9) showing that FR reduces to a regular relaxor.

5.2 Transients

In the long time limit, the sinusoidal driving force continues to operate, the transients die out and a steady-state oscillation at the driving frequency is eventually established. In the case of the driven FR, the decay of the transients is described by a monotonically time-dependent function described by $x_1(t)$ in Eq. (21) and Eq. (21a). No simple closed form expression can be given for this part, but an analysis of the asymptotic behavior can be made and the integral itself can be evaluated numerically. As such a few general remarks can be made about the monotonic contribution to the decay of transients. As is obvious from an examination of Eq. (21a), the kernel K is a function of F , α , τ_0^α , and ω and arises only when there is a forcing function as it depends on the amplitude F of the forcing function. For the case $0 < \alpha < 1$ considered here, the kernel K is always positive. Thus the contribution $x_1(t)$ to the response decreases monotonically to zero and exhibits an asymptotic behavior [3] given by

$$x_1(t) \sim \frac{t^{-\alpha}}{\Gamma(1-\alpha)} \quad (38)$$

5.3 Steady-State Solution

In the steady state solution given by Eq. (22) namely $A_2 \sin(\omega t - \delta)$, the amplitude is constant, and the time-dependent part oscillates at the driving

frequency ω , the oscillations persisting long after the transients have died out. The ratio of the amplitude of the response to the amplitude of the forcing function depends on the forcing frequency and in addition, a phase lag is introduced. Figure 1 shows the variation of the response amplitude (in dimensionless form) as a function of the forcing frequency (also in dimensionless form). Figure 2 shows the variation of the phase angle as a function of the forcing frequency. For purposes of comparison with a fractional oscillator, the amplitude response in Eq. (36) and the phase angle in Eq. (37) for the case $\alpha = 1.5$ are also shown in Fig. 1, and Fig. 2 respectively, with $1/\omega_0$ replacing τ_0 in the dimensionless quantities referred to in the axes.

By comparing the expression for the phase lag factor in Eq. (24) with the corresponding factor for a regular relaxor given in Eq. (9), it is suggested that a frequency-dependent relaxation parameter be defined so as to yield a phase lag given by

$$\delta = \tan^{-1}(\omega \tau_{dyn}) \quad (39)$$

This means that the dynamic relaxation parameter is to be defined as:

$$\tau_{dyn} = \left(\frac{\omega^{\alpha-1} \tau_0^\alpha \sin(\pi\alpha/2)}{\omega^\alpha \tau_0^\alpha \cos(\pi\alpha/2) + 1} \right) \quad (40)$$

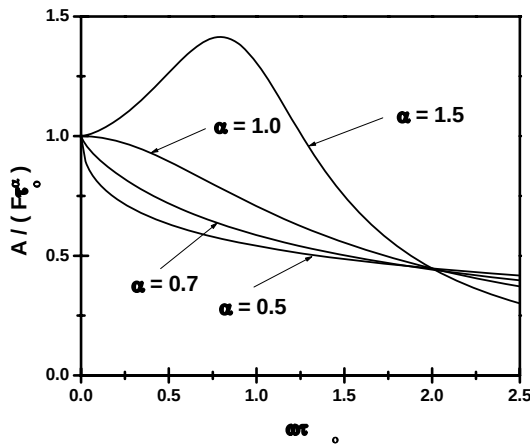


Fig. 1. Steady-state amplitude response for sinusoidal forcing.

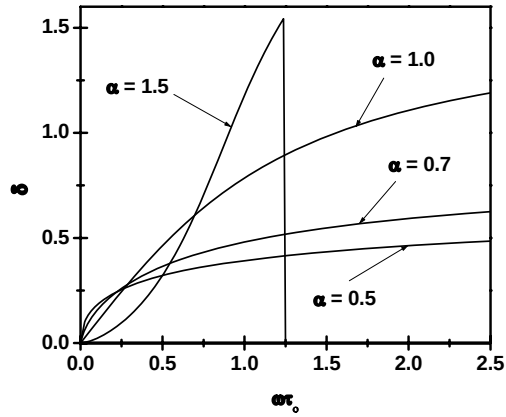


Fig. 2. Steady-state phase angle for sinusoidal forcing.

This is analogous to the frequency-dependent dynamic damping factor in the case of the to be discussed later. Figure 3 shows the variation of the dynamic relaxation parameter as a function of the forcing frequency.

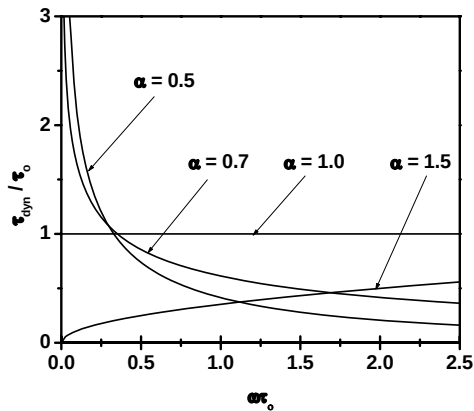


Fig. 3. Frequency-dependent relaxation parameter.

6 Damping Processes in a Driven FO

The damping processes in a FO are summarized below from detailed studies in [10,11].

6.1 Limits $\alpha \rightarrow 1$ and $\alpha \rightarrow 2$

Since the Green's function for the driven FO in Eq. (26) is identical to the Green's function in Eq. (16) for the FR, it is obvious that the dynamical behavior of the FO approaches that of a regular relaxor, just as the FR as discussed earlier. In the limit $\alpha \rightarrow 2$, the Green's function in Eq. (26) reduces to

$$G(t) = tE_{2,2}(-\omega_0^2 t^2) = \frac{\sin(\omega_0 t)}{\omega_0},$$

which is the Green's function for a simple harmonic oscillator. Thus in this limit, the dynamical behavior of the FO reduces to that of a simple harmonic oscillator with no damping.

6.2 Dynamic response

As already noted the dynamic response of the FO to sinusoidal forcing consists of three contributions given in Eq. (28), of which there are two transient contributions:

(i) a monotonically time-dependent part $x_1(t)$ given in Eq. (29) with an asymptotic behavior [3] given by

$$x_1(t) \sim -\frac{t^{-\alpha-1}}{\Gamma(-\alpha)} \quad (41)$$

and the contribution goes to zero in the limit $\alpha \rightarrow 2$.

(ii) an oscillatory contribution to the transient given in Eq. (31) namely,

$$A_2 \exp(-\beta t) \cos(\omega_0 t \sin(\pi / \alpha) - \phi_2), \quad (42)$$

which oscillates with a frequency

$$\omega' = \omega_0 \sin(\pi / \alpha) \quad (43)$$

but with an exponential decay is described by the parameter β given in Eq. (33)

(iii) The third contribution is a steady-state part given by Eq. (35), which oscillates with the driving frequency ω . The damping parameter involved in the steady state, is given by

$$\beta_{dyn} = \frac{\omega^\alpha \sin(\pi\alpha/2)}{2\omega} \quad (44)$$

as the effective damping parameter in the steady state. This dynamic damping parameter is different from the one defined in Eq. (22) of [11]. Figure 3 also shows the variation of the dynamic damping parameter in Eq. (44) for a particular value of the parameter, $\alpha = 1.5$, as a function of the frequency (with the appropriate change in dimensionless quantities for the axes as noted earlier).

7 Discussion

It is clear that the dynamic behavior of both FR and FO approach that of a regular relaxor in the limit of $\alpha \rightarrow 1$ and that the FO approaches the dynamical behavior of a simple harmonic oscillator in the limit $\alpha \rightarrow 2$. It is obvious that the dynamic relaxation parameter, τ_{dyn} in Eq. (40) is different from the relaxation parameter τ_0^α in Eq. (11) for the FR. It is also obvious that the dynamic damping parameter β_{dyn} in Eq. (49) is different from the damping parameter β in Eq. (5) for the FO. The parameter in Eq. (11) pertains to the discussion of “natural” relaxation of the FR as well as the decay of the transients when the FR is driven sinusoidally. Similarly, the parameter in Eq. (5) pertains to the decay of “free” oscillations of the FO, as well as the decay of the transients arising in the case of the sinusoidally driven FO. In contrast, τ_{dyn} and β_{dyn} pertain to the discussion of steady-state oscillations when the energy supplied by the driving force is “dissipated” at the same rate so as to maintain a constant amplitude in the response. In addition, there is introduced a phase lag in the response. In the limit of $\alpha \rightarrow 1$ the phase lags for the FR and FO approach each other. In the limit $\alpha \rightarrow 2$, $\beta_{dyn} \rightarrow 0$ just as expected for a simple harmonic oscillator. It is also obvious that a single “quality factor” cannot be defined as a measure of the quality of resonance for a FO in view of the frequency dependence of the damping parameter. Finally it has been shown that there are different power laws describing the asymptotic decay tails.

8 Conclusions

A rich variety of relaxation and damping characteristics is exhibited by the fractional relaxor–oscillator system. While the “free” relaxation/oscillations can be described by means of a single relaxation/damping parameter, the “forced” oscillations by a sinusoidal driving force are characterized by a frequency-dependent relaxation parameter τ_{dyn} in the case of the FR, and by a frequency-dependent damping parameter β_{dyn} in the case of the FO. The phase lag changes continuously across the whole range of values of the fractional parameter. The dynamic behavior approaches that of a regular relaxor or a simple harmonic oscillator in the appropriate limits.

References

1. Hilfer R (2000) Fractional time evolution, in: Hilfer R (ed.), *Applications of Fractional Calculus in Physics*. World Scientific, Singapore, pp. 87–130.
2. Metzler R, Klafter J (2000) The Random walk’s Guide to Anomalous Diffusion: A Fractional dynamics Approach, *Physics Reports*, 339, pp. 1–77.
3. Gorenflo R, Mainardi F (1997) Fractional calculus: integral and differential equations of fractional order, in: Carpinteri A, Mainardi F (eds.), *Fractals and Fractional Calculus in Continuum Mechanics*, Springer, Wien, pp. 223–276; <http://www.fracalmo.org>. Gorenflo R, Mainardi F (1996) Fractional Oscillations and Mittag-Leffler Functions, Preprint No A-14/96, Fachbereich Mathematik und Informatik, Freie Universitaet Berlin, (<http://www.math.fuberlin.de/publ/index.html>), and Gorenflo R, Mainardi F (1996) Fractional Oscillations and Mittag-Leffler Functions, International Workshop on the Recent Advances in Applied Mathematics (RAAM96), State of Kuwait May 4–7, 1996, Proceedings, Kuwait University, pp. 193–208.
4. Gorenflo R, Rutman R (1995) On Ultraslow and on intermediate processes, in: Rusev P, Dimovski I, Kiryakova V (eds.), *Transform Methods and Special Functions*, Sophia 1994, Science Culture and Technology, Singapore.
5. Mainardi F (1996) Chaos, Solitons Fractals, 7(9):1461–1477.
6. Caputo M, Mainardi F (1971) *Linear Models of Dissipation in Anelastic Solids*, Rivista. Nuovo Cimento (Ser II), 1, pp. 161–198.
7. Mainardi F (1994) Fractional relaxation in anelastic solids, *J. Alloys Comp.*, 211/212:534–538.
8. Gloeckle WG, Nonenmacher TF (1991) Fractional integral operators and fox functions in the theory of viscoelasticity, *Macromolecules*, 24:6426–6434.

9. Narahari Achar BN, Hanneken JW, Enck T, Clarke T (2001) Dynamics of the fractional oscillator, *Physica A*, 297:361–367.
10. Narahari Achar BN, Hanneken JW, Clarke T, (2002), Response characteristics of the fractional oscillator, *Physica A*, 309:275–288. There is a factor r^α missing in Eq. (27) of this reference. The corrected form of the kernel is given in Eq. (21a) in the current paper.
11. Narahari Achar BN, Hanneken JW, Clarle T (2004) Damping characteristics of the fractional oscillator, *physica A*, 339:311–319. A typographical error in Eq. (13) of this reference has been corrected and the corrected equation appears as Eq. (37) in the current paper.
12. Tofighi A (2003) The intrinsic damping of the fractional oscillator, *Physica A*, 329:29–34.
13. Oldham KB, Spanier J (1974) *The Fractional Calculus*. Academic Press, New York.
14. Miller K, Ross B (1993) *An Introduction to the Fractional Calculus and Fractional Differential Equations*. Wiley, New York.
15. Podlubny I (1999) *Fractional Differential Equations*. Academic Press, San Diego.
16. Samko SG, Kilbas AA, Marichev OI (1993) *Fractional Integrals and Derivatives: Theory and Applications*. Gordon and Breach, Amsterdam.
17. Sneddon IN (1972) *The Use of Integral Transforms*. McGraw-Hill, New York.
18. Erdélyi A (1955) *Higher Transcendental Functions*, Vol. III. McGraw- Hill, New York.
19. Marion JB, Thornton ST (1995) *Classical Dynamics of Particles and Systems* 4th edition. Saunders, Fort Worth.

A DIRECT APPROXIMATION OF FRACTIONAL COLE–COLE SYSTEMS BY ORDINARY FIRST-ORDER PROCESSES

Markus Haschka and Volker Krebs

Institut für Regelungs- und Steuerungssysteme, Universität Karlsruhe (TH),
Karlsruhe, Germany; E-mail: haschka@irs.uni-karlsruhe.de,
krebs@irs.uni-karlsruhe.de

Abstract

Cole–Cole systems are widely used in electrochemistry to represent impedances of galvanic elements like fuel cells. For system analysis of Cole–Cole systems in the time domain fractional calculus has to be applied. A plain representation of fractional differential equations of Cole–Cole systems by means of conventional ordinary differential equations is addressed in this contribution. Usually in literature, the operator for the fractional derivation is approximated to ensure that the fractional system can be represented by conventional differential equations of integer order. This article presents a new approach which results in a direct approximation of the Cole–Cole system itself by conventional linear time invariant systems. The method considered is based on the distribution density of relaxation times of conventional first-order processes. This distribution density is an alternative representation of the transfer behavior of such systems. Several approximation methods, based on an analysis of the distribution density, are presented in this work. The feasibility of these methods will be demonstrated by a comparison of the approximation model to a reference model for a solid oxide fuel cell (SOFC), respectively.

Keywords

Cole–Cole systems, fractional Calculus.

1 Introduction

Cole–Cole systems are commonly used in electrochemistry to model impedance measurements in the frequency domain [1, 2]. They have been introduced by Cole and Cole [1] to represent the dispersion and absorption in liquids and dielectrics. The transfer function of a Cole–Cole system is given by

$$Z_{CC}(j\omega) = \frac{R}{1 + (j\omega\tau_0)^\alpha} \quad (1)$$

with the parameters $R, \tau_0 \in \mathbb{R}_+$ and $\alpha \in (0, 1]$. Today, this kind of system representation is widely used to analyze electrical impedance spectroscopy (EIS) measurements of fuel cells and will be used for model-based diagnosis purposes. In [3] it is shown that a serial connection of Cole–Cole systems is sufficient to represent the impedance of a solid oxide fuel cell (SOFC). The application of Cole–Cole systems can be explained physically by the occurrence of a fractal geometry in porous materials.

Cole–Cole system from measurement data possible to apply a nonlinear optimization program to minimize a performance index for the deviations of the impedance measurement from the model data [3]. This procedure yields a highly accurate result, but is very time-consuming. Additionally, the measurement equipment needed is quite expensive. Thus, often a frequency domain approach is not appropriate for a broad application of impedance-based online diagnosis for SOFCs. Alternative approaches are proposed in [3], where fractional calculus was used to represent the input–output behavior of the impedance in the time domain. However, a drawback of the time domain analysis using fractional calculus in [3] is the difficult numerical computation of the so-called Mittag-Leffler function which is required to compute the impulse response of the impedance model. Oustaloup developed a procedure in [4] to represent the fractional derivation by phase-lead and phase-lag elements. Using those elements, it is possible to simulate and to identify Cole–Cole type systems in the time domain.

In this article, a different approach is used which is based on the concept of the distribution density of relaxation times of first-order models. With this procedure it will be possible to represent the fractional-order impedance

$$Z(j\omega) = \sum_{\kappa=1}^N \frac{R_\kappa}{1 + (j\omega\tau_{0,\kappa})^{\alpha_\kappa}} \quad (2)$$

with N fractional Cole–Cole elements by a finite sum of conventional first order processes

$$Z(j\omega) \simeq \hat{Z}(j\omega) = \sum_{k=1}^M \frac{r_k}{1 + j\omega\tau_k}, \quad M \gg N. \quad (3)$$

Hence, with this new direct approximation method, the parameters r_k and τ_k ($k = 1, 2, \dots, M$) of the approximation model (3) will be determined in a direct dependency of the parameters R_κ , $\tau_{0,\kappa}$ and α_κ ($\kappa = 1, 2, \dots, N$) of the fractional-order model (2).

In the following section 2, the concept of the distribution density of relaxation times is outlined. Based on the distribution density of Cole–Cole systems, it

is possible to derive the approximation (3) of Cole–Cole type systems. The development of this direct approximation is accomplished in section 3. In section 4, the impedance of a SOFC is simulated using a direct approximation model to demonstrate the feasibility of the proposed method. Finally, the last section gives a conclusion and an outlook on future work and suggests possible applications.

2 Distribution Density of Relaxation Times

The proposed procedure for the approximation of the fractional impedance (2) is based on the concept of the distribution density of relaxation times [2]. The connection between an impedance $Z(j\omega)$ and its related distribution density $\gamma(\tau)$ of relaxation times τ is defined by the complex-valued integral

$$Z(j\omega) = \int_0^\infty \frac{\gamma(\tau)}{1 + j\omega\tau} d\tau. \quad (4)$$

A relaxation process in the impedance $Z(j\omega)$ is indicated by a peak in the density $\gamma(\tau)$. The impedance is a linear time-invariant and causal system. It can be shown that all systems with this properties are uniquely represented by the real part $\text{Re}\{Z(j\omega)\}$ or the imaginary part $\text{Im}\{Z(j\omega)\}$ of the transfer function $Z(j\omega)$, respectively. This can be explained by the possibility to obtain the transfer function by using the relation

$$Z(j\omega) = \mathcal{H}\{\text{Im}\{Z(j\omega)\}\} + j\text{Im}\{Z(j\omega)\}, \quad (5)$$

where $\mathcal{H}$ denotes the Hilbert-transformation operator, defined by the Cauchy principal-value integral

$$\mathcal{H}\{\phi(x)\} = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\phi(x)}{y - x} dx.$$

Hence, if the imaginary part of an impedance is known, the complete impedance can be deduced using relation (5). After a decomposition of (4) into real and imaginary part, the noncomplex integral

$$\text{Im}\{Z(j\omega)\} = \int_0^\infty \gamma(\tau) \frac{-\omega\tau}{1 + \omega^2\tau^2} d\tau \quad (6)$$

can be written, which gives the relation between the imaginary part $\text{Im}\{Z(j\omega)\}$ and its corresponding distribution density $\gamma(\tau)$.

Usually, the time constants of relaxation processes differ in decades, which motivates the use of a logarithmic scale for the relaxation time τ and the frequency ω . Therefore, the new variables

$$\Omega = \ln\left(\frac{\omega}{\omega_0}\right) \quad \text{and} \quad T = \ln(\omega_0\tau)$$

are introduced in order to achieve a logarithmic scaling. The frequency ω_0 is an arbitrary normalization frequency which is usually set equal to 1s^{-1} . Using the new variables Ω and T , the integral (6) can finally be transformed to

$$\text{Im} \{ Z(j\omega_0 e^\Omega) \} = - \int_{-\infty}^{\infty} \underbrace{\gamma \left(\frac{e^T}{\omega_0} \right) \frac{e^T}{\omega_0}}_{= \tilde{\gamma}(T)} \frac{e^{\Omega+T}}{1 + e^{2\Omega+2T}} dT. \quad (7)$$

In the following, the new function $\tilde{\gamma}(T)$ is used as a measure for the distribution density of relaxation processes over the logarithmic time T and is very important to construct the intended approximation (3). By an application of the identity $2/(e^\xi + e^{-\xi}) = \text{sech}(\xi)$, Eq. (7) can be simplified to

$$z(\Omega) = \text{Im} \{ Z(j\omega_0 e^\Omega) \} = -\frac{1}{2} \int_{-\infty}^{\infty} \tilde{\gamma}(T) \text{sech}(\Omega + T) dT. \quad (8)$$

The previous integral Eq. (8) can be solved for $\tilde{\gamma}(T)$ by the equation

$$\tilde{\gamma}(T) = -\frac{1}{\pi} \left[(z(\Omega))|_{\Omega=T+j\frac{\pi}{2}} + (z(\Omega))|_{\Omega=T-j\frac{\pi}{2}} \right] \quad (9)$$

which was first published in [5]. By using (9), it is possible to obtain – for a known impedance $Z(j\omega)$ – the corresponding distribution densities $\gamma(\tau)$ or $\tilde{\gamma}(T)$, respectively. It is not feasible to use Eq. (9) to calculate the distribution density for given impedance measurement data, because the course of the impedance has to be known exactly, which cannot be assumed for system identification purposes. However, using (9) the distribution density $\tilde{\gamma}(T)$ for a single Cole–Cole system will be determined. The transfer function (1) is separated into its real and its imaginary part, which leads to the equation:

$$\begin{aligned} Z_{CC}(j\omega) &= \frac{R}{1 + (j\omega\tau_0)^\alpha} \\ &= \frac{R [1 + (\omega\tau_0)^\alpha (\cos(\alpha\frac{\pi}{2}) - j \sin(\alpha\frac{\pi}{2}))]}{1 + 2(\omega\tau_0)^\alpha \cos(\alpha\frac{\pi}{2}) + (\omega\tau_0)^{2\alpha}}. \end{aligned} \quad (10)$$

In order to obtain $\tilde{\gamma}(T)$, it is necessary to determine the imaginary part $z_{CC}(\Omega)$ in dependency of the logarithmic frequency Ω :

$$\begin{aligned} z_{CC}(\Omega) &= -\frac{R}{2} \frac{\sin(\alpha\frac{\pi}{2})}{\cos(\alpha\frac{\pi}{2}) + \cosh(\alpha(\Omega + T_0))}, \\ &\text{with } T_0 = \ln(\omega_0\tau_0). \end{aligned} \quad (11)$$

Combined with Eq. (9), the imaginary part (11) of the impedance can be used to determine

$$\tilde{\gamma}_{CC}(T) = \frac{R}{2\pi} \frac{\sin((1-\alpha)\pi)}{\cosh(\alpha(T-T_0)) - \cos((1-\alpha)\pi)} \quad (12)$$

as the distribution density of a single Cole–Cole system. According to [3], the impedance of a single SOFC can be represented by a serial connection of a finite number of N Cole–Cole systems according to Eq. (2) in section 1. Depending on the intended accuracy, the number of Cole–Cole systems necessary for an appropriate model is between two and five. Therefore, the impedance can be represented by (2). As it was shown before, the imaginary part $z(\Omega)$ of the impedance is relevant for the calculation of the distribution density $\tilde{\gamma}(T)$. Due to the linearity of Eq. (9), the distribution densities of each single Cole–Cole system have to be summed up to form the resulting density of the complete impedance

$$\tilde{\gamma}(T) = \sum_{\kappa=1}^N \tilde{\gamma}_{CC,\kappa}(T),$$

where $\tilde{\gamma}_{CC,\kappa}(T)$ denotes the distribution density of the κ th Cole–Cole system

$$Z_{CC,\kappa}(j\omega) = \frac{R_\kappa}{1 + (j\omega\tau_{0,\kappa})^{\alpha_\kappa}}.$$

For $\alpha = 1$, a Cole–Cole system changes to a conventional RC-element, which is represented by its impedance

$$Z_{RC} = \frac{R}{1 + j\omega\tau_0}. \quad (13)$$

Using the sifting property of the Dirac δ -impulse, it can be shown that

$$\gamma_{RC}(\tau) = R\delta(\tau - \tau_0) \quad (14)$$

is the distribution density of the impedance (13) of a single RC-element, because the integral (4) evaluated using the distribution density (14) gives the impedance

$$\int_0^\infty \frac{R\delta(\tau - \tau_0)}{1 + j\omega\tau} d\tau = \frac{R}{1 + j\omega\tau_0}.$$

Hence, a δ -impulse in the distribution density $\gamma(\tau)$ indicates a conventional RC-process.

Next, the occurrence of the δ -impulse in the $\tilde{\gamma}(T)$ -distribution density has to be investigated. In the following calculation, the identity $\tilde{\gamma}(T) = R\delta(T - T_0)$ is used:

$$\begin{aligned}
\operatorname{Im}\{Z(j\omega)\} = z(\Omega) &= -\frac{1}{2} \int_{-\infty}^{\infty} \tilde{\gamma}(T) \operatorname{sech}(\Omega + T) dT \\
&= -\frac{1}{2} \int_{-\infty}^{\infty} R \delta(T - T_0) \operatorname{sech}(\Omega + T) dT \\
&= -\frac{1}{2} R \operatorname{sech}(\Omega + T_0) \\
&= -R \frac{\omega \tau_0}{1 + \omega^2 \tau_0^2}.
\end{aligned} \tag{15}$$

Equation (15) is the imaginary part of the impedance of a single RC-element (13) with the time constant $\tau_0 = e^{T_0}/\omega_0$. In the following, this fact will be used to derive an integer-order approximation for a fractional-order Cole–Cole system. For a better understanding of the concept of the distribution of relaxation times, an example with usual values for a SOFC will be given in the next paragraph.

2.1 Example: distribution density for a SOFC

The distribution of relaxation times is commonly used in the analysis of impedance measurements of fuel cells. Let

$$\begin{aligned}
Z_{SOFC}(j\omega) &= \frac{0.0017}{1 + (j\omega 9.1)^{0.54}} + \frac{0.023}{1 + (j\omega 1.1)^{0.9}} \\
&\quad + \frac{0.011}{1 + (j\omega 0.00017)^{0.54}}
\end{aligned} \tag{16}$$

be the impedance of a SOFC without its purely ohmic component. These values are taken from [3], where the fractional impedances of SOFCs were investigated. In Fig. 1, the distribution density $\tilde{\gamma}(T)$ of the impedance (16) is given.

There are peaks in the density at the logarithmic time constants

$$\begin{aligned}
T_{0,1} &= \ln(\omega_0 \tau_{0,1}) = 2.2 \\
T_{0,2} &= \ln(\omega_0 \tau_{0,2}) = 0.095 \\
T_{0,3} &= \ln(\omega_0 \tau_{0,3}) = -8.68.
\end{aligned}$$

Hence, it is possible to detect dynamic processes in the distribution density $\tilde{\gamma}(T)$ by finding these peaks for related relaxation times $T_{0,\kappa}$. The first Cole–Cole process with the time constant $T_{0,1}$ has a very small peak, which cannot be separated from the large peak of the second Cole–Cole process with the time-constant $T_{0,2}$. With this basic knowledge about the concept of the distribution density of relaxation times, it is possible to derive the new method to determine the intended approximation (3).

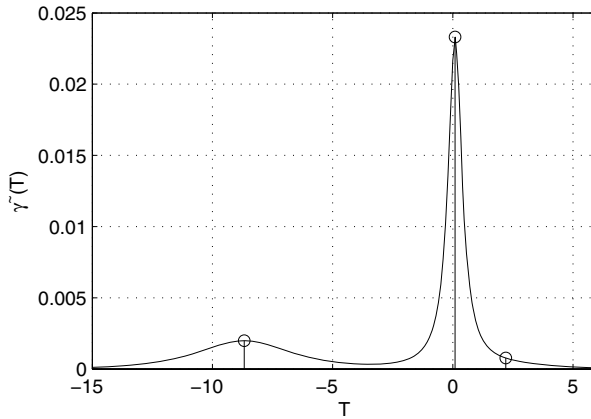


Fig. 1. Distribution density of relaxation times.

3 Derivation of the Direct Approximation

The approach for the direct approximation method is based on the integral

$$z(\Omega) = \text{Im} \{ Z(j\omega_0 e^{\Omega}) \} = -\frac{1}{2} \int_{-\infty}^{\infty} \tilde{\gamma}(T) \text{sech}(\Omega + T) dT, \quad (17)$$

which can be interpreted as a superposition of an infinite set of dynamical processes, which is able to generate the system property of the system/impedance (2). To find a finite-order approximation of this system behavior, a method for order reduction has to be applied. Therefore, only dynamical processes with a significant contribution to the system behavior have to be considered. The distribution density $\tilde{\gamma}(T)$ is converging to zero for $T \rightarrow -\infty$ and $T \rightarrow \infty$. Hence, finite integration bounds will be sufficient to represent the system (2) with an acceptable accuracy. This loss of accuracy has to be investigated by analyzing the integral

$$I(L) = \int_{T_0-L}^{T_0+L} \tilde{\gamma}(T) dT \quad (18)$$

over the distribution density (12) of a single Cole–Cole system. The quotient

$$q = \frac{I(L)}{\lim_{L \rightarrow \infty} I(L)}$$

gives the fraction of the area between the distribution density $\tilde{\gamma}(T)$ and the T -axis for limited integration bounds $T_0 \pm L$ compared to the total area for infinite bounds. This quotient

$$q(\alpha, L) = 2 \sin((\alpha - 1)\pi) \cdot \frac{\left(\arctan\left(\frac{e^{-\alpha L} + \cos(\pi\alpha)}{\sqrt{1 - \cos^2(\pi\alpha)}}\right) - \arctan\left(\frac{e^{\alpha L} + \cos(\pi\alpha)}{\sqrt{1 - \cos^2(\pi\alpha)}}\right) \right)}{\sin(\pi\alpha) \left(\pi - 2 \arctan\left(\frac{\cos(\pi\alpha)}{\sqrt{1 - \cos^2(\pi\alpha)}}\right) \right)} \quad (19)$$

can be calculated by an integration of (18).

By solving (19) for L , it is possible to determine an appropriate value for the integration limits $T_0 - L$ and $T_0 + L$ in dependency of the parameter α and the requested value for $0 < q(\alpha, L) < 1$. The selected finite lower bound of the integral (17) is denoted as T_{LB} and the finite upper bound as T_{UB} . The decomposition of the integral (17) with respect to the integration variable T is the next important step. Hence, integral (17) will be rewritten as the sum

$$\begin{aligned} z(\Omega) &\simeq -\frac{1}{2} \int_{T_{LB}}^{T_{UB}} \tilde{\gamma}(T) \operatorname{sech}(\Omega + T) dT \\ &= -\frac{1}{2} \sum_{k=1}^M \left(\int_{T_k - \Delta T/2}^{T_k + \Delta T/2} \tilde{\gamma}(T) \operatorname{sech}(\Omega + T) dT \right). \end{aligned} \quad (20)$$

This decomposition is depicted in Fig. 2.

The limited integration interval is fragmented into equidistant M slices with the width ΔT . The slices are centered at the values

$$\begin{aligned} T_1 &= T_{LB} + \Delta T/2, T_2 = T_1 + \Delta T, T_3 = T_2 + \Delta T, \dots, \\ T_{M-1} &= T_{M-2} + \Delta T, T_M = T_{UB} - \Delta T/2 \end{aligned}$$

with the constant distance $\Delta T = (T_{UB} - T_{LB})/M$. Each slice gives a contribution

$$C_k(\Omega) = \int_{T_k - \Delta T/2}^{T_k + \Delta T/2} \tilde{\gamma}(T) \operatorname{sech}(\Omega + T) dT$$

to the complete system dynamic $z(\Omega)$. In order to achieve a finite model order of the approximation model, the contribution (21) of each slice has to be replaced by a single first-order process

$$\tilde{C}_k(\Omega) = r_k \operatorname{sech}(\Omega + T_k) \quad (21)$$

with a sharp time constant T_k and a constant resistance r_k . This stipulation is depicted in Fig. 3. For the equivalence of the contribution of a single slice and the contribution of a sharp first-order process, the equation

$$\underbrace{\int_{T_k - \Delta T/2}^{T_k + \Delta T/2} \tilde{\gamma}(T) \operatorname{sech}(\Omega + T) dT}_{C_k(\Omega)} = \underbrace{r_k \operatorname{sech}(\Omega + T_k)}_{\tilde{C}_k(\Omega)}$$

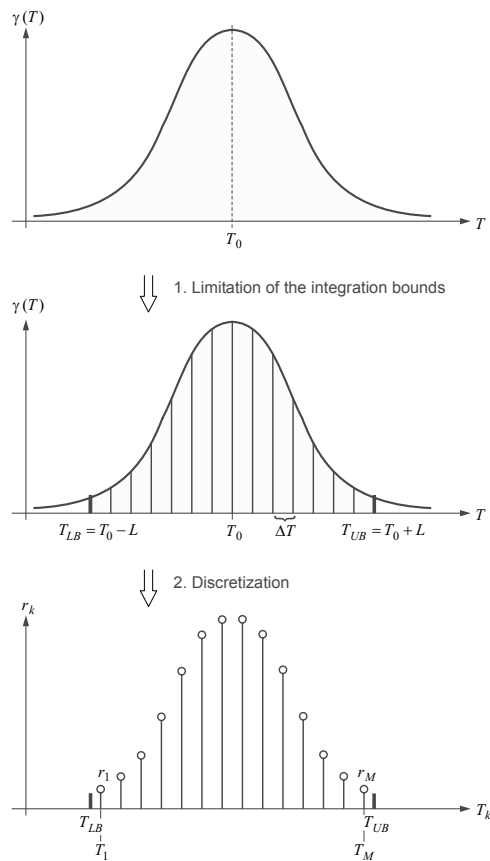


Fig. 2. Schematic depiction of the direct approximation method.

has to be met, which leads to the equation

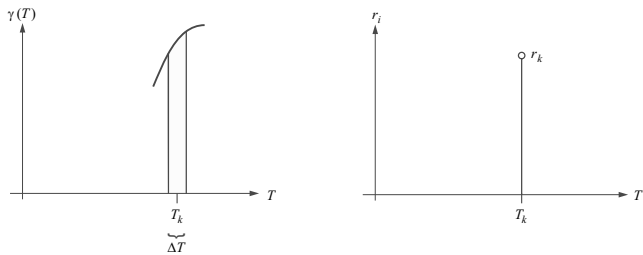


Fig. 3. Schematic depiction of the discretization step.

$$r_k(\Omega) = \frac{1}{\text{sech}(\Omega + T_k)} \int_{T_k - \Delta T/2}^{T_k + \Delta T/2} \tilde{\gamma}(T) \text{sech}(\Omega + T) dT, \quad (22)$$

for a frequency dependent value for r_k . The frequency dependency of $r_k(\Omega)$ in (22) contradicts Eq. (21), where r_k was assumed to be a constant resistance. This problem can be solved by a replacement of the T -dependent term $\text{sech}(\Omega + T)$ by the T -independent term $\text{sech}(\Omega + T_k)$. This simplification can be made, because ΔT is usually very small. Thus, integral (22) can be simplified in order to yield a constant coefficient

$$r_k = \int_{T_k - \Delta T/2}^{T_k + \Delta T/2} \tilde{\gamma}(T) dT. \quad (23)$$

Hence, the transfer function of the integer-order direct approximation system is given by

$$\hat{Z}(j\omega) = \sum_{k=1}^M \frac{r_k}{1 + j\omega\tau_k}, \quad M \gg N. \quad (24)$$

with its model parameters

$$r_k = \underbrace{\frac{\sum_{\kappa=1}^N R_{\kappa}}{\sum_{k=1}^M \tilde{r}_k}}_K \underbrace{\int_{T_k - \Delta T/2}^{T_k + \Delta T/2} \tilde{\gamma}(T) dT}_{\tilde{r}_k}, \quad (25)$$

$$\tau_k = \frac{1}{\omega_0} e^{T_k}. \quad (26)$$

The normalization factor K is necessary to achieve identical steady-state behavior for the actual and for the approximation system. This significant property is lost during the simplifications. In Fig. 2, a schematic depiction of the complete direct approximation method is given. Only the two steps of the direct approximation (limitation and discretization) are causing approximation errors. These two errors can be decreased by a selection of a higher value for L and a smaller width ΔT for the slices. Both measures are leading to a higher model order $M = 2L/\Delta T$.

In the following, three different methods are presented to determine the parameters r_k . These methods are:

1. A direct integration of the distribution density $\tilde{\gamma}(T)$
2. Application of the Simpson-integration rule
3. A simple rectangular integration

The integral in Eq. (25) can be evaluated analytically, if the impedance is represented by a serial connection of Cole–Cole systems. In this case, a primitive exists for the integrand

$$\tilde{\gamma}(T) = \sum_{\kappa=0}^N \tilde{\gamma}_{CC_{\kappa}}(T). \quad (27)$$

A formula for the determination of the coefficients for the approximation of a single Cole–Cole system is given by

$$r_{CC,k} = R \sin((\alpha - 1)\pi) e^{\alpha T_0} \cdot \left(\frac{\arctan\left(\frac{e^{\alpha T_k - \alpha \Delta T/2} + \cos(\pi\alpha) e^{\alpha T_0}}{\sqrt{e^{2\alpha T_0} (1 - \cos^2(\pi\alpha))}}\right)}{\pi\alpha \sqrt{e^{2\alpha T_0} (1 - \cos^2(\pi\alpha))}} - \frac{\arctan\left(\frac{e^{\alpha T_k + \alpha \Delta T/2} + \cos(\pi\alpha) e^{\alpha T_0}}{\sqrt{e^{2\alpha T_0} (1 - \cos^2(\pi\alpha))}}\right)}{\pi\alpha \sqrt{e^{2\alpha T_0} (1 - \cos^2(\pi\alpha))}} \right). \quad (28)$$

If the distribution density (27) is used to evaluate (25), the result for the coefficients of the approximation is given by

$$r_k = \sum_{\kappa=0}^N r_{CC_{\kappa,k}}, \quad k = 1, 2, \dots, M.$$

The parameter $r_{CC_{\kappa,k}}$ denotes the k th parameter of the approximation model of the κ th Cole–Cole system, which contributes to the impedance (2). According to (25), an integral has to be solved, in order to determine the coefficients r_k of the approximation. The exact evaluation of this integral is not necessary, because ΔT is very small in order to justify the simplification (23). An alternative method to determine r_k is to use the Simpson-integration rule. Hence, the parameters r_k can alternatively be determined by

$$r_k = \frac{\Delta T}{6} (\tilde{\gamma}(T_k - \Delta T/2) + 4\tilde{\gamma}(T_k) + \tilde{\gamma}(T_k + \Delta T/2)). \quad (29)$$

Finally, the most simple approach to determine the r_k -parameters is the rectangular integration. This simplification leads to a larger error compared to the Simpson-rule, if it is used to evaluate a finite integral. A rectangular integration for the computation of the model parameters r_k gives $r_k = \Delta T \tilde{\gamma}(T_k)$.

These presented three approximation methods will be compared in the subsequent section.

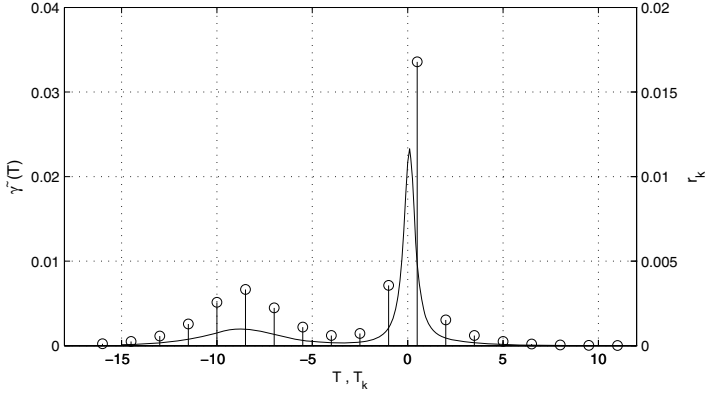


Fig. 4. Distribution density and coefficients of the direct approximation.

4 Simulation of a SOFC-Impedance in the Time Domain

The fractional-order impedance (16) which is a typical one for SOFC is simulated in the time domain using the proposed direct approximation methods presented in this contribution. The $\tilde{\gamma}(T)$ -distribution density of this impedance is depicted in Fig. 1. In order to cover at least $q = 99\%$ of the area between distribution density and T -axis for each Cole–Cole system, the lower bound $T_{LB} = -17.5$ and the upper bound $T_{UB} = 1$ have to be chosen. The resulting model order is $M = 19$, if $\Delta T = 1.5$ is selected. In Fig. 4, the distribution density of the considered fractional impedance model (16) and the $M = 19$ r_k -coefficients are given to depict the relation between the distribution density $\tilde{\gamma}(T)$ and the direct approximation. Figure 5 demonstrates the resulting impedance of the approximation systems compared to the ideal impedance (16). The impedance plots of the approximations using the direct integration and the Simpson rule are very close to each other. Hence, the impedance of the approximation by direct integration according to formula (28) is not given in Fig. 5 to improve the legibility. The approximation of the impedance using the rectangular method is closer to the ideal impedance than the other two approximations. A simulation of the temporal course of the voltage for a given current course $i(t)$ created by a pulse generator, is shown in Figs. 6 and 7. This simulation result demonstrates the capacity of the approximation methods for time domain analysis. The output signal of the systems was computed by MATLAB using a numerical solver for ordinary differential equations. The exact course of the output signal of system (16) is not known. Thus, a reference solution is determined by using the approximation method for fractional systems developed by Oustaloup et al. in [4].

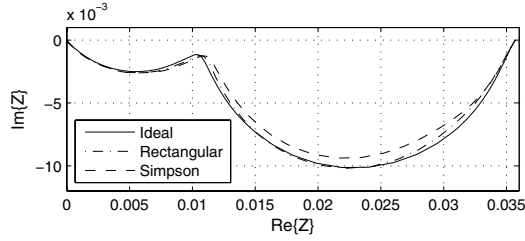


Fig. 5. Impedances of the ideal system and of the approximations.

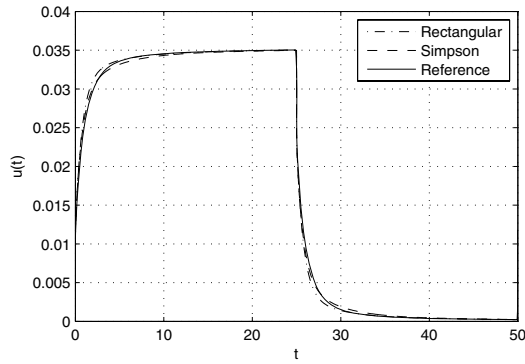


Fig. 6. Time domain simulation of the fractional SOFC-impedance.

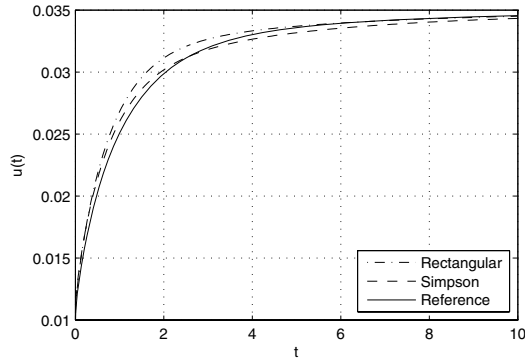


Fig. 7. Time domain simulation of the fractional SOFC-impedance.

For the numerical evaluation of the reference signal with the Oustaloup method, a very high system order of $M_{reference} = 39$ was used for the determination of the reference signal depicted in Figs. 6 and 7.

The plotted signals are very similar to each other, although the system order of $M = 19$ of the considered direct approximations is relatively low for

a representation of a dynamic system with time constants distributed over ten decades. It is not possible to appoint one approximation method as the best one, because Fig. 7 shows that the accuracy of the approximations changes with respect of the simulation time. An increase of M improves the result significantly and makes numerical simulations of a high accuracy possible.

5 Conclusion and Outlook

A new direct approximation method for Cole–Cole systems was presented in this work. This new method is based on the distribution density of relaxation times and takes advantage of its bell-shape. Hence, a limitation on a finite interval of relaxation times is possible. The discretization of the density leads to three different methods for the determination of the approximation model. Even the least complex method yields good simulation results. With the proposed direct approximation it is possible to represent impedances of galvanic elements in the time domain. As an example, the transient behavior of the impedance of a SOFC was simulated in this article. The fast and efficient determination of the approximation model can be used for the identification of impedances of SOFCs using current-interrupt measurements. Other applications like fault diagnosis of electrochemical devices are also possible. A reduction of the model order is possible if a lower approximation accuracy is acceptable for diagnosis or modeling.

Acknowledgment

Thanks go to the American Society of Mechanical Engineers (ASME) for the permission to publish this revised contribution of an ASME article.

References

1. Cole KS, Cole RH (1941) Dispersion and absorption in dielectrics – I. Alternating current characteristics, *J. Chem. Phys.*, 9(April):341–351.
2. Macdonald J (1987) *Impedance Spectroscopy*. Wiley, New York.
3. Haschka M, Rüger B, Krebs V (2004) Identification of the electrical behavior of a solid oxide fuel cell in the time-domain. *Proceedings of Fractional Differentiation and its Applications 2004*, Bordeaux, France., pp. 327–333.
4. Oustaloup A, Levron F, Nanot FM (2000) Frequency-band complex noninteger differentiator: characterization and synthesis. *IEEE Trans Circuits Syst.*, 47(January):25–39.
5. Fuoss RM, Kirkwood JG (1941) Electrical properties of solids. VIII. Dipole moments in polyvinyl chloride-diphenyl systems. *J. Am. Chem. Soc.*, 63(April):385–394.

FRACTIONAL MULTIMODELS OF THE GASTROCNEMIUS MUSCLE FOR TETANUS PATTERN

Laurent Sommacal¹, Pierre Melchior¹, Jean-Marie Cabelguen², Alain Oustaloup¹, and Auke Ijspeert³

¹ LAPS - UMR 5131 CNRS, Université Bordeaux 1 - ENSEIRB, 351 cours de la Libération, F33405 TALENCE Cedex, France; Tél: +33 (0)5 40 00 66 07, Fax: +33 (0)5 40 00 66 44, E-mail: surname.lastname@laps.u-bordeaux1.fr, URL: <http://www.laps.u-bordeaux1.fr>

² INSERM E 0358, Institut Magendie, 1 rue Camille St Saëns, F33077 BORDEAUX Cedex, France; Tel: +33 (0)557 574 052, E-mail: Jean-marie.cabelguen@bordeaux.inserm.fr

³ EPFL, Swiss Federal Institute of Technology, School of Computer and Communication Sciences IC-ISIM-LSL INN 241, CH-1015, LAUSANNE, Switzerland; Tel: (+41) 216 932 658, E-mail: auke.ijspeert@epfl.ch

Abstract

This study talks about gastrocnemius muscle identification. During biological activation, every contractile structure is unsynchronized. Likewise, contraction and relaxation phases depend on all contractile elements, the activation type and the state of health. Moreover, gastrocnemius muscle is composed of three fibre types: fast (IIB), resistant (IIA), and slow (I) fibres.

Some recent works highlight a fractal structure of the muscle, which consolidate the approach based on the use of a noninteger (or fractal) model to characterize its dynamic behavior. A fractional structure model, due to its infinite dimension nature, is particularly adapted to model complex systems with few parameters and to obtain a real-time exploitable model.

According to its complexity, muscle structure and activation mechanisms, and to these previous considerations, an identification based on fractional model is presented.

A model is proposed for the tetanus pattern response in a high tiredness state. It is based on a multimodel structure, which corresponds to the decomposition in contraction and relaxation phases. This multimodel

structure is expected to be included subsequently in agonist–antagonist structure.

Keywords

System identification, modeling, multi-models, fractal systems, biomedical muscle.

1 Introduction

The importance of the striated muscles contraction in animal organism physiology is considerable. Indeed, these muscles are involved during partial or global organism moves (locomotion). The knowledge of the effective contribution of muscular contractions to locomotor activity, makes possible to associate cinematic changes to physiological (tiredness) or pathological origin modifications of the muscle fibre properties (myopathies).

The rhythmic and stereotyped nature of locomotor moves, as well as the conservation of several of their cinematic characteristics within vertebrate, including man, justifies studies that have been devoted to the striated muscles behavior during locomotion. Electromyographic (EMG) recordings carried out on various species, permit to specify the muscular activations sequence during the locomotor cycle [13].

The contribution of muscular activations recorded for the observed locomotor movements remains nevertheless very badly known, in particular, due to the lack of data, concerning both the mechanical properties of muscles working under dynamic conditions and those coming from the locomotor apparatus (articulations, skin).

Thus, this paper aims for determining the muscular activation contribution to the locomotor movements for an inferior vertebrate (urodele) by means of a mathematical modeling of the muscular contraction in dynamic conditions. It is based on EMG and kinematic data obtained during a former study [3, 4].

The striated muscle structure is widely described in biology. Three fibre types [15] make up this muscle type. Myofilaments groups are the constituents of each fibre, which can be reduced to actin and myosin filaments. The fibre types present different characteristics, like contraction and relaxation delays and feeding types.

Some others recent works [1, 2, 5, 8, 12] present a fractal structure of the muscle, which consolidate the approach based on the use of noninteger (or fractal) model to characterize its dynamic behavior. A fractional structure

model, due to its infinite dimension nature, is particularly adapted to model complex systems with few parameters, and to obtain real-time exploitable model: muscle complexity can be described thanks to fractional models, with a model containing a reduced number of parameters.

After this introduction, section 2 explains the fractional differentiation. The experimental protocol is detailed in section 3. Then, section 4 presents the muscular response to tetanus pattern identification. Finally, a conclusion is given in section 5.

2 Fractional Differentiation

The fractional derivative of the function $f(t)$ at ν order is defined as [6],

$$D^\nu f(t) \approx \frac{1}{h^\nu} \sum_{k=0}^{\infty} (-1)^k a_k(\nu) f(t - kh), \quad (1)$$

with $t = Kh, K \in \mathbb{N}^+$ and h is the sampling period.

Assuming that $f(t) = 0 \forall t < 0$, the D^ν Laplace transform is [7, 10, 11, 14],

$$\mathcal{L}(D^\nu f(t)) = s^\nu \cdot \mathcal{L}(f(t)), \quad (2)$$

where ν can be real or imaginary number order.

Linear model described with the fractional differential equation,

$$\sum_{l=1}^L a_l \frac{d^{n_l}}{dt^{n_l}} \hat{y}(t) = \sum_{q=1}^Q b_q \frac{d^{m_q}}{dt^{m_q}} u(t), \quad (3)$$

where

$$n_1, \dots, n_L, m_1, \dots, m_Q \in \mathbb{R}^{L+Q}, \quad (4)$$

can be modeled as the following fractional transfer function, providing that the system is relaxed at $t = 0$,

$$\hat{y}(s) = \frac{b_1 s^{m_1} + b_2 s^{m_2} + \dots + b_Q s^{m_Q}}{a_1 s^{n_1} + a_2 s^{n_2} + \dots + a_L s^{n_L}} u(s). \quad (5)$$

The output model can then be simulated:

$$\hat{y}(Kh) = \frac{1}{\sum_{l=1}^L \frac{a_l}{h^{n_l}} \binom{n_l}{0}} \left(\sum_{k=0}^K \sum_{q=1}^Q (-1)^k \frac{b_q}{h^{n_l}} \binom{n_q}{k} u((K-k)h) - \sum_{k=1}^K \sum_{l=1}^L (-1)^k \frac{a_l}{h^{n_l}} \binom{n_l}{k} \hat{y}((K-k)h) \right). \quad (6)$$

Generally system identification of fractional model (linear or nonlinear) can be carried by minimizing the quadratic output error criterion J ,

$$J = \sum_{k=1}^N (\varepsilon(kh))^2, \quad (7)$$

where

$$\varepsilon(kh) = y^*(kh) - \hat{y}(kh) \quad (8)$$

is the output error.

A nonlinear optimization algorithm can then be used, for example, Newton, Marquard, Simplex... Thus the nonlinear Simplex [16, 17, 18] is chosen to be included in the output error model (Fig. 1).

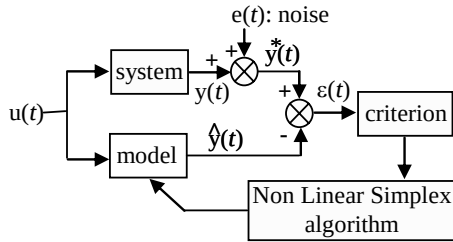


Fig. 1. Output error model.

3 Experimental Protocol and Models to Isolated Pulses

3.1 Experimental protocol

Experiments about this project are performed on *Rana esculenta* frog muscles, and more particularly on the gastrocnemius one. The structure and the composition of this muscle are identical to those of salamanders, with which next experiments will be conducted. The frog muscle advantage is a bigger length, easier to experiment. The aim is to model muscle behavior thanks to tetanus, swimming, and walking salamander cycles.

Frogs are demedulated and decerebrated to eliminate arc-back. Legs skin is removed. Exposed muscles and skin are covered with some paper moisted with Ringer solution. Control electrodes are applied on sciatic nerve, between the pelvis and the gastrocnemius. The tendon between the gastrocnemius and the ankle is severed and fixed on an isometric sensor (Phymep UF-1). Experimentally, the muscle length is the *in situ* length measured in the median posture. Pulses are sent on the muscle thanks to the sciatic nerve. The pulse

time width is 1 ms. The amplitude is fixed, during a test phase beginning every experiment, to obtain maximum amplitude. In this way, each fibre is excited. The isometric sensor allows us to translate the muscular tension into a potential variation, which is displayed on an oscilloscope. Electric signals (pulses for the stimulation and the contraction responses) are recorded thanks to an A/N converter (CED 1401) and the software Spike 2. Data are received as text files. The sampling frequency for excitation and isometric responses is 10 kHz. Data text files will be used for the muscle identification. The schematic protocol is presented on Fig. 2.

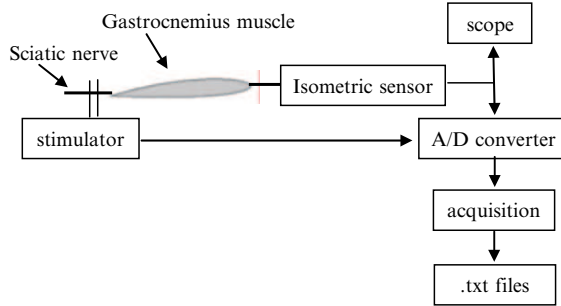


Fig. 2. Schematic protocol.

Eight pattern sequences are applied on muscle to obtain a predictable and representative behavior. For each pattern, different regularly spaced pulses sequences are applied. Each response is studied in order to identify the muscular behavior.

3.2 Single pulse model

An integer model is obtained thanks to parametric estimation so as to study the muscle behavior. A minimum amount of 11 parameters is adopted to have a good modeling:

$$\begin{aligned}
 & \hat{y}(t) + 1.2 \cdot 10^{-1} D^1 \hat{y}(t) + 4.1 \cdot 10^{-3} D^2 \hat{y}(t) + 6.7 \cdot 10^{-5} D^3 \hat{y}(t) \\
 & + 5.7 \cdot 10^{-7} D^4 \hat{y}(t) + 2.8 \cdot 10^{-9} D^5 \hat{y}(t) - 6.9 \cdot 10^{-11} D^6 \hat{y}(t) \\
 & = 1.4 \cdot 10^{-3} u(t) + 3.1 \cdot 10^{-2} D^1 u(t) + 2.1 \cdot 10^{-3} D^2 u(t) \\
 & - 2.6 \cdot 10^{-6} D^3 u(t) + 8.8 \cdot 10^{-8} D^4 u(t).
 \end{aligned} \tag{9}$$

with $u(t)$, the signal applied on sciatic nerve and $\hat{y}(t)$, the muscular strength response.

To test the ability of the noninteger tool to report this behavior, a same model structure based on fractional orders is defined. As we have no particular

knowledge about the derivative orders, the half integer order step is chosen (0.5), which is the extreme noninteger case.

$$\begin{aligned}
 & \hat{y}(t) - 2.6 \cdot 10^{-1} D^{0.5} \hat{y}(t) + 9.0 \cdot 10^{-2} D^1 \hat{y}(t) - 1.4 \cdot 10^{-2} D^{1.5} \hat{y}(t) \\
 & + 1.9 \cdot 10^{-3} D^2 \hat{y}(t) - 1.3 \cdot 10^{-4} D^{2.5} \hat{y}(t) + 6.8 \cdot 10^{-6} D^3 \hat{y}(t) \\
 & = -1.5 \cdot 10^{-2} u(t) + 2.6 \cdot 10^{-2} D^{0.5} u(t) + 1.4 \cdot 10^{-2} D^1 u(t) \\
 & - 1.8 \cdot 10^{-3} D^{1.5} u(t) + 2.6 \cdot 10^{-4} D^2 u(t).
 \end{aligned} \tag{10}$$

represents one of 74 muscular responses, the integer and fractional model responses.

In both cases, the structure models, obtained using a parametric estimation, allow to obtain a muscular response with a very small error, in either of integer or fractional cases (Fig. 3).

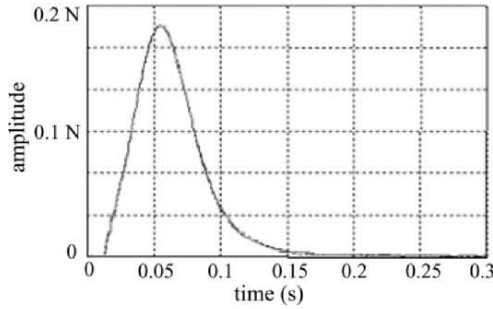


Fig. 3. Muscular response to a pulse (solid), integer (dotted), and fractional models (dashed).

3.3 A model for the amplitude response

For isolated pulses, muscular responses show an amplitude variation. It appears also few dynamic variations. To have a global modeling to a pulse pattern, a mean response model is determined after the normalization of each response. Amplitude variations are described thanks to a second model, function of the delay between two successive pulses,

Figure 4 shows the normalized mean response with the fractional model response. The maximum and minimum muscular responses are also shown in this figure. The identification is applied on this normalized mean response.

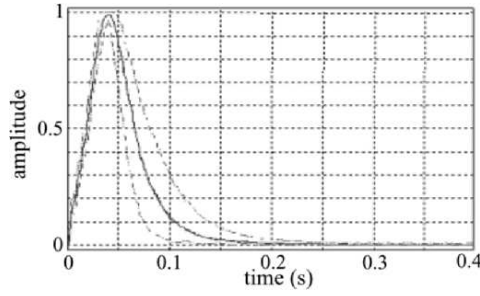


Fig. 4. Normalized mean muscular response (solid), fractional model response (dotted), and minimum/maximum experiment responses (dashed).

The obtained model is the following one, with 11 fractional parameters:

$$\begin{aligned}
 & \hat{y}(t) - 1.8 \cdot 10^{-1} D^{0.5} \hat{y}(t) + 7.0 \cdot 10^{-2} D^1 \hat{y}(t) - 9.1 \cdot 10^{-3} D^{1.5} \hat{y}(t) \\
 & + 1.4 \cdot 10^{-3} D^2 \hat{y}(t) - 8.8 \cdot 10^{-5} D^{2.5} \hat{y}(t) + 5.0 \cdot 10^{-6} D^3 \hat{y}(t) \\
 & = -1.3 \cdot 10^{-2} u(t) + 2.0 \cdot 10^{-2} D^{0.5} u(t) + 4.2 \cdot 10^{-2} D^1 u(t) \\
 & - 5.0 \cdot 10^{-3} D^{1.5} u(t) + 2.7 \cdot 10^{-4} D^2 u(t).
 \end{aligned} \tag{11}$$

The behavior model is computed from the average of normalized responses, depending on time elapsed between two successive pulses.

Figure 5 shows amplitude development for eight experiment sequences. When the pattern is reduced to sufficiently near pulses, the maximum muscular amplitude can be reduced to a decreasing exponential function and can be modeled by:

$$A(t) = (1 - K) e^{-t/\tau} + K, \quad \text{with } K \approx 0.7 \text{ and } \tau \approx 2.5 \text{ s.} \tag{12}$$

So, (11) associated with (12), allows to get a global model to closely spaced pulses, where (11) reports dynamic and (12) the amplitude variation.

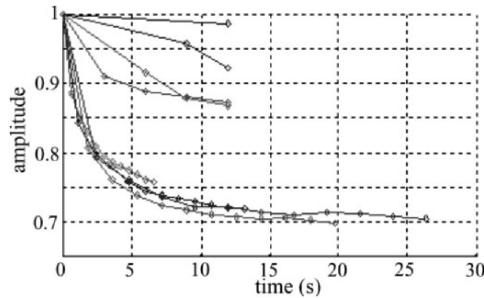


Fig. 5. Amplitude development depends on the delay between two successive pulses.

4 Muscular Response to the Tetanus Pattern Identification

4.1 Tetanus pattern determination

In a second stage, the obtained model will be used with swimming and walking patterns, to verify the model validity. EMG data obtained on *in vivo* salamander during swimming cycle, as seen on Fig. 6, do not seem to be reproducible. The patterns resulting from the application of a threshold on these biological data, a second time, prove these patterns are not yet reproducible.

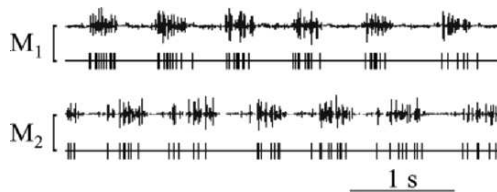


Fig. 6. Activation of two antagonist myomeres (M_1 and M_2) during swimming cycles for the pleurodele. For each myomere, the plot above represents the electromyographic (EMG) activity and the plot below is the pulse train, after EMG threshold.

The tetanus pattern is easier to build, because it makes possible the creation of an artificial pattern which allows to have the minimum rising time and a maximal amplitude response. Firstly, the tetanus pattern is studied and secondly, the obtained model is going to be ratified by means of its pattern (11).

To create this pattern, which implicates the muscle tetanus, the pulses arrangement acts as a substitute for EMG signals propagated on each nerve axons during *in vivo* conditions. The main advantage is this pattern is totally reproducible, since every fibre is activated by pattern. This response pattern results from an iterative pulse addition. The time fixed between two pulses corresponds to the biggest response amplitude and the smallest twitch time. The delay between each exercise is sufficient to let the muscle in the same tiredness state, 15 s minimum.

Until 6 pulses, the optimal response is obtained with the following pattern: 24, 21, 18, 15, 12 ms. Nevertheless, EMG data show the global signal is richer at the beginning than at the end. Thus the next stage consists to study the muscle response to the inversed pattern. Both amplitudes are quite equal but the peak response time is smaller with the second pattern.

Due to this result, a second study has allowed to create the following global pattern, for which, the twitch time is minimized with a maximum amplitude: 12, 12, 16, 16, 16, 16, 18, 18, 18, 18, ... ms.

This pattern, defined by these time gaps, is adopted as the tetanus pattern.

4.2 Identification of the muscular response from the tetanus pattern

A typical tetanus pattern response is presented in the Fig. 7. The time domain identification of this response has been done.

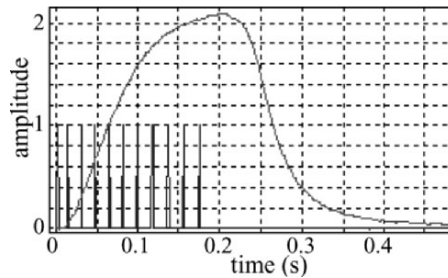


Fig. 7. One typical muscular response obtained with the tetanus pattern.

In a first step, the tetanus pattern is applied to the models (11) and (12) corresponding to an isolated pulse response, compared to the muscular response obtained with the tetanus pattern (Fig. 8).

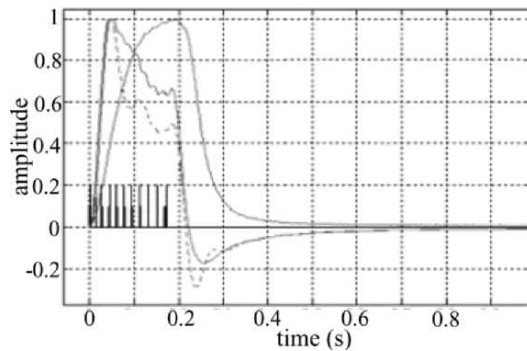


Fig. 8. Comparison of the integer (light, solid), fractional (dotted) models responses corresponding to an isolated pulse response with the muscular responses (dark, solid), to the tetanus pattern (black).

The difference between the experimental response and the models response proves these models are not sufficient to explain all tetanus response.

So, in a second step, another approach consists to try to identify the global muscular response obtained with the tetanus pattern. The fractional model size is 15 parameters. The difference between the experimental response and the model response proves that a global model is not sufficient to explain all tetanus response anew (Fig. 9).

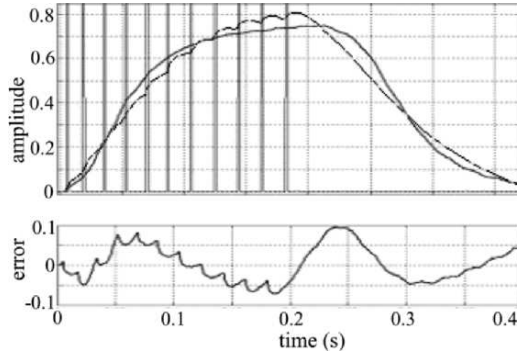


Fig. 9. Comparison of the global model response (light) with the muscular response (black) to the tetanus pattern and the error.

Thus, to approach the biological phenomenon of contraction and relaxation phases, in a third step, a new approach is defined: a multi-model structure [7] has been tested. The contraction and relaxation phases are treated separately (Fig. 10), due to the fact biological literature mentions a physical mechanisms difference [15]. This induces a nonlinear behavior, which can be first approximated with a linear multimodels. This response corresponds to a tired muscle response. Because the muscle and nerve frequency bands are low, the signal input cannot be rich enough in frequencies to allow a reliable frequency-domain system identification. The twitch model (Fig. 11a) is designed from the contraction phase of response, which corresponds to the eleven first pulses of tetanus pattern. Figure 11b presents the relaxation phase. The optimization algorithm is based on a fractional-order model, which includes by nature integer orders. Only the structure size is fixed, orders and coefficients are free. The algorithm optimizes both coefficients and orders, and converges toward fractional model with 6 parameters:

Contraction phase:

$$G_C(s) = \frac{1}{s} \left(\frac{44.27}{s^{1.23} + 40.66} \right) + \frac{1}{s} \left(\frac{-51.13}{s^{1.38} + 52.98} \right), \quad (13)$$

Relaxation phase:

$$G_R(s) = \left(\frac{192.66859}{s^{0.86} + 6.2722972} \right) + \left(\frac{1766.43}{s^{1.13} + 62.50} \right). \quad (14)$$

Since it is a global identification, these parameters (coefficients and orders) have no physiological meaning.

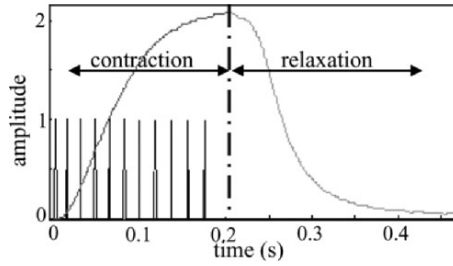


Fig. 10. Contraction and relaxation response phase for tetanus pattern.

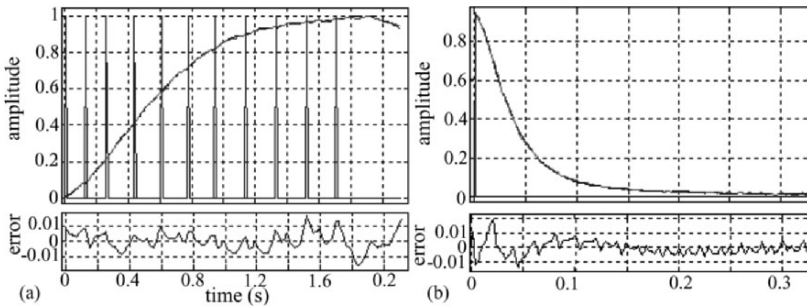


Fig. 11. (a) Twitch phase of muscular response (solid) and twitch fractional model (black, dotted) for tetanus pattern and error, (b) relaxation phase of muscular response (solid) and relaxation fractional model (black, dashed) and error.

The validation phase has been tested respectively with the sixth, seventh, and eleventh pulses of tetanus pattern. Figure 12 presents the comparison of the contraction/relaxation phase model responses with the muscular response, respectively, to the 6 (a)–(d), 7 (b)–(e), and 11 (c)–(f) first pulses of tetanus pattern. In the different cases, the obtained model allows a good reconstruction of the contraction phase. For relaxation phase identification, the obtained model allows also a good reconstruction, but the error is worse than contraction phase identification one. Moreover, the error is not centered on zero, which is significant of a junction choice error.

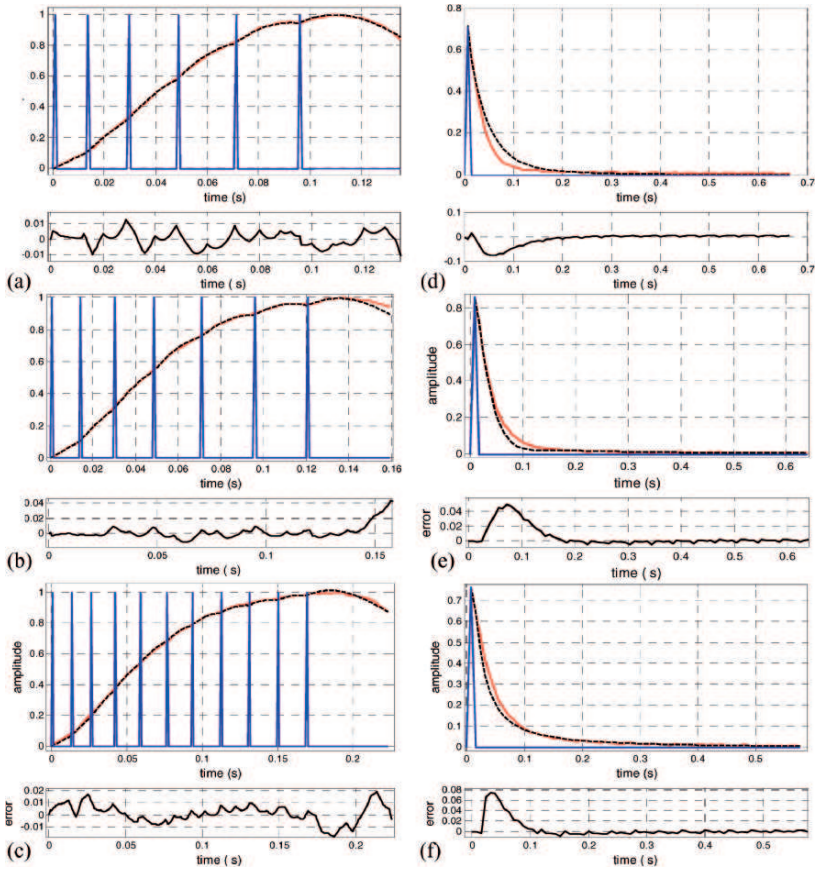


Fig. 12. Validation phase: comparison of the contraction/relaxation phase model responses (black, dotted) with the muscular response (solid) respectively to the 6 (a)–(d), 7 (b)–(e), and 11 (c)–(f) first pulses of tetanus pattern and errors.

5 Conclusion

This paper introduces the gastrocnemius muscle identification by fractional model.

A first model is proposed for the muscular response to a single pulse. The muscular response from the tetanus pattern is also studied. An artificial global tetanus pattern has been determined. A fractional model of the gastrocnemius

muscle response to tetanus pattern is proposed. It is based on a multimodel structure, which corresponds to the decomposition in contraction and relaxation phases, according to physical phenomenon described in biological literature. The optimization algorithm considers fractional-order model, which includes by nature integer orders. Only the structure size is fixed, orders and coefficients are let free. The algorithm optimized both coefficients and orders, and converges toward fractional model with 6 parameters. The validation phase shows the comparison of the model responses with the muscular response for different sizes of tetanus pattern. In the different cases, obtained models allow a good reconstruction of the contraction phase, but the existence of a junction choice error in the relaxation phase.

These models are obtained with a particular muscle tiredness state. Also, it is necessary to study the tiredness to obtain a model which can predict all tetanus response.

Acknowledgment

This paper is a modified version of a paper published in proceedings of IDETC/CIE 2005, September 24–28, 2005, Long Beach, California, USA. The authors would like to thank the American Society of Mechanical Engineers (ASME) for allowing them to publish this revised contribution of an ASME article in this book.

References

1. Arsos GA, Dimitriu PP (2004) A fractal characterization of the type II fibre distribution in the extensor digitorum longus and soleus muscles of the adult rat, *Muscle and Nerve*, 18:961–968.
2. Cross SS (1997) Fractals in pathology, *J. Pathol.* 182:1–8.
3. Delvolvé I, Bem T, Cabelguen J-M (1997) Epaxial and limb muscle activity during swimming and terrestrial stepping in the adult newt, *Pleurodeles waltl*, *J. Neurophysiol.*, 78:638–650.

4. Delvolvé I, Branchereau P, Dubuc R, Cabelguen J-M (1999) Fictive patterns of rhythmic motor activity induced by NMDA in an in vitro brainstem-spinal cord preparation from an adult urodele amphibian, *J. Neurophysiol.*, 82:1074–1078.
5. Goldberger AL, Amaral LAN, Hausdorff JM, Ivanov PCh, Peng C-K, Stanley HE (February 2002) Fractal dynamics in physiology: alterations with disease and aging, *PNAS*, 99(Suppl. 1):2466–2472.
6. Grünwald AK (1867) Ueber begrenzte Derivationen und deren Anwendung, *Z. Angew. Math. Phys.*, 12:441–480.
7. Liouville J (1832) Mémoire sur le calcul des différentielles à indices quelconques, *Ecole Polytechnique*, 13(21):71–162.
8. Lowen SB, Cash SS, Poo M-M, Teich MC (August 1997) Quantal neurotransmitter secretion rate exhibits fractal behavior, *J. Neurosci.*, 17(15):5666–5677.
9. Malti R, Aoun M, Battaglia J-L, Oustaloup A, Madami K (August 1989) Fractional multimodels – Application to heat transfert modelling, 13th IFAC Symposium on System Identification, Rotterdam, The Netherlands.
10. Miller KS, Ross B (1993) *An Introduction to the Fractional Calculus and Fractional Differential Equation*. Wiley, New York.
11. Oustaloup A (1995) *La Derivation Non Entière*, Hermes, Paris.
12. Ravier P, Buttelli O, Couratier P (2005) An EMG fractal indicator having different sensitivities to changes in force and muscle fatigue during voluntary static muscle contractions, *J. Electromyogr. Kinesiol.*, 15(2):210–221.
13. Rossignol S (1996) Neural control of stereotypic limb movements, *International Handbook of Physiology* (Rowell LB, Sheperd JT ed.). American Physiological Society, pp. 173–216.
14. Samko AG, Kilbas AA, Marichev OI (1987) *Fractional Integrals and Derivatives*. Gordon and Breach Science, Minsk.
15. Shepherd GM (1994) *Neurobiology*. Oxford, New york.
16. Sommacal L, Melchior P, Cabelguen J-M, Oustaloup A, et Ijspeert A (2005) Fractional model of a gastrocnemius muscle for tetanus pattern, Fifth ASME International Conference on Multibody Systems, Nonlinear Dynamics and Control, Long Beach, California, USA.

17. Subrahmanyam MB (August 1989) An extension of the simplex method to constrained nonlinear optimization, *Int. J. Optim. Theory Appl.*, 62(2):311–319.
18. Woods DJ (May 1985) An interactive approach for solving multi-objective optimization problems, Technical Report 85–5, Rice University, Houston.

LIMITED-BANDWIDTH FRACTIONAL DIFFERENTIATOR: SYNTHESIS AND APPLICATION IN VIBRATION ISOLATION

Pascal Serrier, Xavier Moreau, and Alain Oustaloup

LAPS-UMR 5131 CNRS, Université Bordeaux1 – ENSEIRB, 351 cours de la Libération, 33405 TALENCE Cedex, France; E-mail: {pascal.serrier, xavier.moreau, alain.oustaloup}@laps.u-bordeaux1.fr

Abstract

The use of fractional differentiation in vehicle suspension design has many interests. This paper presents a hydropneumatic suspensions design method based on fractional differentiation. Once a hydraulic structure has been chosen for the suspension, it is possible to calculate the values of all technological parameters, so that the suspension force – deflection transfer function is a band-limited fractional differentiator. The combination of the CRONE control methodology and hydropneumatic technology leads to remarkable performances of robustness.

Keywords

Fractional differentiation, hydropneumatic technology, CRONE control, hydropneumatic test bench.

1 Introduction

For a long time, fractional derivative was not used because of the lack of physical signification of this concept and the lack of means to synthesize and achieve fractional differentiators. Nowadays, fractional derivative is used in numerous applications, such as heat transfer phenomena [1], dielectric polarization [2], and vibration isolation [3].

This paper presents, in vibration isolation context, two structures of hydropneumatic components allowing to achieve a suspension whose force-deflection transfer function is a given limited-bandwidth fractional differentiator.

In a more general context, two means to achieve limited-bandwidth fractional differentiator in hydropneumatic technology are proposed.

The first part reminds the formulation of vibration isolation problem as robust control synthesis problems and the interest of fractional derivative in vibration isolation.

Through the example of a hydraulic test bench, the second part presents the synthesis of a limited-bandwidth fractional differentiator to satisfy given specifications in a vibration isolation context. Two structures are proposed to achieve a suspension in hydropneumatic technology. A method to determine the corresponding technological parameters is described in each case.

The third part shows the simulated performances obtain with the test bench, taking into account the influence of sprung mass variation on the hydropneumatic components.

The last part concerns the future development of this work.

2 Fractional Derivative in Vibration Isolation

2.1 From vibration isolation to robust control synthesis

Vibration isolation is a usual mechanical problem which consists in limiting vibration transmission between a source and one or some systems. A solution is to isolate the system from the source by using a vibration isolator, also called suspension. This problem can be formalized as a usual problem of control synthesis in the field of Automatics.

Some previous works [4] have shown, from a one degree of freedom (DOF) model (Fig. 1), that a suspension which develops a force $u(t)$ that is links to its deflection $z_{10}(t)$ by the relation:

$$U(s) = -D(s)Z_{10}(s), \quad (1)$$

naturally makes a feedback control around the static equilibrium position.

The block diagram (Fig. 2) which is issued from the modelling clearly shows that the suspension has the same role that the controller of a control loop, that the displacement and force solicitations ($z_0(t)$, $f_0(t)$) can be considered as input and output perturbations which act on the plant. The plant is a double integrator whose transitional pulsation depends on the sprung mass.

The open-loop transfer function $\beta(s)$ is:

$$\beta(s) = D(s)G(s), \text{ with } G(s) = \frac{1}{M s^2}. \quad (2)$$

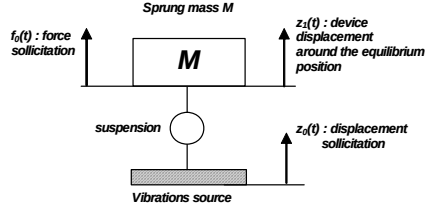


Fig. 1. One degree of freedom (DOF) model.

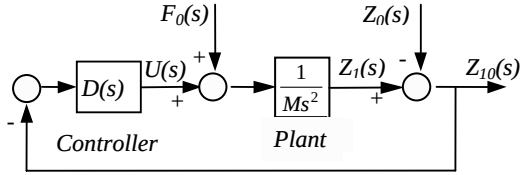


Fig. 2. One degree of freedom model block diagram.

Thus, suspension design can be made by using the classic control synthesis method. In particular, when parameters uncertainties (especially mass variation) are considered, the problem becomes a robust control synthesis problem.

2.2 The CRONE suspension

The CRONE suspension is a suspension whose synthesis is based upon CRONE control one [5]. The CRONE control allows to obtain the stability degree robustness in spite of parameters uncertainties. The second generation CRONE control is used when either parameter uncertainties correspond only to plant gain variations or when parameter and controller uncertainties compensate for themselves and lead to a behaviour like the one that will be develop below. In both cases, the open-loop phase remains the same around the gain crossover pulsation ω_u .

In the particular case of this article, the suspension force-deflection transfer function is a limited-bandwidth fractional differentiator [6], namely:

$$D(s) = D_0 \left(\frac{1 + \frac{s}{\omega_b}}{1 + \frac{s}{\omega_h}} \right)^m, \quad (3)$$

where D_0 is the static gain, ω_b and ω_h are the low and high transitional pulsations and where m is the fractional order, that means that m may be not non-integer.

D_0 , m , ω_b , and ω_h are in the CRONE method the high-level synthesis parameters. These four parameters are determined from the specification sheets.

Bode diagrams of this suspension are given in Fig. 3. The phase is constant between the pulsations ω_A and ω_B .

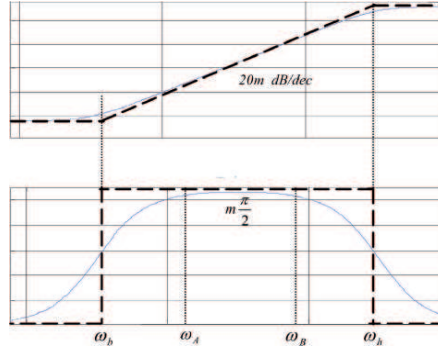


Fig. 3. Bode diagrams of a limited-bandwidth fractional differentiator.

To achieve a real hydropneumatic device whose transfer function is a limited-bandwidth fractional differentiator, or only to simulate a fractional differentiator, it is necessary to synthesize a rational approximation of the fractional differentiator. A method to do this is proposed in [6].

The desired transfer function $D(s)$ can be approximated by N poles and N zeros. This approximation $D_N(s)$ is given by [6]:

$$D_N(s) = D_0 \prod_{i=1}^N \left(\frac{1 + \frac{s}{\omega_i'}}{\frac{\omega_i}{1 + \frac{s}{\omega_i}}} \right), \quad (4)$$

where ω_i' and ω_i are recursively distributed through the recursive coefficients α and η which are defined by [6]:

$$\frac{\omega_{i+1}'}{\omega_i'} = \frac{\omega_{i+1}}{\omega_i} = \alpha\eta > 1, \quad \frac{\omega_i}{\omega_i'} = \alpha > 1 \quad \text{and} \quad \frac{\omega_{i+1}'}{\omega_i} = \eta > 1. \quad (5)$$

In hydropneumatic technology, the CRONE suspension is made of hydraulic accumulator (capacitive components, C, which contain oil and gas, nitrogen, separated by an impermeable diaphragm) and hydraulic dampers (dissipative components, R).

Two structures are considered in the following part: a parallel arrangement of R and C components in series (RC cells) (Fig. 4), and gamma arrangement (Fig. 5).

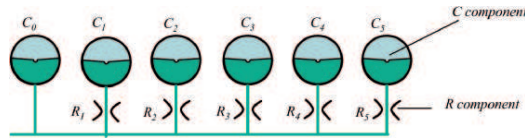


Fig. 4. Parallel arrangement of RC cells.

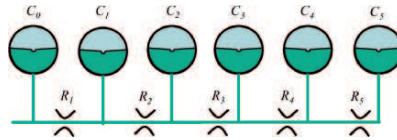


Fig. 5. Gamma arrangement.

3 CRONE Hydropneumatic Test Bench

This part uses a CRONE hydropneumatic test bench as an example to analyse the previous structures.

3.1 Description

The CRONE test bench allows to study the free evolution of a mass (M) after a release test. The mass is mechanically linked to a hydraulic simple effect jack (Fig. 6). The minimal mass of 75 kg can be increased by additional masses. So, M can vary between 75 and 150 kg.

The suspension jack is connected to a two parts hydraulic circuit (Fig. 6). The first part is composed of a pump equipped with a make and brake circuit and a proportional valve. Its aim is to maintain the mass M at a fixed height independently of the mass value thanks to a control feedback. The second is composed of two change over valves which allow to select either a parallel arrangement of two cells including one RC cell ($N = 1$, N is the number of RC cells), or a parallel arrangement of six cells including five RC cell ($N = 5$) or a gamma arrangement of six cells.

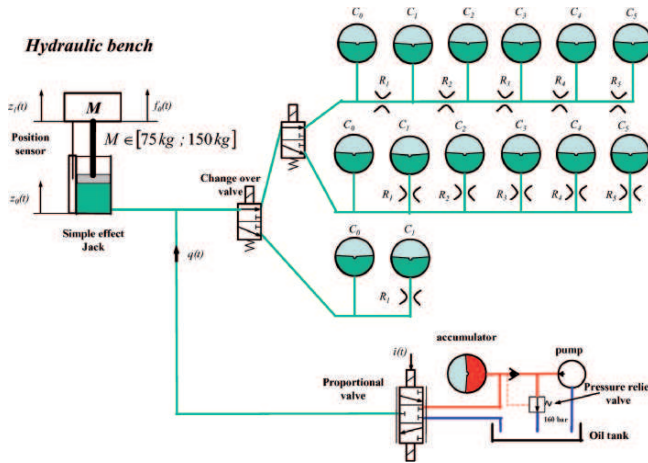


Fig. 6. Hydraulic diagram of the CRONE test bench.

The associated control diagram is presented in Fig. 7. The external loop which regulates the static equilibrium position at a value equal to half stroke of the jack has the same rapidity as the self-leveller device of a hydro-pneumatic suspension. This rapidity is characterized by an open-loop gain crossover frequency of 0.1 rad/s. The internal loop has a rapidity characterized by an open-loop gain crossover frequency of 6 rad/s, the same rapidity as the vertical mode of a usual vehicle. So, both loops are dynamically uncoupled. That is why only the internal loop is considered in the following parts of this article.

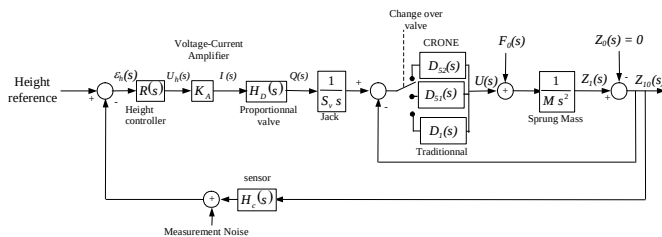


Fig. 7. Control diagram of the CRONE test bench.

3.2 CRONE suspension achievement in hydropneumatic technology

The following steps are necessary to determine the technological parameters of the hydropneumatic components:

From the specification sheets (rapidity, stability...), the suspension desired transfer function $D_N(s)$ is determined, according to the CRONE control methodology [5]. This ideal transfer function is characterized by the four high-level synthesis parameters.

A rational approximation with N poles and N zeros is then established thanks to relation (5).

Two sets of relations (one for each structures: parallel arrangement of RC cells or gamma arrangement), which are established in the two following paragraphs, link the N poles and N zeros to the $2N+1$ physical parameters (N resistances and $N+1$ capacities) of the hydropneumatic suspension.

Two others relations (which are detailed in the next part) allow to obtain the $2(2N+1)$ technological parameters from the physical parameters.

Finally, if the suspension is achieved with five RC cells ($N=5$), twenty-two technological parameters are obtained from the four high-level synthesis parameters.

3.3 Relations between technological parameters and the poles and the zeros recursive distribution

Parallel Arrangement of RC Cells

The input hydraulic impedance of each parallel arrangement is characterized by an expression of the form:

$$\frac{P_e(s)}{Q_e(s)} = \frac{1}{C_0 s + \sum_{i=1}^N \frac{1}{R_i + \frac{1}{C_i s}}}, \quad (6)$$

where $P_e(s)$ and $Q_e(s)$ are the pressure and the flow at the input point, where R_i and C_i are the i th cell resistance and capacity whose expression is obtained by linearizing the hydraulic accumulator pressure–volume characteristic around the equilibrium point. The equilibrium point is defined by the static pressure P_s and the gas volume V_{Si} . The capacity is thus given by [3]:

$$\frac{1}{C_i} = - \left. \frac{\partial P}{\partial V} \right|_{\substack{P=P_s \\ V=V_{Si}}} = \gamma \frac{P_s}{V_{Si}}, \quad (7)$$

γ is the thermodynamic coefficient which characterizes the gas evolution ($\gamma=1$ for an isotherm evolution, $\gamma=1.4$ for an adiabatic evolution). Knowing that the product between the pressure and gas volume is constant ($PV = cst$), the gas volume V_{Si} can be expressed with the i th accumulator calibration pressure P_{0i} and the volume V_{0i} (gas initial volume), namely:

$$V_{si} = \frac{P_{0i}}{P_s} V_{0i}, \quad (8)$$

so the capacity C_i expression:

$$C_i = \frac{P_{0i} V_{0i}}{\gamma P_s^2}. \quad (9)$$

The hydraulic resistor R_i is dimensioned so that the flow is laminar [7]:

$$R_i = \frac{128}{\pi} \mu \frac{l_{Ri}}{d_{Ri}^4}, \quad (10)$$

where μ is the oil dynamic viscosity, l_{Ri} and d_{Ri} are respectively the hydraulic resistor length and diameter (Fig. 8).

Expressions (9) and (10) define the relations between physical hydro-pneumatic parameters and technological parameters.

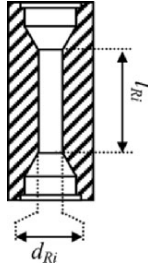


Fig. 8. Hydraulic resistance.

More over, if the pressure drop due to the valve is considered negligible, the arrangement input pressure $P_e(s)$ is equal to the jack pressure. That is why the pressure $P_e(s)$ is linked to the force $U(s)$, which is applied on the mass M by the jack, by the relation:

$$P_e(s) = \frac{U(s)}{S_v}, \quad (11)$$

where S_v is the jack section.

When the self-leveller device is not in action, the arrangement input flow $Q_e(s)$ depends on the jack displacement $Z_{10}(s)$:

$$Q_e(s) = S_v s Z_{10}(s). \quad (12)$$

The introduction of relation (11) and (12) in Eq. (6) leads to the force-deflection transfer function expression $D_N(s)$, namely:

$$D_N(s) = -\frac{U(s)}{Z_{10}(s)} = S_v^2 s \frac{P_e(s)}{Q_e(s)} = \frac{S_v^2 s}{C_0 s + \sum_{i=1}^N \frac{1}{R_i + \frac{1}{C_i s}}}. \quad (13)$$

By dividing the denominator by $S_v^2 s$, the expression (13) becomes:

$$D_N(s) = \frac{1}{\frac{C_0}{S_v^2} + \sum_{i=1}^N \frac{1}{R_i S_v^2 s + \frac{S_v^2}{C_i}}}, \quad (14)$$

expression of the form:

$$D_N(s) = \frac{1}{\frac{1}{k_0} + \sum_{i=1}^N \frac{1}{b_i s + k_i}}, \quad (15)$$

by introducing:

$$k_0 = \frac{S_v^2}{C_0}, \quad k_i = \frac{S_v^2}{C_i} \quad \text{and} \quad b_i = S_v^2 R_i, \quad (16)$$

where k_0 and k_i are homogeneous to stiffnesses, which are expressed in Newton per meter, (N/m), and where b_i are homogeneous to viscous damping coefficients which are expressed in Newton second per meter (Ns/m). The hydraulic arrangement equivalent mechanical diagram is given in Fig. 9.

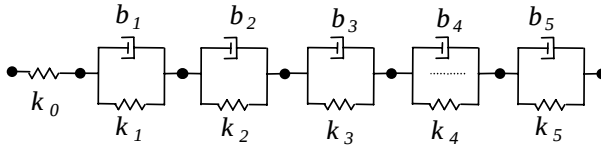


Fig. 9. Hydraulic arrangement equivalent mechanical diagram.

Finally, $D_N(s)$ can be written:

$$D_N(s) = \frac{1}{\frac{1}{k_0} + \sum_{i=1}^N \frac{1/b_i}{s + \omega_{zi}}}, \quad (17)$$

by introducing

$$\omega_{zi} = \frac{k_i}{b_i}. \quad (18)$$

In order to establish the relations between the mechanical parameters k_i and b_i (or hydropneumatic C_i and R_i) and the recursive distribution of the transitional frequencies ω_i and ω'_i , the inverse of relation (17), namely:

$$D_N^{-1}(s) = \frac{1}{k_0} + \sum_{i=1}^N \frac{1/b_i}{s + \omega_{zi}}, \quad (19)$$

is interpreted as the decomposition in simple elements of the inverse of relation (4), that is to say:

$$D_N^{-1}(s) = \frac{1}{D_0} \prod_{i=1}^N \left(\frac{1 + \frac{s}{\omega_i}}{1 + \frac{s}{\omega_i'}} \right) = \frac{1}{D_0} \prod_{i=1}^N \frac{\omega_i'}{\omega_i} + \sum_{i=1}^N \frac{A_i}{s + \omega_i'}, \quad (20)$$

with

$$A_i = \frac{1}{D_0} \left(\prod_{l=1}^N \frac{\omega_l'}{\omega_l} \right) \left(\frac{\prod_{l=1, l \neq i}^N (\omega_l - \omega_i')}{\prod_{l=1, l \neq i}^N (\omega_l' - \omega_i)} \right). \quad (21)$$

Member with member identification of the relations (19) and (20) makes it possible to determine the mechanical parameters k_0 , b_i , and k_i , namely:

$$k_0 = D_0 \prod_{i=1}^N \frac{\omega_i'}{\omega_i}, \quad b_i = \frac{1}{A_i} \quad \text{and} \quad k_i = \omega_i' b_i, \quad (22)$$

as well as hydropneumatic parameters C_0 , R_i , and C_i , that is to say, taking into account the relations (16).

Lastly, the technological parameters such as the calibration pressure P_{0i} and the volume V_{0i} of each accumulator, the diameter d_{Ri} and the length l_{Ri} of each resistance, are deduced from relations (9) and (10) by taking into account the technological constraints associated to each component.

Gamma Arrangement

For simplification purpose, the following relations are established in the case of six hydropneumatic accumulators and five hydraulic resistors. It is, of course, possible to extend these relations to more or less components.

The input hydraulic impedance of the gamma arrangement is characterized by:

$$\frac{P_e(s)}{Q_e(s)} = \frac{1}{C_0 s + \frac{1}{R_1 + \frac{1}{C_1 s + \frac{1}{\dots + \frac{1}{R_5 + \frac{1}{C_5 s}}}}}}. \quad (23)$$

To compare this expression to the desired hydraulic impedance, which is obtained by dividing relation (4) by $S_v^2 s$, the two transfer functions have to be written under the same form. The most valuable form for comparison is continuous fraction.

Expressions (4) and (25) can be converted to the same simple continuous fraction form, namely:

$$\frac{P_e(s)}{Q_e(s)} = \frac{1}{A_0 s + B_0 + \frac{1}{A_2 s + B_2 + \frac{1}{A_4 s + B_4 + \frac{1}{A_5 s + B_5}}}}. \quad (24)$$

Member to member identification makes it possible to determine the mechanical parameters k_0 , b_i , and k_i in the case of gamma arrangement as well as hydropneumatic parameters C_0 , R_i , and C_i , that is to say, taking into account the relations (16).

Technological parameters are obtained thanks to relations (9) and (10).

3.4 Important note

Capacities C_i depend on the static pressure (9). The static pressure P_s can be expressed according to the weight Mg and of the jack section S_v , namely:

$$P_s = \frac{M g}{S_v}. \quad (25)$$

By replacing in relation (9) P_s by its expression (24), one notes that capacities C_i , and thus stiffnesses k_i , depend on the square of the sprung mass M , that is to say:

$$C_i = \frac{S_v^2 P_{0i} V_{0i}}{\gamma (Mg)^2} \quad \text{and} \quad k_i = \gamma \frac{(Mg)^2}{P_{0i} V_{0i}}. \quad (26)$$

Thus, the variations or uncertainties of the sprung mass M not only affect the plant $G(s)$ as defined in paragraph 2, but also the real form $D_N(s)$ of the controller because of the relations between the physical parameters and the parameters of $D_N(s)$.

This result leads to a new problematic in automatic control. Usually, controller uncertainties are not taken into account because they are much smaller than the plant uncertainties. In the particular case of an hydropneumatic achievement, controller uncertainties are not negligible any more.

A method to take into account the characteristics related to hydropneumatic technology during the synthesis of the limited-bandwidth fractional differentiator has been developed in [8]. It is based on two remarkable properties of the limited-bandwidth fractional differentiator achieved in hydropneumatic technology. Indeed, it was shown that, in this case, the recursive parameters α and η are independent of the variations of the mass M . So, the fractional-order asymptotic behaviour of $D_N(s)$, characterized for the gain diagram by a slope of m20dB/dec and for the phase diagram by a phase blocking of $m \frac{\pi}{2}$ rad, is not modified; only the frequency domain where this asymptotic behaviour exists is relocated towards the high frequencies when the mass increases (and reciprocally towards the low frequencies when it decreases). Moreover, it is shown that the open-loop crossover frequency can remain insensitive to the variations of M . It is then possible to take into account these two results in the determination of the high-level synthesis parameters of the limited-bandwidth fractional differentiator to translate them, not only by the robustness of the degree of stability (intrinsic property with CRONE approach), but also by the robustness of the rapidity (intrinsic property to hydropneumatic technology).

From the following specifications [7]:

- For the rapidity, an open-loop crossover frequency ω_u of 6 rad/s
- For the stability, a phase margin M_ϕ of 45°
- For the uncertainties, $M \in [75 \text{ kg}; 150 \text{ kg}]$

the four high-level synthesis parameters of the ideal form $D(s)$ for the minimal mass $M = 75 \text{ kg}$ are calculated [7], namely:

$$m = 0.5, \quad \omega_b = 0.1 \text{ rad/s}, \quad \omega_h = 90 \text{ rad/s}$$

$$\text{and} \quad D_0 = 349 \text{ N/m} \quad (27)$$

Then, thanks to the relations between the four high-level synthesis parameters and the $2N$ real form parameters, the transitional frequency and

the recursive distribution can be calculated, always considering the minimal mass, with $N = 5$.

In the case of a parallel arrangement of RC cells, the physical parameters are calculated from relations (22) and in the case of a gamma arrangement from relations (23) and (24).

Lastly, from the jack section, and the relations (9) and (10), the technological parameters are deduced. (Volumes V_{oi} come from constructor data sheets.)

The numeric values of all parameters can be found in [8].

3.5 Performances

Within the framework of a comparative study, the parameters of arrangement with two cells (whose RC, N is = 1 for the traditional suspension) are calculated starting from the same specifications as previously. Thus, for the minimal mass of 75 kg, the three systems present the same dynamic.

Figure 10 presents the frequency responses obtained with the traditional suspension ((a), (c), and (e)) and CRONE suspension ((b), (d), and (f)) for the two extreme values of the sprung mass M (blue dotted line: $M = 75$ kg, in green $M = 150$ kg). The results obtain with the parallel arrangement of RC cells and with the gamma arrangement are exactly the same. This first significant result explains why only one diagram is shown the CRONE suspension and not one for each structure.

The transfer functions $D_1(s)$ (traditional suspension (a)) and $D_5(s)$ (CRONE suspension (b)) Bode diagrams highlights the influence of a mass M increase. Indeed, in both cases, the static gain D_0 increases and the transitional frequencies is relocated towards the high frequencies without the maximum in lead-phase not being modified. Thus, the length of the frequency domain which characterizes the second generation CRONE control [5] is dimensioned so that the open-loop crossover frequency belongs to this fractional-order asymptotic behaviour whatever the mass M values is between 75 and 150 kg. This result is illustrated by the open-loop bode diagrams (Fig. 10d) and Black-Nichols loci (Fig. 10f) in the case of CRONE suspension.

Moreover, around the crossover frequency ω_u (Fig. 10d and 10f Bode diagrams) for the two extreme values of M , one observes that the controller gain variations and the plant gain variations compensate for each other. This means an open-loop crossover frequency insensitive to the mass variations. Thus, the insensitivity of ω_u ensures the robustness of the rapidity (intrinsic property of hydropneumatic technology) and the constancy of the phase margin M_ϕ ensures the robustness of the stability degree (intrinsic property of CRONE approach) towards the sprung mass variations.

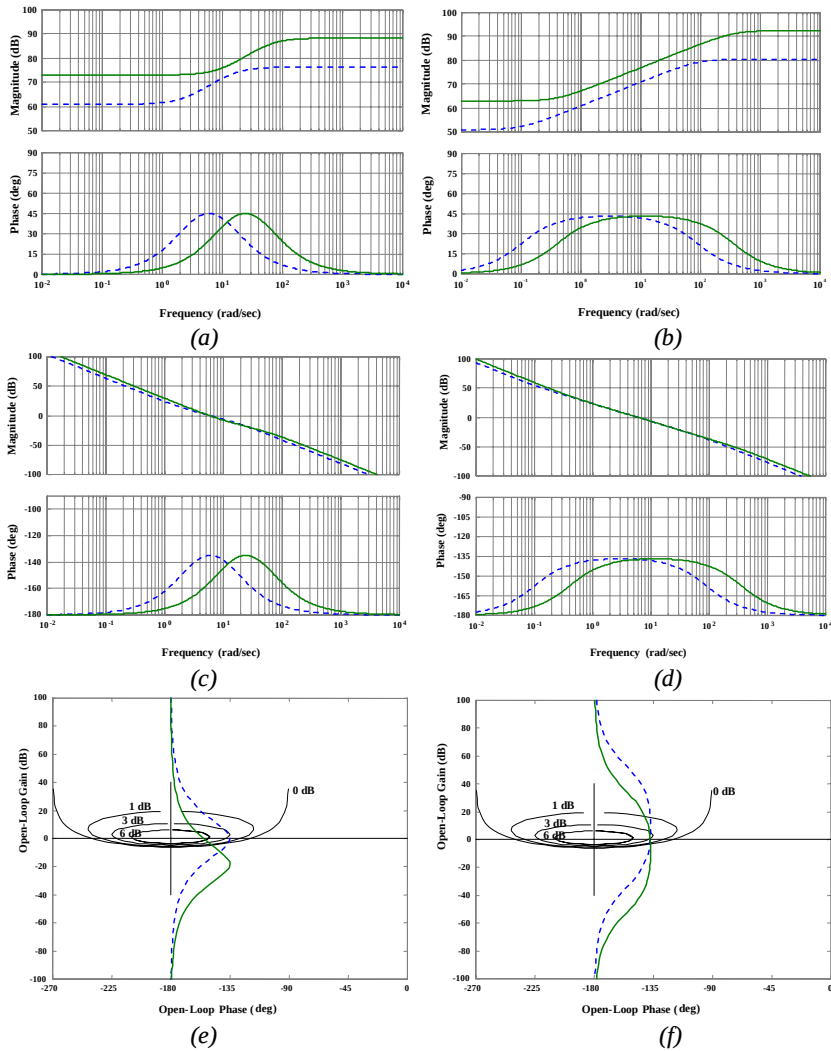


Fig. 10. Frequency responses obtained with the traditional suspension ((a), (c), and (e)) and the CRONE suspension ((b), (d), and (f)) for the extreme mass values M .

These properties are illustrated in Fig. 11. Figure 11 presents the time responses of a release test obtained with the traditional suspension (a) and with the CRONE suspension (b) for the two sprung mass extreme values (*blue dotted line*: $M = 75$ kg, in *green* $M = 150$ kg). The mass is initially held in a high position and released at $t = 0$ ($z_1(0^-) = 1$).

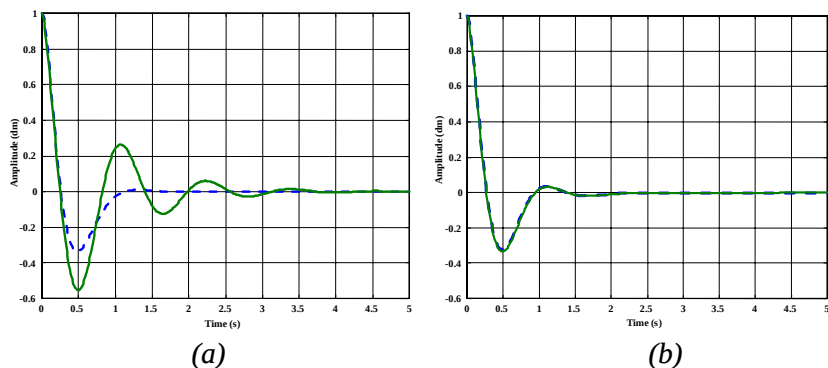


Fig. 11. Time responses of the traditional suspension (a) and of the CRONE suspension (b) with the two extreme sprung mass values M (blue dotted line: $M = 75$ kg, in green $M = 150$ kg).

4 Conclusion

The performances of the test bench presented in this article make it possible to highlight the interest of fractional derivative in vibration isolation. The performances obtained are remarkable, in particular when suspension CRONE is achieved in hydropneumatic technology from a method of synthesis based on the frequential recursivity. Indeed, the association of CRONE approach for the synthesis and of hydropneumatic technology for the achievement makes it possible to obtain the robustness of the stability degree, but also the robustness of the rapidity towards the mass variations. Two structures have been presented with the advantages and performance. An interesting point is that the gamma structure induces less technological parameters dispersion than the parallel arrangement of RC cells. This can allow to achieve more easily industrial applications of the CRONE suspension because of the standardization of the hydraulic accumulator in the gamma arrangement case. The next step of this work consists in taking into account the component non-linearities. The final objective is to find a design rule for vehicle hydraulic resistors which are non-linear for functional reasons.

The principal industrial applications of this work are the automobile suspensions in particular with the Hydractive CRONE suspension [9]. It should be noted that the Hydractive CRONE suspension which presents three operating modes (a comfort mode, an intermediate mode and a safety mode) is the origin of the definition of a new class of systems, namely the hybrid fractional dynamic systems (HFDS) [10].

Acknowledgment

Thanks go to the American Society of Mechanical Engineers (ASME) for the permission to publish this revised contribution of an ASME article.

References

1. Le Mehauté A (1991) *Fractal Geometries*. CRC Press, Ann Arbor, London.
2. Onaral B, Schwan HP (1982) Linear and non linear properties of platinum electrode polarization, Part I, Frequency dependence at very low frequencies, *Med. Bio. Eng. Comput.*, 20:299–306.
3. Moreau X, Oustaloup A, Nouillant M (1999) From analysis to synthesis of vehicle suspensions: the CRONE approach, *Proceedings of ECC'99, Karlsruhe*.
4. Moreau X, Ramus-Serment C, Oustaloup A (2002) Fractional Differentiation in Passive Vibration Control, Special Issue on Fractional Calculus in the *J. Nonlinear Dyn.*, 29:343–362.
5. Oustaloup A (1991) *La commande CRONE*. Edition Hermès, Paris.
6. Oustaloup A (1995) *La dérivation non entière: théorie, synthèse et applications*. Edition Hermès, Paris.
7. Binder RC (1973) *Fluid Mechanics*. Prentice-Hall, Englewood Cliffs, NJ.
8. Serrier P, Moreau X, Oustaloup A (2005) Synthesis of a limited bandwidth fractional differentiator made in hydropneumatic technology. *ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Long Beach, CA.
9. Serrier P (2004) Synthèse fondée sur la récursivité fréquentielle d'un dérivateur d'ordre non entier borné en fréquence réalisé en technologie hydropneumatique – Application à la suspension CRONE Hydractive, *Mémoire de stage MASTER EEA Recherche*, Ecole Doctorale des Sciences Physiques et de l'Ingénieur de l'Université Bordeaux I.
10. Altet O, Nouillant C, Moreau X, Oustaloup A (2003) Hydractive CRONE suspension as hybrid system, *Int. J. Hybrid Syst.*, 3(2,3):165–188.

Part 5

Electrical Systems

A FRACTIONAL CALCULUS PERSPECTIVE IN THE EVOLUTIONARY DESIGN OF COMBINATIONAL CIRCUITS

Cecília Reis¹, J. A. Tenreiro Machado¹, and J. Boaventura Cunha²

¹ Institute of Engineering of Porto, Rua Dr. António Bernardino de Almeida,
Porto, Portugal; E-mail: cmr,jtm@isep.ipp.pt

² University of Trás-os-Montes and Alto Douro, Engenharias II, Vila Real,
Portugal; E-mail: jboavent@utad.pt

Abstract

This paper analyses the performance of a genetic algorithm (GA) in the synthesis of digital circuits using two novel approaches. The first concept consists in improving the static fitness function by including a discontinuity evaluation. The measure of variability in the error of the Boolean table has similarities with the function continuity issue in classical calculus. The second concept extends the static fitness by introducing a fractional-order dynamical evaluation. The dynamic-fitness function results from an analogy with control systems where it is possible to benefit the proportional algorithm by including a differential scheme. It is investigated the GA performance when adopting each concept separately. The experiments reveal superior results, in terms of speed and convergence of the number of iterations required to achieve a solution. In a final phase the two concepts are integrated in the GA fitness function leading to the best performance.

Keywords

Circuit design, fractional systems, genetic algorithms, logic circuits.

1 Introduction

In the last decade genetic algorithms (GAs) have been applied in the design of electronic circuits, leading to a novel area of research called evolutionary Electronics (EE) or evolvable hardware (EH) [1]. EE considers the concept for automatic design of electronic systems. Instead of using human-conceived models, abstractions, and techniques, EE employs search algorithms to develop good designs [18].

One decade ago Sushil and Rawlins [9] applied GAs to the combinational circuit design problem. They combined knowledge-based systems with the GA

and defined a genetic operator called masked crossover. This scheme leads to other kinds of offspring that can not be achieved by classical crossover operators.

John Koza [2] adopted genetic programming to design combinational circuits.

In the sequence of this work, Coello, Christiansen, and Aguirre [19] presented a computer program that automatically generates high-quality circuit designs. They use five possible types of gates (AND, NOT, OR, XOR, and WIRE) with the objective of finding a functional design that minimizes the use of gates other than WIRE.

Miller, Thompson, and Fogarty [3] applied evolutionary algorithms for the design of arithmetic circuits. The technique was based on evolving the functionality and connectivity of a rectangular array of logic cells, with a model of the resources available on the Xilinx 6216 FPGA device.

Kalganova, Miller, and Lipnitskaya [10] proposed a new technique for designing multiple-valued circuits.

In order to solve complex systems, Torresen [11] proposed the method of increased complexity evolution. The idea is to evolve a system gradually as a kind of divide-and-conquer method. Evolution is first undertaken individually on simple cells. The evolved functions are the basic blocks adopted in further evolution of more complex systems.

A major bottleneck in the evolutionary design of electronic circuits is the problem of scale. This refers to the very fast growth of the number of gates, used in the target circuit, as the number of inputs of the evolved logic function increases. This results in a huge search space that is difficult to explore even with evolutionary techniques. Another related obstacle is the time required to calculate the fitness value of a circuit [13, 3]. A possible method to solve this problem is to use building blocks either than simple gates. Nevertheless, this technique leads to another difficulty, which is how to define building blocks that are suitable for evolution.

Gordon and Bentley [14] suggest an approach that allows evolution to search for good inductive bases for solving large-scale complex problems. This scheme generates, inherently, modular, and iterative structures, that exist in many real-world circuit designs but, at the same time, allows evolution to search innovative areas of space.

The idea of using memory to achieve better fitness function performances was first introduced by Sano and Kita [15]. Their goal was the optimization of systems with randomly fluctuating fitness function and they developed a genetic algorithm with memory-based fitness evaluation (MFEGA). The key ideas of the MFEGA are based on storing the sampled fitness values into memory as a search history, introducing a simple stochastic model of fitness values to be able to estimate fitness values of points of interest using the history for selection operation of the GA.

Following this line of research, and looking for better performance GAs, this paper proposes a GA for the design of combinational logic circuits using fractional-order dynamic fitness functions.

The area of fractional calculus (FC) deals with the operators of integration and differentiation to an arbitrary (including noninteger) order and is as old as the theory of classical differential calculus [4, 5]. The theory of FC is a well-adapted tool to the modelling of many physical phenomena, allowing the description to take into account some peculiarities that classical integer-order models simply neglect. Nevertheless, the application of FC has been scarce until recently, but the advances on the theory of chaos motivated a renewed interest in this field. In the last two decades we can mention research on viscoelasticity/damping, chaos/fractals, biology, signal processing, system identification, diffusion and wave propagation, electromagnetism, and automatic control [6, 7, 20, 8].

Bearing these ideas in mind the article is organized as follows. Section 2 describes the adopted GA as well as the fractional-order dynamic fitness functions. Section 3 presents the simulation results and finally, section 4 outlines the main conclusions and addresses perspectives towards future developments.

2 The Adopted Genetic Algorithm

In this section we present the GA in terms of the circuit encoding as a chromosome, the genetic operators, and the static and dynamic fitness functions.

2.1 Problem definition

The circuits are specified by a truth table with input bits ordered according with the Gray code. The goal is to implement a functional circuit with the least possible complexity. Two sets of logic gates have been defined, as shown in Table 1, being Gset a the simplest one (i.e., a RISC-like set) and Gset b a more complex gate set (i.e., a CISC-like set).

For each gate set the GA searches the solution space, based on a simulated evolution aiming the survival of the fittest strategy. In general, the best individuals of any population tend to reproduce and survive, thus improving successive generations. However, inferior individuals can, by chance, survive and also reproduce. In our case, the individuals are digital circuits, which can evolve until the solution is reached (in terms of functionality and complexity).

2.2 Circuit encoding

In the GA scheme the circuits are encoded as a rectangular matrix $\mathbf{A}$ (row $\times$ column = $r \times c$) of logic cells as represented in Fig. 1, having inputs $\mathbf{X}$ and outputs $\mathbf{Y}$.

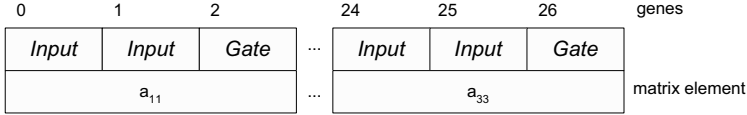


Fig. 2. Chromosome for the 3×3 matrix of Fig. 1.

The crossover rate CR represents the percentage of the population P that reproduces in each generation. Likewise the mutation rate MR is the percentage of the population P circuits that can mutates in each generation.

2.4 The static and the dynamic fitness functions

The goal of this study is to find new ways of evaluating the individuals of the population in order to achieve GAs with superior performance.

In this paper we propose two concepts for the improvement of the standard static fitness function F_s by:

- Introducing a discontinuity measure δ
- Including a differential term leading to a dynamical fitness function F_d

The calculation of the F_s in (1) is divided in two parts, f_1 and f_2 , where f_1 measures the circuit functionality and f_2 measures the circuit simplicity. In a first phase, we compare the output $\mathbf{Y}$ produced by the GA-generated circuit with the required values $\mathbf{Y}_R$, according with the truth table, on a bit-per-bit basis. By other words, f_{11} is incremented by *one* for each correct bit of the output until f_{11} reaches the maximum value f_{10} , that occurs when we have a functional circuit (Eq. 1a and 1b).

In order to measure the output error variability (subsections 3.1 and 3.3) f_{11} is decremented by δ for each $\mathbf{Y}_R - \mathbf{Y}$ error discontinuity (i.e., when passing from $\mathbf{Y}_R - \mathbf{Y} = 0$ to $\mathbf{Y}_R - \mathbf{Y} = 1$, or *vice-versa*) by comparing two consecutive levels of the truth table (Eq. 1c).

Once the circuit is functional, in a second phase, the GA tries to generate circuits with the least number of gates. This means that the resulting circuit must have as much genes $\langle gate\ type \rangle \equiv \langle wire \rangle$ as possible. Therefore, the index f_2 , that measures the simplicity (the number of null operations), is increased by *one* (*zero*) for each *wire* (*gate*) of the generated circuit (Eq. 1d), yielding:

- First phase, circuit functionality:

$$f_{10} = 2^{ni} \times no \quad (1a)$$

$$f_{11} = f_{11} + 1, \text{ if } \{bit\ i\ of\ \mathbf{Y}\} = \{bit\ i\ of\ \mathbf{Y}_R\}, i = 1, \dots, f_{10} \quad (1b)$$

$$f_1 = f_{11} - \delta, \text{ if } error_i \neq error_{i-1}, i = 1, \dots, f_{10} \quad (1c)$$

(when measuring discontinuity)

- Second phase, circuit simplicity:

$$f_2 = f_2 + 1, \text{ if gate type} = \text{wire} \quad (1d)$$

$$F_s = \begin{cases} f_1, & F_s < f_{10} \\ f_1 + f_2, & F_s \geq f_{10} \end{cases} \quad (1e)$$

where $i = 1, \dots, f_{10}$, ni , and no represent the number of inputs and outputs of the circuit.

The concept of dynamic-fitness function F_d results from an analogy with control systems, where we have a variable to be controlled, similarly with the GA case, where we master the population through the fitness function. The simplest control system is the proportional algorithm; nevertheless, there can be other control algorithms as for example, the proportional and the differential scheme.

In this line of thought, applying the static fitness function corresponds to using a kind of proportional algorithm. Therefore, to implement a proportional-derivative evolution the fitness function needs a scheme of the type:

$$F_d = F_s + K D^\mu [F_s] \quad (2)$$

where $0 \leq \mu \leq 1$ is the differential fractional-order and K is the gain of the dynamical term.

The generalization of the concept of derivative $D^\mu[f(x)]$ to noninteger values of μ goes back to the beginning of the theory of differential calculus. In fact, Leibniz, in his correspondence with Bernoulli, L'Hôpital, and Wallis, had several notes about its calculation for $\mu = 1/2$ [4, 5]. Nevertheless, the adoption of the FC in control algorithms has been recently studied using the frequency and discrete-time domains [6, 7, 20].

The mathematical definition of a derivative of fractional-order μ has been the subject of several different approaches. For example, Eq. (3) and Eq. (4), represent the Laplace (for zero initial conditions) and the Grünwald–Letnikov definitions of the fractional derivative of order μ of the signal $x(t)$:

$$D^\mu [x(t)] = L^{-1} \{s^\mu X(s)\} \quad (3)$$

$$D^\mu [x(t)] = \lim_{h \rightarrow 0} \frac{1}{h^\mu} \sum_{k=0}^{\infty} \frac{(-1)^k \Gamma(\mu+1)}{k! \Gamma(\mu-k+1)} x(t-kh) \quad (4)$$

where Γ is the gamma function and h is the time increment. This formulation [20] inspired a discrete-time calculation algorithm, based on the approximation of the time increment h through the sampling period T and a r -term truncated series yielding the equation:

$$D^\mu [x(t)] \approx \frac{1}{T^\alpha} \sum_{k=0}^r \frac{(-1)^k \Gamma(\mu+1)}{k! \Gamma(\mu-k+1)} x(t-kT) \quad (5)$$

3 Experiments and Simulation Results

A reliable execution and analysis of a GA usually requires a large number of simulations to provide that stochastic effects have been properly considered. Therefore, in this study are executed $n = 100$ simulations for each case.

The experiments consist on running the GA in order to generate a combinational logic circuit with the gate sets presented in Table 1, $CR = 95\%$, $MR = 20\%$ and $P = 100$ circuits, using the fitness scheme described previously.

In this section are adopted three case studies corresponding to a 4-bit parity checker ($PC4$), a 2-to-1 multiplexer ($M2-1$) and a 1-bit full adder ($FA1$) as follows: and a 2-bit multiplier ($MUL2$) as follows:

- the $PC4$ circuit, has 4 inputs $\mathbf{X} = \{A_3, A_2, A_1, A_0\}$ and 1 output $\mathbf{Y}_R = \{P\}$. The matrix $\mathbf{A}$ size is 4×4 , and the length of each string representing a circuit (i.e., the chromosome length) is $CL = 48$,
- the $M2-1$ circuit, has 3 inputs $\mathbf{X} = \{S_0, I_1, I_0\}$ and 1 output $\mathbf{Y}_R = \{O\}$. The matrix $\mathbf{A}$ size is 3×3 , and $CL = 27$,
- the $FA1$ circuit, has 3 inputs $\mathbf{X} = \{A, B, C_{in}\}$ and 2 outputs $\mathbf{Y}_R = \{S, C_{out}\}$. The matrix $\mathbf{A}$ size is 3×3 , and $CL = 27$.

The input bits are grouped according with the Gray code and Table 2 presents the Boolean truth tables for the circuits under study.

The implementation of the differential fractional-order-operator adopts Eq. (5) with a series truncation of $r = 50$ terms.

Having these ideas in mind, a superior GA performance means achieving solutions with a smaller number N of generations, in order to accelerate convergence and a smaller variability, deviation in order to reduce the stochastic nature of the algorithm.

Due to the huge number of possible combinations of the GA parameters, in the sequel we evaluate only a limited set of cases. Therefore, *a priori*, other values can lead to different results. Nevertheless, the authors developed an extensive number of numerical experiments and concluded that the following cases are representative.

Table 2. The truth tables of the *PC4*, *M2 – 1* and *FA1* circuits

<i>PC4</i>					<i>M2 – 1</i>				<i>FA1</i>			
A_3	A_2	A_1	A_0	P	S_0	I_1	I_0	O	A	B	C_{in}	S C_{out}
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	1	0	0	1	1	0	0	1	1
0	0	1	1	0	0	1	1	1	0	1	1	0
0	0	1	0	1	0	1	0	0	0	1	0	1
0	1	1	0	0	1	1	0	1	1	1	0	1
0	1	1	1	1	1	1	1	1	1	1	1	1
0	1	0	1	0	1	0	0	0	1	0	0	0
0	1	0	0	1	1	0	0	0	1	0	0	1
1	1	0	0	0								
1	1	0	1	1								
1	1	1	1	0								
1	1	1	0	1								
1	0	1	0	0								
1	0	1	1	1								
1	0	0	1	0								
1	0	0	0	1								

3.1 Fitness with discontinuity measure

In this first subsection we analyze the GA improvement when adopting a static fitness function including the discontinuity measure δ error.

Figures 3–5 show the average number of generations $AV(N)$ and the corresponding standard deviation $SD(N)$ to achieve the solution versus the discontinuity factor $\delta \in [0,1]$, using Gsets {a, b} for the *PC4*, *M2 – 1* and the *FA1* circuits, respectively.

The results reveal that [21] Gset a presents better performance than Gset b for all values of δ . On the other hand, analyzing the influence of δ we conclude that the GA response is best mostly in the region $\delta \approx 0.5$ for the three circuits and for the two gate sets.

3.2 Fitness with dynamical term

In this second subsection we analyze the GA performance when adopting a dynamical scheme for the fitness function.

The simulations investigate the differential scheme $\mu = \{0.0, 0.25, 0.5, 0.75, 1.0\}$ in F_d for gains in the range $K \in [0,1]$.

Figures 6–8 show the average number of generations $AV(N)$ and the standard deviation $SD(N)$ to achieve a solution versus K with F_d , for the *PC4*, *M2 – 1* and *FA1* circuits, using the Gsets {a, b}, respectively. For comparison the charts include the case $\mu = 0$, that corresponds to the static function F_s . We verify that F_d produces better results than the classical F_s .

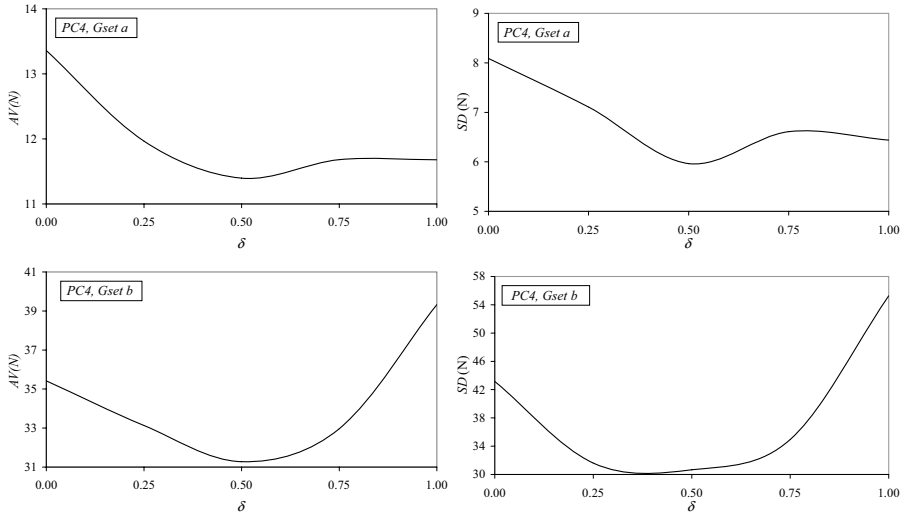


Fig. 3. Average number of generations $AV(N)$ and standard deviation $SD(N)$ to achieve a solution for the $PC4$ circuit versus $\delta \in [0,1]$, with Gsets $\{a, b\}$ and F_s .

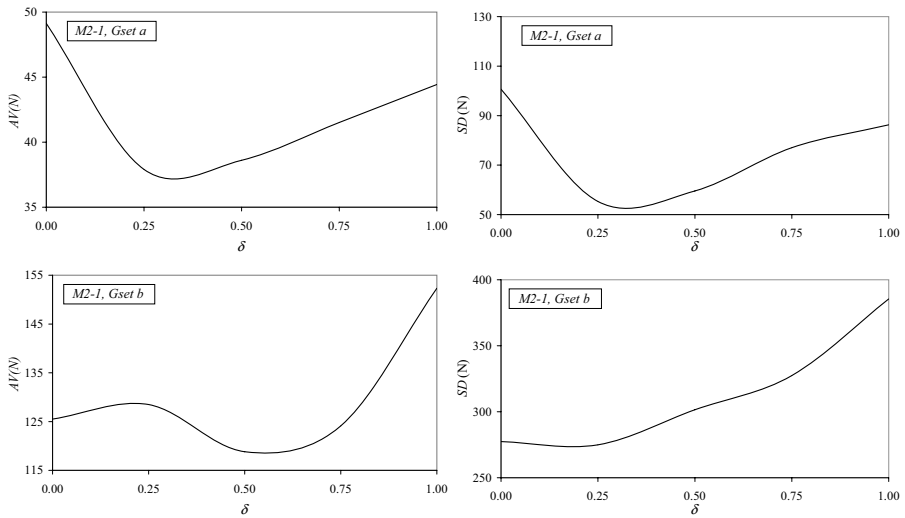


Fig. 4. Average number of generations $AV(N)$ and standard deviation $SD(N)$ to achieve a solution for the $M2-1$ circuit versus $\delta \in [0,1]$, with Gsets $\{a, b\}$ and F_s .

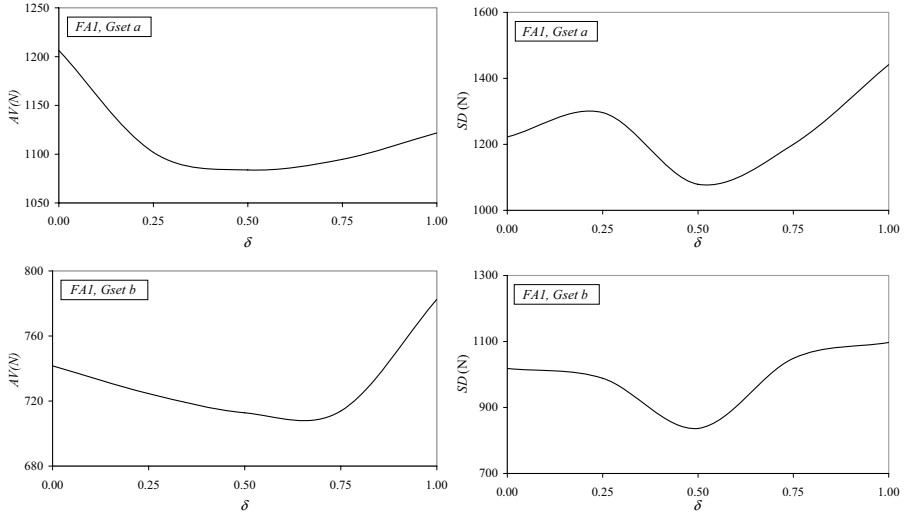


Fig. 5. Average number of generations $AV(N)$ and standard deviation $SD(N)$ to achieve a solution for the *FA1* circuit versus $\delta \in [0,1]$, with Gsets $\{a, b\}$ and F_s .

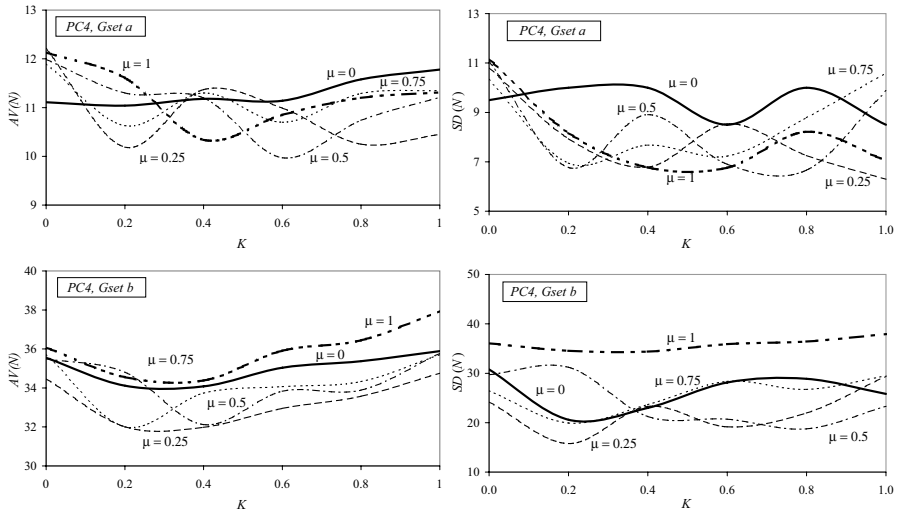


Fig. 6. Average number of generations $AV(N)$ and standard deviation $SD(N)$ to achieve a solution for the *PC4* circuit versus $K \in [0,1]$, with Gsets $\{a, b\}$ and F_d .

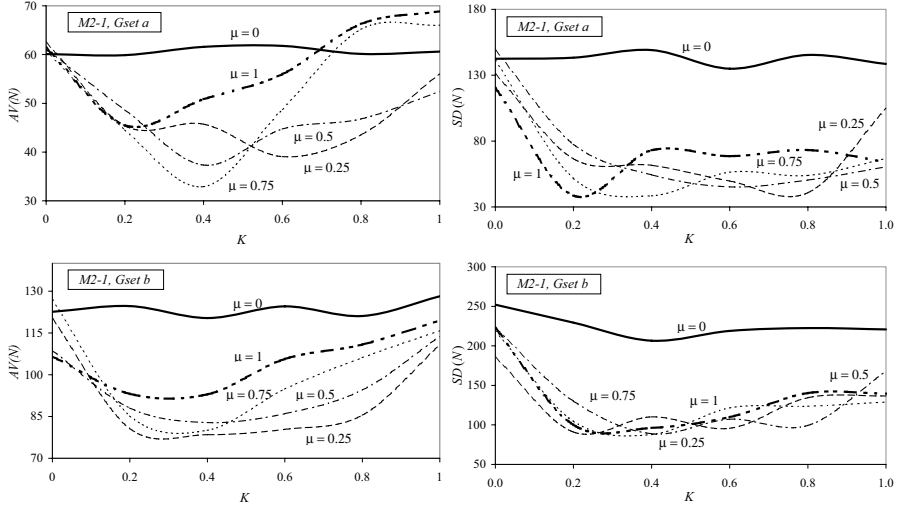


Fig. 7. Average number of generations $AV(N)$ and standard deviation $SD(N)$ to achieve a solution for the $M2-1$ circuit versus $K \in [0,1]$, with Gsets. {a, b} and F_d .

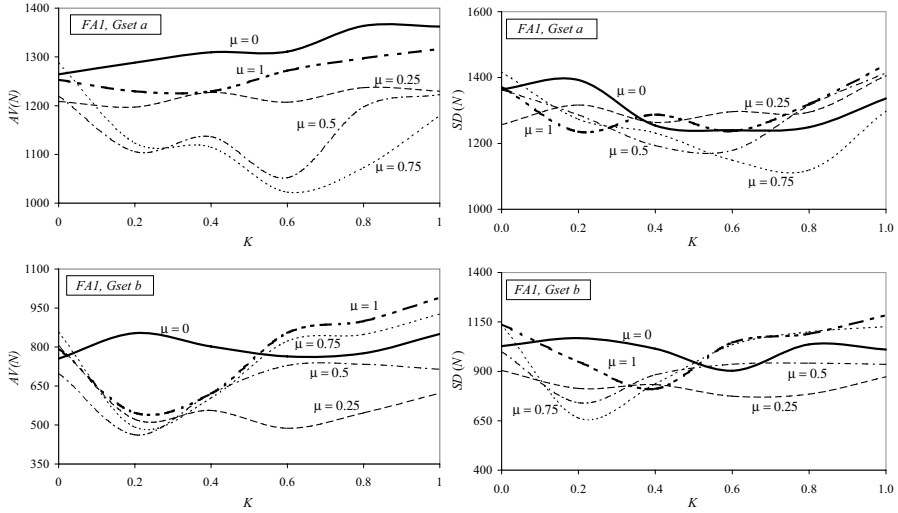


Fig. 8. Average number of generations $AV(N)$ and standard deviation $SD(N)$ to achieve a solution for the $FA1$ circuit versus $K \in [0,1]$, with Gsets {a, b} and F_d .

3.3 Fitness with discontinuity and dynamical information

In this third set of simulations, we integrate the two new concepts, namely the error discontinuity measure and the fractional-order dynamical scheme. Figures 9–11 depict the average number of generations $AV(N)$ and the standard deviation $SD(N)$ to achieve a solution versus δ , for the *PC4*, *M2-1* and *FA1* circuits, using the Gsets {a, b} with $\mu = 0.5$ and $K = 0.5$. We observe that the results are superior to the previous cases, being the best results for $\delta \approx 0.5$.

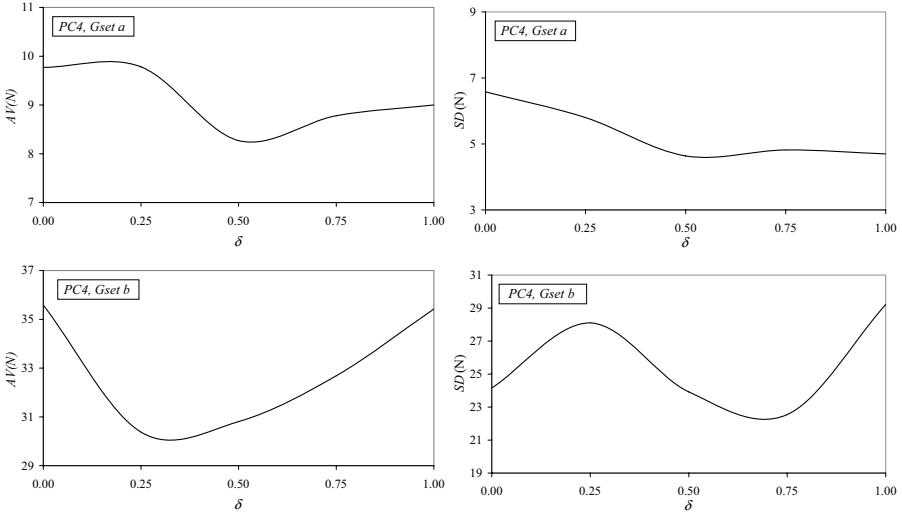


Fig. 9. Average number of generations $AV(N)$ and standard deviation $SD(N)$ to achieve a solution for the *PC4* circuit versus $\delta \in [0,1]$, with Gsets {a, b} and F_d with $\mu = 0.5$ and $K = 0.5$.

Figure 12 presents the circuits obtained by the GA.

3.4 Other circuits

In this section are addressed two complementary case studies corresponding to a 5-bit parity checker (*PC5*), a 1-bit full subtractor (*FS1*) and a 2-bit multiplier (*MUL2*), as follows:

- the *PC5* circuit, has 5 inputs $\mathbf{X} = \{A_4, A_3, A_2, A_1, A_0\}$ and 1 output $\mathbf{Y}_R = \{P\}$. The matrix $\mathbf{A}$ size is 5×5 , and the length of each string representing a circuit (i.e., the chromosome length) is $CL = 75$,
- the *FS1* circuit, has 3 inputs $\mathbf{X} = \{A, B, B_{in}\}$ and 2 outputs $\mathbf{Y}_R = \{S, B_{out}\}$. The matrix $\mathbf{A}$ size is 3×3 , and $CL = 27$,

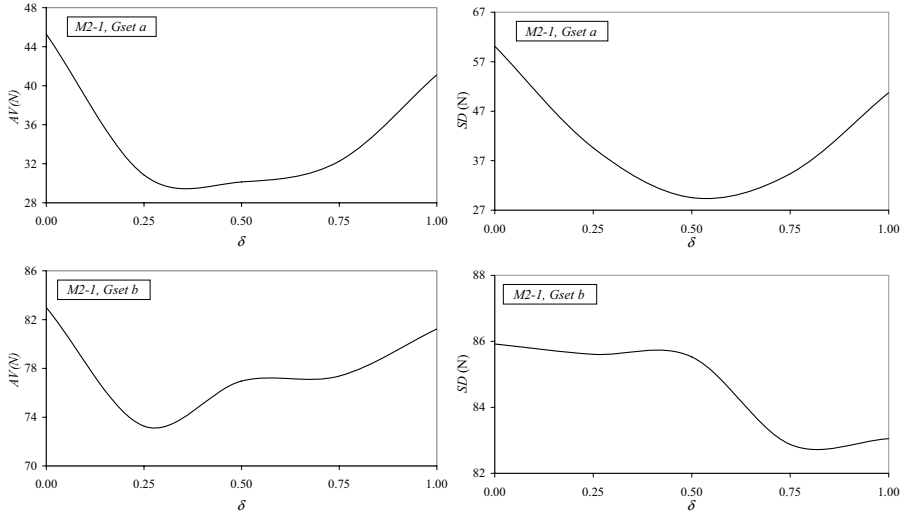


Fig. 10. Average number of generations $AV(N)$ and standard deviation $SD(N)$ to achieve a solution for the $M2-1$ circuit versus $\delta \in [0,1]$ with Gsets $\{a, b\}$ and F_d with $\mu = 0.5$ and $K = 0.5$.

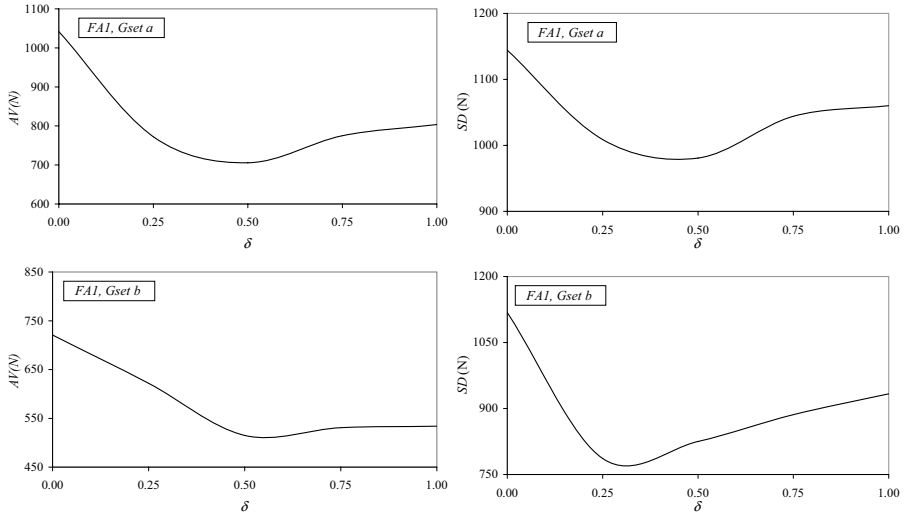


Fig. 11. Average number of generations $AV(N)$ and standard deviation $SD(N)$ to achieve a solution for the $FA1$ circuit versus $\delta \in [0,1]$ with Gsets $\{a, b\}$ and F_d with $\mu = 0.5$, $K = 0.5$.

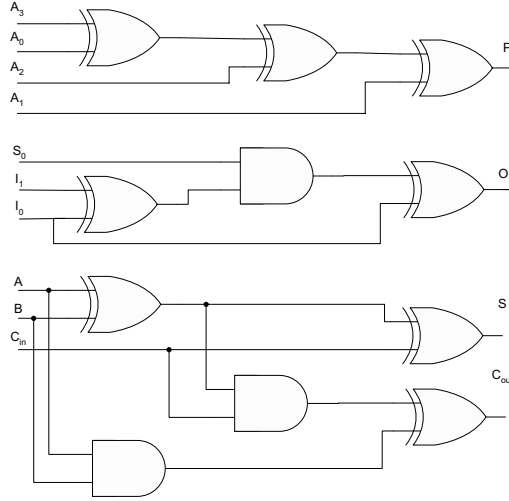


Fig. 12. The *PC4*, the *M2-1* and the *FA1* circuits.

- the *MUL2* circuit, has 4 inputs $\mathbf{X} = \{A_1, A_0, B_1, B_0\}$ and 4 outputs $\mathbf{Y}_R = \{C_3, C_2, C_1, C_0\}$. The matrix $\mathbf{A}$ size is 4×4 , and $CL = 48$.

The experiments consist on running the GA in order to generate a combinational logic circuit with the gate sets presented in Table 1, $CR = 95\%$, $MR = 20\%$, and $P = 1000$ circuits, using the fitness scheme described in section 2. Table 3 shows the truth tables for these circuits.

Figures 13–15 show $AV(N)$ and $SD(N)$ for the two gate sets when applying the different schemes with $\delta = 0.5$, $\mu = 0.5$ and $K = 0.5$ for the *PC5*, the *FS1*, and the *MUL2* circuits. Once more the charts reveal a remarkable improvement when adopting the proposed concepts.

Figure 16 shows the *PC5*, the *FS1*, and the *MUL2* circuits generated by the GA.

In conclusion, the modification of the standard fitness function concept, by introducing the discontinuity and the dynamical effects improves significantly the GA performance.

4 Conclusions

This paper presented two techniques for improving the GA performance [17]. In a first phase, we concluded that we get superior results by measuring the error discontinuity. In a second phase, we verified that the concept of fractional-order dynamical fitness function constitutes an important method to outperform the classical static approach. In a third phase the two methods

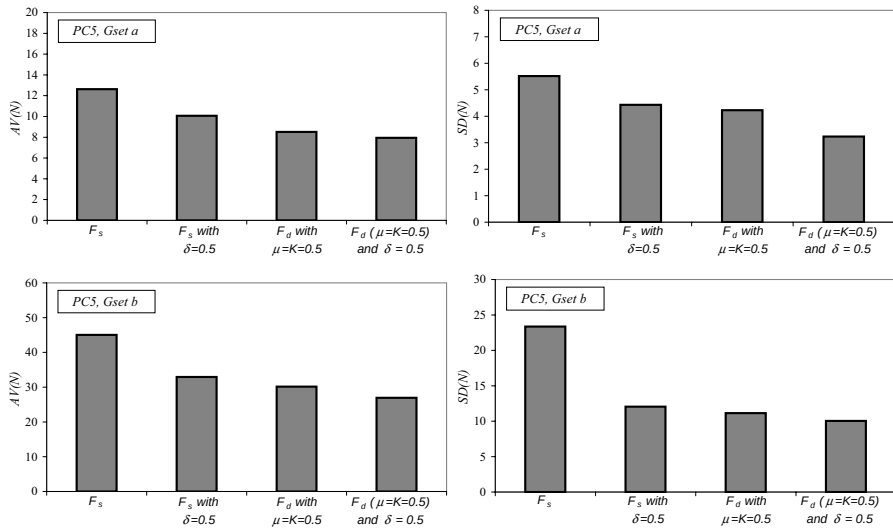


Fig. 13. Average number of generations $AV(N)$ and standard deviation $SD(N)$ to achieve a solution for the *PC5* circuit with Gsets {a, b} and the four proposed schemes.

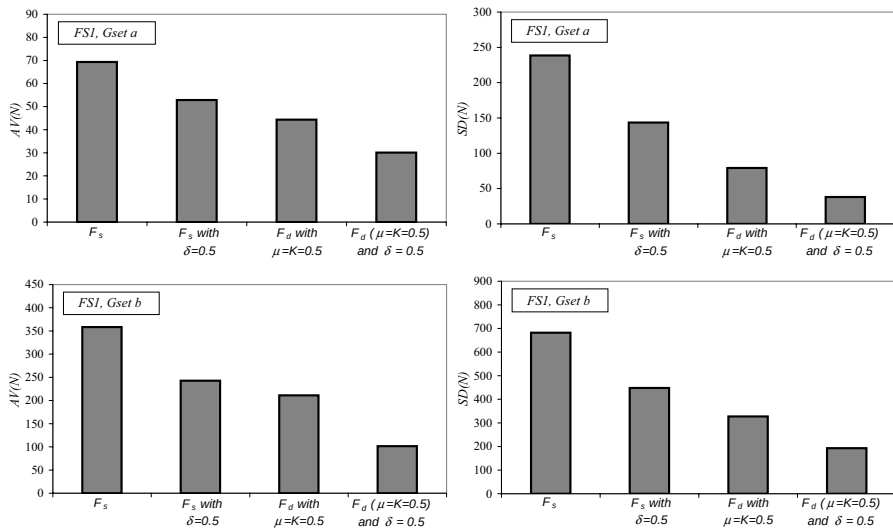


Fig. 14. Average number of generations $AV(N)$ and standard deviation $SD(N)$ to achieve a solution for the *FS1* circuit with Gsets {a, b} and the four proposed schemes.

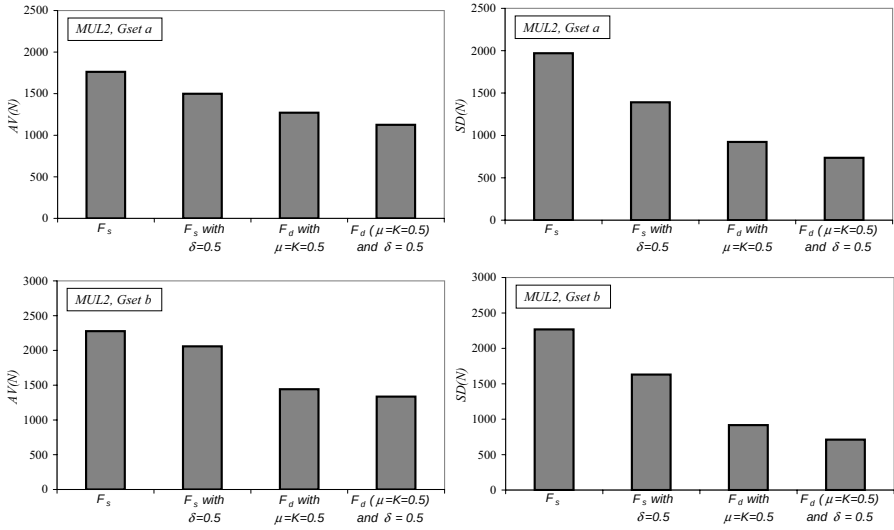


Fig. 15. Average number of generations $AV(N)$ and standard deviation $SD(N)$ to achieve a solution for the *MUL2* circuit with Gsets {a, b} and the four proposed schemes.

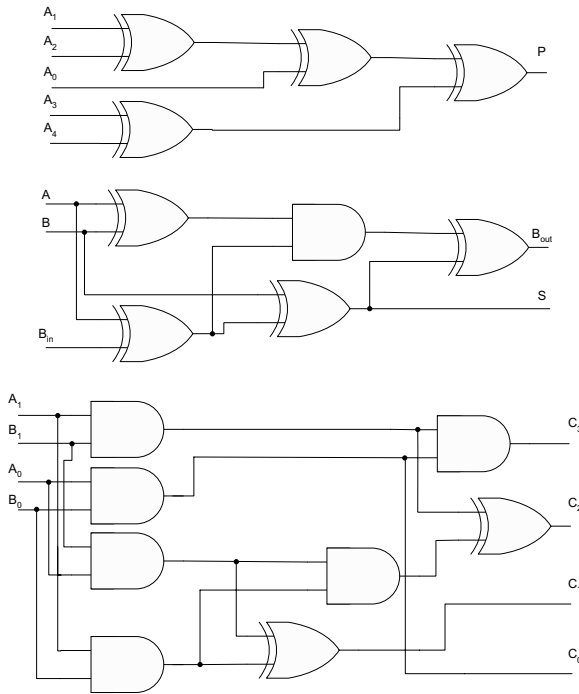


Fig. 16. The *PC5*, the *FS1*, and the *MUL2* circuits.

Table 3. The truth tables of the *PC5*, *FS1*, and *MUL2* circuits

<i>PC5</i>						<i>FS1</i>					<i>MUL2</i>							
A_4	A_3	A_2	A_1	A_0	P	A	B	B_{in}	S	B_{out}	A_1	A_0	B_1	B_0	C_3	C_2	C_1	C_0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	1	0	0	1	1	1	0	0	0	1	0	0	0	0
0	0	0	1	1	0	0	1	1	0	1	0	0	1	1	0	0	0	0
0	0	0	1	0	1	0	1	0	1	1	0	0	1	0	0	0	0	0
0	0	1	1	0	0	1	1	0	0	0	0	1	1	0	0	0	1	0
0	0	1	1	1	1	1	1	1	1	1	0	1	1	1	0	0	1	1
0	0	1	0	1	0	1	0	0	0	0	0	1	0	1	0	0	0	1
0	0	1	0	0	1	1	0	0	1	0	0	1	0	0	0	0	0	0
0	1	1	0	0	0	0					1	1	0	0	0	0	0	0
0	1	1	0	1	1	1					1	1	0	1	0	0	1	1
0	1	1	1	1	0	0					1	1	1	1	1	0	0	1
0	1	1	1	0	1	1					1	1	1	0	0	1	1	0
0	1	0	1	0	0	0					1	0	1	0	0	1	0	0
0	1	0	1	1	1	1					1	0	1	1	0	1	1	0
0	1	0	0	1	0	0					1	0	0	1	0	0	1	0
0	1	0	0	0	1	1					1	0	0	0	0	0	0	0
1	1	0	0	0	0	0												
1	1	0	0	1	1	1												
1	1	0	1	1	0	0												
1	1	0	1	0	1	1												
1	1	1	1	1	0	0												
1	1	1	1	1	1	1												
1	1	1	1	0	1	0												
1	1	1	1	0	0	1												
1	0	1	0	0	0	0												
1	0	1	0	1	1	1												
1	0	1	1	1	0	0												
1	0	0	1	0	0	0												
1	0	0	1	1	1	1												
1	0	0	0	1	0	0												
1	0	0	0	1	1	1												
1	0	0	0	0	1	0												
1	0	0	0	0	0	1												

were integrated leading to the best GA performance. The tuning of the “optimal” parameters (μ , K) was established by numerical evaluation. Therefore, future research will address the problem of having a more systematic design method. Furthermore, these conclusions encourage further studies using other fractional order dynamical schemes.

References

1. Zebulum R, Pacheco M, Vellasco M (2001) *Evolutionary Electronics: Automatic Design of Electronic Circuits and Systems by Genetic Algorithms*. CRC Press, Boca Raton, FL.

2. Koza J (1992) *Genetic Programming: On the Programming of Computers by means of Natural Selection*. MIT Press, Cambridge, MA.
3. Miller J, Thompson P, Fogarty T (1997) *Algorithms and Evolution Strategies in Engineering and Computer Science: Recent Advancements and Industrial Applications*. Wiley, New York.
4. Oldham K, Spanier J (1974) *The Fractional Calculus: Theory and Application of Differentiation and Integration to Arbitrary Order*. Academic Press, New York.
5. Miller K, Ross B (eds.) (1993) *An Introduction to the Fractional Calculus and Fractional Differential Equations*. Wiley, New York.
6. Oustaloup A (1995) *Dérivation Non Entier: Théorie, Synthèse et Applications*. HERMES, Paris.
7. Méhauté A (1991) *Fractal Geometries: Theory and Applications*. Penton Press, London.
8. Westerlund S (2002) Dead Matter Has Memory! Causal Consulting, Kalmar, Sweden.
9. Louis S, Rawlins G (1991) Designer Genetic Algorithms: Genetic Algorithms in Structure Design. In: *Proceedings of the Fourth International Conference on Genetic Algorithms*, University of California, San Diego.
10. Kalganova T, Miller J, Lipnitskaya N (1998) Multiple Valued Combinational Circuits Synthesised using Evolvable Hardware. In: *Proceedings of the Seventh Workshop on Post-Binary Ultra Large Scale Integration Systems*, Fukuoka, Japan.
11. Torresen J (1998) A Divide-and-Conquer Approach to Evolvable Hardware. In: *Proceedings of the Second International Conference on Evolvable Hardware*, Lausanne, Switzerland.
12. Hollingworth G, Smith S, Tyrrell A (2000) The Intrinsic Evolution of Virtex Devices Through Internet Reconfigurable Logic. In: *Proceedings of the Third International Conference on Evolvable Systems*, Edinburgh, UK.
13. Vassilev V, Miller J (2000) Scalability Problems of Digital Circuit Evolution. In: *Proceedings of the Second NASA/DOD Workshop on Evolvable Hardware*, Palo Alto, CA.
14. Gordon T, Bentley P (2002) Towards Development in Evolvable Hardware. In: *Proceedings of the NASA/DOD Conference on Evolvable Hardware*, Washington DC.
15. Sano Y, Kita H (2000) Optimization of Noisy Fitness Functions by means of Genetic Algorithms using History of Search. In: *Proceedings of the PPSN VI*
16. Morrison R (2003) Dispersion-Based Population Initialization. In: *Proceedings of the Genetic and Evolutionary Computation Conference*, Chicago, IL.
17. Reis C, Machado J, Cunha J (2005) Evolutionary Design of Combinational Circuits using Fractional-Order Fitness. In: *Proceedings of the Fifth EUROMECH Nonlinear Dynamics Conference*, Eindhoven University of Technology, Holland.
18. Thompson A, Layzell P (1999) Analysis of unconventional evolved electronics. *Communications of the ACM*, 42(4):71–79.
19. Coello C, Christiansen A, Aguirre A (1996) Using genetic algorithms to design combinational logic circuits. *Intelligent Engineering through Artificial Neural Networks*. ASME Press, St. Louis, Missouri, pp. 391–396.
20. Machado J (1997) Analysis and design of fractional-order digital control systems. *SAMS J. Syst. Anal., Model. Simul.* 27:107–122.
21. Reis C, Machado J, Cunha J (2004) Evolutionary design of combinational logic circuits. *J. Ad. Comput. Intell. Inform.* 8(5):507–513.

ELECTRICAL SKIN PHENOMENA: A FRACTIONAL CALCULUS ANALYSIS

J. A. Tenreiro Machado¹, Isabel S. Jesus¹, Alexandra Galhano¹,
J. Boaventura Cunha², and J. K. Tar³

¹ Department of Electrical Engineering, Institute of Engineering of Porto
Rua Dr. António Bernardino de Almeida, 4200-072 Porto, Portugal;
E-mail: jtm,isj,amf@isep.ipp.pt

² University of Trás-os-Montes and Alto Douro, Institute of Intelligent
Engineering Systems, Vila-Real, Portugal; E-mail: jboavent@utad.pt

³ Institute of Intelligent Engineering Systems, Budapest Tech, John von Neumann
Faculty of Informatics, Budapest, Hungary; E-mail: tar@nik.bmf.hu

Abstract

The internal impedance of a wire is the function of the frequency. In a conductor, where the conductivity is sufficiently high, the displacement current density can be neglected. In this case, the conduction current density is given by the product of the electric field and the conductance. One of the aspects of the high-frequency effects is the skin effect (*SE*). The fundamental problem with *SE* is it attenuates the higher frequency components of a signal.

The *SE* was first verified by Kelvin in 1887. Since then many researchers developed work on the subject and presently a comprehensive physical model, based on the Maxwell equations, is well established.

The Maxwell formalism plays a fundamental role in the electromagnetic theory. These equations lead to the derivation of mathematical descriptions useful in many applications in physics and engineering. Maxwell is generally regarded as the 19th century scientist who had the greatest influence on 20th century physics, making contributions to the fundamental models of nature.

The Maxwell equations involve only the integer-order calculus and, therefore, it is natural that the resulting classical models adopted in electrical engineering reflect this perspective. Recently, a closer look of some phenomena present in electrical systems and the motivation towards the development of precise models, seem to point out the requirement for a fractional calculus approach. Bearing these ideas in mind, in this study we address the *SE* and we re-evaluate the results demonstrating its fractional-order nature.

Keywords

Skin effect, eddy currents, electromagnetism, fractional calculus.

1 Introduction

Some experimentation with magnets was beginning in the late 19th century. By then reliable batteries had been developed and the electric current was recognized as a stream of charge particles. Maxwell developed a set of equations expressing the basic laws of electricity and magnetism, and he demonstrated that these two phenomenas are complementary aspects of electromagnetism. He showed that electric and magnetic fields travel through space, in the form of waves, at a constant velocity.

The skin effect (*SE*) is one subject who can be explained by the Maxwell's equations. The *SE* is the tendency of a high-frequency electric current to distribute itself in a conductor so that the current density near the surface is greater than that at its core. This phenomenon increases the effective resistance of the conductor with the frequency of the current. The effect is most pronounced in radio-frequency systems, especially antennas and transmission lines [1], but it can also affect the performance of high-fidelity sound equipment, by causing attenuation in the treble range. The first study of *SE* was explained by Lord Kelvin in 1887, but many other scientists had made contributions to improve the comprehension of this theme.

The *SE* can be reduced by using stranded rather than solid wire. This increases the effective surface area of the wire for a given wire gauge. It is simple to see that the spatial variation of the fields in vacuum is much smaller than the special variation in the metal. Therefore, in usual study, for the purposes of evaluating the fields in the conductor, the spatial variation from the wave length outside the conductor can be ignored. For the usual case the radii of curvature of the surface should be much larger than a skin depth, the solution is straightforward. To analyse this phenomenon, we apply the Maxwell's equations that relate the solutions for these fields. More often, however, some of the parameters that tend to be considered are the capacitance per length, inductance per length, and their relationship with the signals, the nominal propagation velocity, and the characteristic impedance of the system.

In our study we apply the Bessel functions to compute values of cable impedance Z . For the sake of clarity we plot some values of the low and high-frequency approximations of impedance. We verify the fractional order of these systems, namely the half-order nature of dynamic phenomenon.

Having these ideas in mind this paper is organized as follows. Section 2 summarizes the mathematical description of the *SE*. Section 3 re-evaluates the *SE* demonstrating its fractional-order dynamics. After clarifying the fundamental concepts, section 4 addresses the case of eddy (or Foucault) currents that occur in electrical machines. Finally, section 5 draws the main conclusions.

2 The Skin Effect

In the differential form the Maxwell equations are [2]:

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \quad (1a)$$

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \quad (1b)$$

$$\nabla \cdot \mathbf{D} = \rho \quad (1c)$$

$$\nabla \cdot \mathbf{B} = 0 \quad (1d)$$

where $\mathbf{E}$, $\mathbf{D}$, $\mathbf{H}$, $\mathbf{B}$, and $\mathbf{J}$ are the vectors of electric field intensity, electric flux density (or electric displacement), magnetic field intensity, magnetic flux density and the current density, respectively, and ρ and t are the charge density and time. Moreover, for a homogeneous, linear, and isotropic media, we can establish the relationships:

$$\mathbf{D} = \varepsilon \mathbf{E} \quad (2a)$$

$$\mathbf{B} = \mu \mathbf{H} \quad (2b)$$

$$\mathbf{J} = \sigma \mathbf{E} \quad (2c)$$

where ε , μ , and σ are the electrical permittivity, the magnetic permeability and the conductivity, respectively.

In order to study the *SE* we start by considering a cylindrical conductor with radius r_0 conducting a current I along its longitudinal axis. In a conductor, even for high frequencies, the term $\partial \mathbf{D} / \partial t$ is negligible in comparison with the conduction term $\mathbf{J}$ or, by other words, the displacement current is much lower than the conduction current. Therefore, for a radial distance $r < r_0$ the application of the Maxwell's equations with the simplification of (1b) leads to the expression [3, 4]:

$$\frac{\partial^2 E}{\partial r^2} + \frac{1}{r} \frac{\partial E}{\partial r} = \sigma \mu \frac{\partial E}{\partial t} \quad (3)$$

For a sinusoidal field we can adopt the complex notation $E = \sqrt{2} \tilde{E} e^{i\omega t}$, where $i = \sqrt{-1}$, yielding:

$$\frac{d^2 \tilde{E}}{dr^2} + \frac{1}{r} \frac{d\tilde{E}}{dr} + q^2 \tilde{E} = 0 \quad (4)$$

with $q^2 = -i\omega\sigma\mu$.

Equation (4) is a particular case of the Bessel equation that, for the case under study, has solution of the type:

$$\tilde{E} = \frac{q}{2\pi r_0 \sigma} \frac{J_0(qr)}{J_1(qr_0)} I, \quad 0 \leq r \leq r_0 \quad (5)$$

where J_0 and J_1 are complex valued Bessel functions of the first kind of orders 0 and 1, respectively.

Equation (5) establishes the so-called *SE* that consists on having a non-uniform current density, namely a low density near the conductor axis and an high density on surface, the higher the frequency ω .

An important measure of the *SE* is the so-called skin depth $\delta = \left(\frac{2}{\omega\mu\sigma}\right)^{1/2}$, corresponding to the distance δ below the conductor surface, for which the field reduces to e^{-1} of its value.

The total voltage drop is $\tilde{Z}\tilde{I} = \tilde{E}\tilde{I}$ that, for a conductor of length l_0 , results:

$$\tilde{Z} = \tilde{E} = \frac{ql_0}{2\pi r_0\sigma} \frac{J_0(qr_0)}{J_1(qr_0)} \quad (6)$$

where $\tilde{Z}$ is the equivalent electrical complex impedance.

Knowing [5] the Taylor series:

$$J_0(x) = 1 - \frac{x^2}{2^2} + \dots, \quad J_1(x) = \frac{x}{2} - \frac{x^3}{2^2 4} + \dots \quad (7)$$

and, for large values of x , the asymptotic expansion:

$$J_n(x) = \sqrt{\frac{2}{\pi x}} \cos\left(x - n\frac{\pi}{2} - \frac{\pi}{4}\right), \quad n = 0, 1, \dots \quad (8)$$

we can obtain the low and high-frequency approximations of $\tilde{Z}$:

$$\omega \rightarrow 0 \Rightarrow \tilde{Z} \approx \frac{l_0}{\pi r_0^2 \sigma} \quad (9a)$$

$$\omega \rightarrow \infty \Rightarrow \tilde{Z} \approx \frac{l_0}{2\pi r_0} \sqrt{\frac{\omega\mu}{2\sigma}} (1 + i) \quad (9b)$$

In the classical *SE* the mean free path l that the electrons can travel between subsequent scattering events is less than the skin depth δ . Therefore, for $\delta \gg l$ we have a local relation and the value of $\mathbf{J}$ at a given point is determined simply by the value of $\mathbf{E}$ at that point. The Ohm's law (2c) is valid, the normal *SE* yields $\delta \sim \omega^{-1/2}$, and the impedance $Z = R + iX$ such that $R = X \sim \omega^{1/2}$.

For very low temperatures the *SE* behaves somewhat differently. In the anomalous skin effect (*ASE*) $\delta \ll l$ the relation between $\mathbf{J}$ and $\mathbf{E}$ is non-local and the electrons are subjected to the field for only a part of its transit time between two collisions between the metal ions [12, 13]. Consequently, it is equivalent to a smaller electron concentration in the skin layer and that causes a poorer conductivity. The anomalous skin depth yields $\delta \sim \omega^{-1/3}$, and the impedance $Z = R + iX$ is such that $R = X/\sqrt{3} \sim \omega^{2/3}$.

In this paper we will focus only on the *SE* but the extension of the proposed methods to the *ASE* is straightforward.

3 The Eddy Currents

The previous physical concepts and mathematical tools can be adopted in more complex systems. The “Eddy Currents” phenomenon common in electrical machines, such as transformers and motors, can be modelled using an identical approach.

Let us consider the magnetic circuit of an electrical machine constituted by a laminated iron core. Each ferromagnetic metal sheet with permeability μ has thickness d and width b ($b \gg d$) making a closed magnetic circuit with an average length l_0 . The total pack of ferromagnetic metal sheet make a height a while embracing a coil having n turns with current I .

The contribution of the magnetic core to the coil impedance is (for details see [3]):

$$\tilde{Z} = \frac{2\mu ab j\omega n^2}{(1+i)\beta Ld} \tanh \left[(1+i)\beta \frac{d}{2} \right] \quad (10)$$

where $\beta = \sqrt{\omega\sigma\mu/2}$.

Alternatively, expression (12) can be re-written as:

$$\tilde{Z} = \frac{\mu ab n^2}{l_0} \omega \cdot \frac{[\sinh(\beta d) - \sin(\beta d)] + i[\sinh(\beta d) + \sin(\beta d)]}{(\beta d)[\cosh(\beta d) + \cos(\beta d)]} \quad (11)$$

We can obtain the low and high-frequency approximations of $\tilde{Z}$:

$$\omega \rightarrow 0 \Rightarrow \tilde{Z} \approx i\omega \frac{\mu ab n^2}{l_0} \quad (12a)$$

$$\omega \rightarrow \infty \Rightarrow \tilde{Z} \approx \frac{\mu ab n^2}{l_0} \frac{1}{d} \sqrt{\frac{2\omega}{\sigma\mu}} (1+i) \quad (12b)$$

Once more we have a clear half-order dependence of $\tilde{Z}$ (i.e., $\tilde{Z} \sim \omega^{1/2}$) while the standard approach is to assign frequency-dependent “equivalent” resistance R and inductance L given by $R + i\omega L = \tilde{Z}$.

4 A Fractional Calculus Perspective

In this section we re-evaluate the expressions obtained for the SE and the Eddy phenomena, in the perspective of fractional calculus.

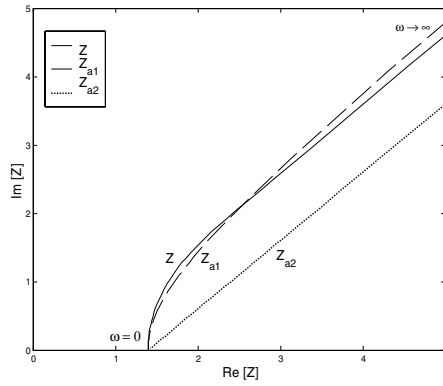
In the SE , to avoid the complexity of the transcendental Eq. (6), the standard approach in electrical engineering is to assign a resistance R and inductance L given by $R + i\omega L = \tilde{Z}$. Nevertheless, although widely used, this method is clearly inadequate because the model parameter values $\{R, L\}$ vary with the frequency. Moreover, (9b) points out the half-order nature of the dynamic phenomenon, at high frequencies (i.e., $\tilde{Z} \sim \omega^{1/2}$), which is not captured by an integer-order approach. A possible approach that eliminates

those problems is to adopt the fractional calculus [6, 7, 8, 9, 10]. Joining the two asymptotic expressions (9) we can establish several types of approximations [11], namely the two expressions:

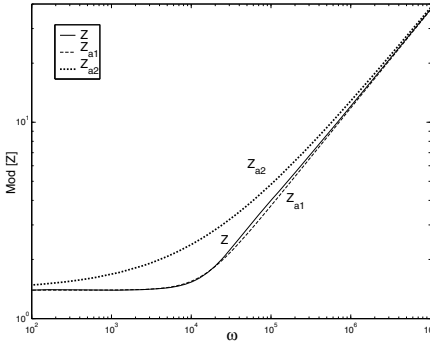
$$\tilde{Z}_{a1} \approx \frac{l_0}{\pi r_0^2 \sigma} \left[i\omega \left(\frac{r_0}{2} \right)^2 \mu \sigma + 1 \right]^{1/2} \quad (13a)$$

$$\tilde{Z}_{a2} \approx \frac{l_0}{\pi r_0^2 \sigma} \left\{ \left[i\omega \left(\frac{r_0}{2} \right)^2 \mu \sigma \right]^{1/2} + 1 \right\} \quad (13b)$$

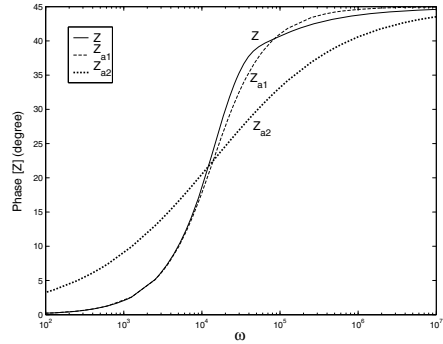
In order to analyse the feasibility of (13) we define the polar, amplitude, and phase relative errors as:



(a)



(b)



(c)

Fig. 1. Diagrams of the theoretical electrical impedance $\tilde{Z}(i\omega)$ and the two approximate expressions $\tilde{Z}_{a1}$, $\tilde{Z}_{a2}$ (10) with: $\sigma = 5.7 \cdot 10^7 \Omega^{-1} \text{ m}$, $l_0 = 10^3 \text{ m}$, $r_0 = 2.0 \cdot 10^{-3} \text{ m}$, $\mu = 1.257 \cdot 10^{-6} \text{ Hm}^{-1}$ (a) Polar, (b) Bode amplitude, and (c) Bode phase.

$$\varepsilon_{Rk}(\omega) = (\tilde{Z} - \tilde{Z}_{ak}) / |\tilde{Z}| \quad (14a)$$

$$\varepsilon_{Mk} = \text{Mod} \{ \varepsilon_{Rk}(\omega) \} \quad (14b)$$

$$\varepsilon_{\phi k} = \text{Phase} \{ \varepsilon_{Rk}(\omega) \} \quad (14c)$$

where the index $k = \{1, 2\}$ represents the two types of approximation.

Figure 1 compares the polar and Bode diagrams of amplitude and phase for expressions (6) and (13) revealing a very good fit in the two cases. On the other hand, Fig. 2 depicts the errors in the charts of polar, amplitude, and phase relative errors, respectively. These figures reveal that the results obtained with the expression (13a) have an better approximation than Eq. (13b), that presents an larger error in the middle of the frequency range.

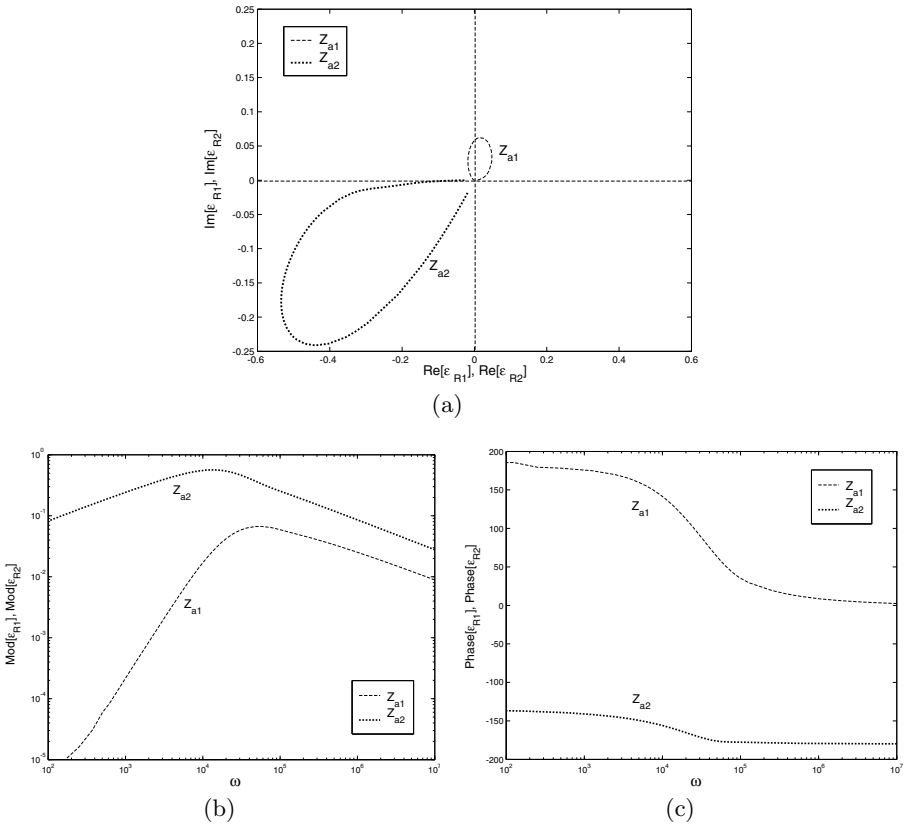


Fig. 2. (a) Polar, (b) amplitude, and (c) phase relative errors for the two approximate expressions $\tilde{Z}_{a1}$ and $\tilde{Z}_{a2}$.

Now we re-evaluate also expressions (10) having in mind the tools of fractional calculus.

A possible approach that avoids the problems posed by the transcendental expression (10) is to joint the two asymptotic expressions (12). Therefore, we can establish several types of approximations, namely the two fractions:

$$\tilde{Z}_{a1} \approx \frac{i\omega \mu ab n^2}{l_0} \left[i\omega \left(\frac{d}{2} \right)^2 \mu \sigma + 1 \right]^{-1/2} \quad (15a)$$

$$\tilde{Z}_{a2} \approx \frac{i\omega \mu ab n^2}{l_0} \left\{ \left[i\omega \left(\frac{d}{2} \right)^2 \mu \sigma \right]^{-1/2} + 1 \right\} \quad (15b)$$

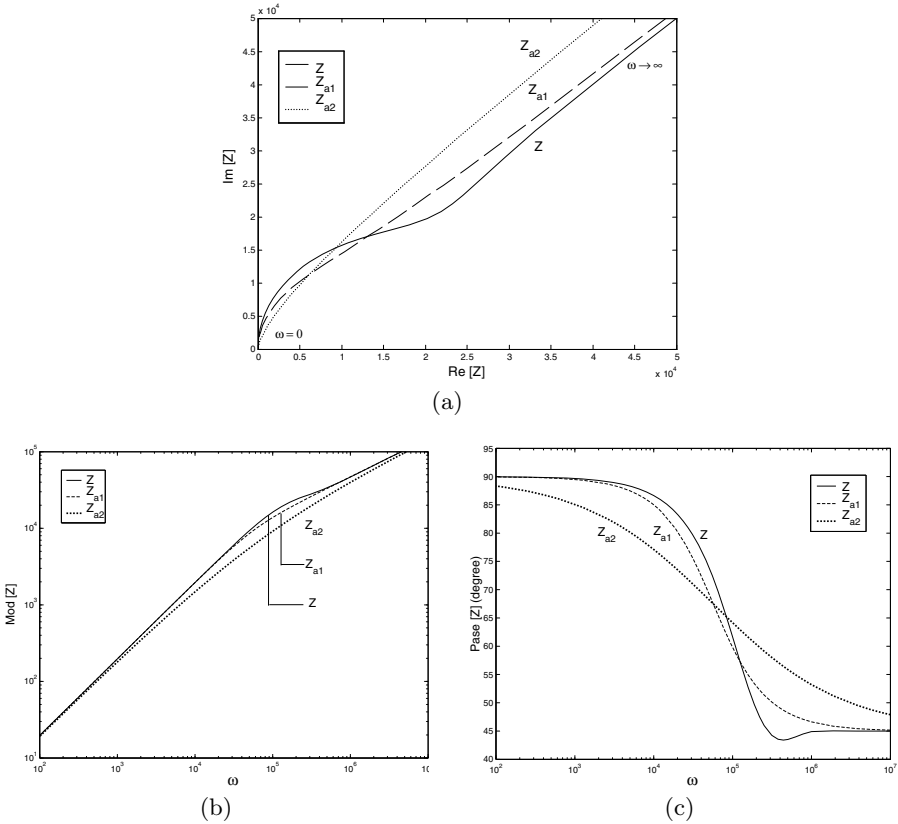


Fig. 3. Diagrams of the theoretical electrical impedance $\tilde{Z}(i\omega)$ and the two approximate expressions $\tilde{Z}_{a1}$ and $\tilde{Z}_{a2}$ (15), with: $l_0 = 1.0$ m, $a = 0.28$ m, $b = 0.28$ m, $d = 2.0 \cdot 10^{-3}$ m, $n = 100$, $\sigma = 7.0 \cdot 10^4 \Omega^{-1}\text{m}$, $\mu = 200 \cdot 1.257 \cdot 10^{-6} \text{ Hm}^{-1}$ (a) polar, (b) Bode amplitude, and (c) Bode phase.

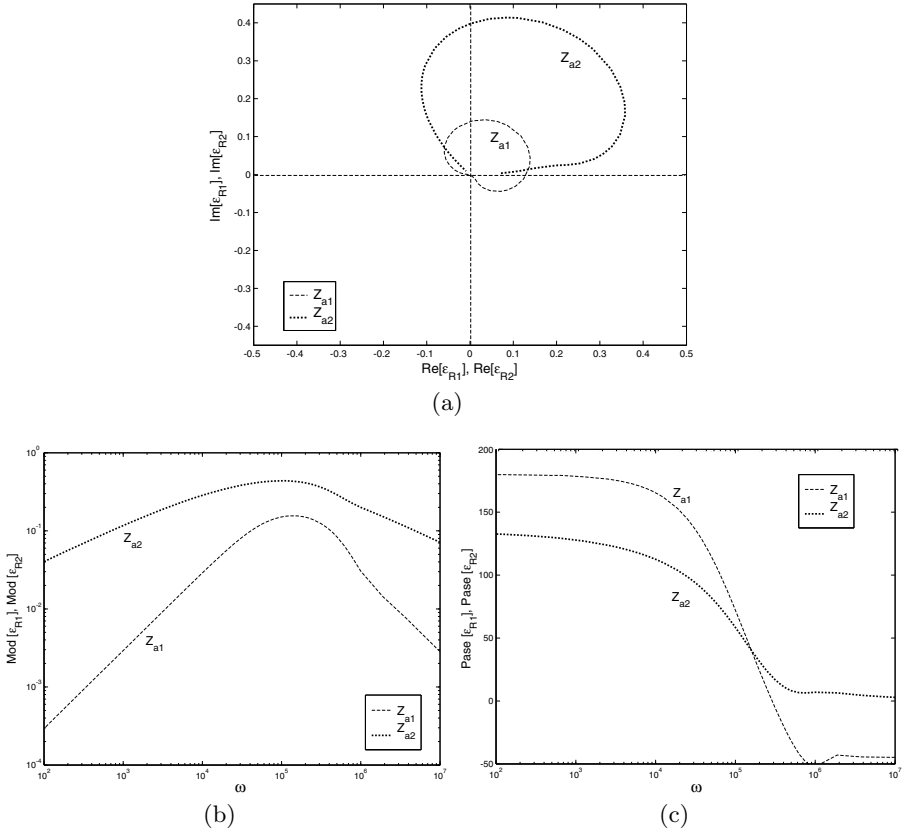


Fig. 4. (a) Polar, (b) amplitude, and (c) phase relative errors for the two approximate expressions $\tilde{Z}_{a1}$ and $\tilde{Z}_{a2}$.

Figure 3 compares the polar and Bode diagrams of amplitude and phase for expressions (10) and (15) revealing a very good fit in the two cases. Figure 4 depicts the relative errors in the charts of polar, amplitude, and phase, respectively. These figures, reveal that the results obtained with expression (15a) have an better approximation, comparatively with Eq. (15b), that presents larger errors in the middle of the frequency range.

5 Conclusions

The classical electromagnetism and the Maxwell equations involve integer-order derivatives, but lead to models requiring a fractional calculus perspective to be fully interpreted. Another aspect of interest is that in all cases we get half-order models. Recent results point out that this is due to the particular

geometry of the addressed problems. Therefore, the analysis of different conductor geometries and its relationship with distinct values of the fractional-order model is under development.

References

1. Chu-Sun Y, Zvonko F, Richard LW (1982) Time - domain Skin - Effect model for Transient Analysis of Lossy Transmission Lines. *Proceedings of the IEEE*, 70(7): 750–757.
2. Richard PF, Robert BL, Matthew S (1964) The Feynman lectures on physics, in: *Mainly Electromagnetism and Matter*. Addison-Wesley, Reading, MA.
3. Küpfmüller KE (1939) *Theoretische Elektrotechnik*. Springer, Berlin.
4. Bessonov L (1968) *Applied Electricity for Engineers*. MIR Publishers, Moscow.
5. Milton A, Irene AS (1965) *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York.
6. Aubourg M, Mengue S (1998) Singularités du Champ Électromagnétique. In: *Proceedings of the Action thématique Les systèmes à dérivées non entières: théorie et applications*. France, 10 June.
7. Sylvain C, Jean F (2003) Fractional Order: Frequential Parametric Identification of the Skin Effect in the Rotor Bar of Squirrel Cage Induction Machine. In: *Proceedings of the ASME 2003 Design Engineering Technical Conference and Computers and Information in Engineering Conference* Chicago, USA, Sept. 2–6.
8. Tenreiro JA, Isabel SJ (2004) A suggestion from the past? FCAA - J. Fract. Calc. Appl. Anal. 7(4).
9. Albert WM, Fernando Silva J, Tenreiro Machado J, Correia de Barros MT (2004) Fractional Order Calculus on the Estimation of Short-Circuit Impedance of Power Transformers. In: *1st IFAC Workshop on Fractional Differentiation and its Application*. France, 19–21 July.
10. Benchellal A, Bachir S, Poinot T, Trigeassou J-C (2004) Identification of a Non-Integer Model of Induction Machines. In: *1st IFAC Workshop on Fractional Differentiation and its Application*. Bordeaux, France, July 19–21.
11. Machado JT, Isabel J, Alexandra G, Albert WM, Fernando S, József KT (2005) Fractional Order Dynamics In Classical Electro-magnetic Phenomena. In: *Fifth EUROMECH Nonlinear Dynamics Conference - ENOC 2005*. Eindhoven, 7–12 August pp. 1322–1326.
12. Sara C, Desy H (2005) Electrodynamics of Superconductors and Superconducting Cavities. In: *6th Scenet school on superconducting materials and applications*. Finland, 18–29 July.
13. Boris P (2003) National Synchrotron Light Source Brookhaven National Lab. In: *Workshop on Superconducting Undulators and Wigglers*. France, 1 July.

IMPLEMENTATION OF FRACTIONAL-ORDER OPERATORS ON FIELD PROGRAMMABLE GATE ARRAYS

Cindy X. Jiang, Joan E. Carletta, and Tom T. Hartley

Department of Electrical & Computer Engineering, The University of Akron, Akron, OH 44325; E-mail: xj2@uakron.edu, carlett@uakron.edu, thartley@uakron.edu

Abstract

Hardware implementation of fractional-order differentiators and integrators requires careful consideration of issues of system quality, hardware cost, and speed. This paper proposes using field programmable gate arrays (FPGAs) to implement fractional-order systems, and demonstrates the advantages that FPGAs provide. The fundamental operator s^a is realized via two different approximations. By applying the binomial expansion, the fractional operator is realized as a high-order finite impulse response (FIR) filter mapped onto a pipelined multiplierless architecture. An IIR approximation is also developed as a parallel combination of first-order filters using the embedded hardware multipliers available on FPGAs. Unlike common fixed-point digital implementations in which all filter coefficients have the same word length, our method quantizes each coefficient using a custom word length chosen in accordance with the filter's sensitivity to perturbations in the coefficient's value. The systems are built based on Xilinx's low-cost Spartan-3 FPGA. They show that the FPGA is an effective platform on which to implement high quality, high throughput approximations to fractional-order systems that are low in cost and require only short design times.

Keywords

Fractional operators, field programmable gate arrays, finite impulse response filters, infinite impulse response filters.

1 Introduction

Fractional derivatives are useful tools for identifying and modeling many dynamic systems. While they have many advantages in the analytical world and much progress has been made in the theory, little has been done to realize them physically. Fractional-order systems are difficult to translate into hardware

because their mathematical properties dictate the use of high-order systems. There are few attempts to develop hardware implementations of fractional-order systems in the literature. Caponetto (2004) proposed a neural network implementation. The resulting integrated circuit design is complicated and restricted to a specific range of integrators, and the system must be trained.

The contribution of our work lies in providing practical and efficient ways to implement fractional-order systems. Recent advances in technology have made digital hardware implementations less expensive, faster, and easier to design. Our techniques exploit the parallel structure and versatility of field programmable gate arrays (FPGAs) in order to yield high-performance and yet low-cost implementations. Our FPGA-based strategies offer simple solutions, and while they are demonstrated on a half-order integrator here, they can be applied to any fractional-order system. This work is a first step towards development of a design flow to overcome the existing barrier between software-based simulations of fractional-order systems and real-time hardware solutions.

In what follows, we discuss the advantages of implementing approximations to fractional-order systems as digital filters using fixed-point mathematics on FPGA. We then discuss approximations to fractional-order derivatives; those based on polynomial functions lead to finite impulse response (FIR) filter implementations, and those based on rational functions lead to infinite impulse response (IIR) implementations. Then, methods for choosing appropriate fixed-point formats for the filter coefficients for both the FIR and IIR approximations are presented. The corresponding FPGA architectures and implementations are described. The performance of a half-order integrator is analysed, and conclusions are drawn.

2 Advantages of FPGA-Based Implementation

Digital hardware designers can choose from a number of different computational platforms when implementing digital signal processing functions of the sort needed to approximate fractional-order systems. Historically, microprocessors and digital signal processors (DSPs) have dominated in low-rate applications for which it is not crucial to save space and power. However, recent advances in technology and in the availability of system-level design tools from vendors have led to a rise in the popularity of FPGAs as a computational platform for digital signal processing applications. FPGAs are general-purpose integrated circuits with tens of thousands of programable logic cells interconnected by wires and programable switches. The main advantage of an FPGA over a microprocessor or DSP is

its versatile, highly parallel structure. Because the programable hardware can be tailored to implement the computation at hand in a maximally parallel way, it can outperform microprocessors and DSPs, which must run the computation serially on general purpose hardware. Modern FPGA system clock rates run in the hundreds of MHz, and an FPGA-based implementation can outperform a DSP-based implementation by a ratio of 100 to 1. Overall, an FPGA has computational power similar to that of an application specific integrated circuit (ASIC), but unlike an ASIC, an FPGA is reconfigurable and has low nonrecurring engineering cost and short design time.

Traditionally, implementation of high-order systems demands the use of floating-point computations. Floating-point mathematics is accurate and easy to work with in the early phases of the design, but make for slow, expensive hardware in production. Fixed-point mathematics is used widely for hardware implementations to save cost and increase speed. The disadvantage of fixed-point mathematics is that the hardware design requires careful consideration of the precision required for each individual application. Thus, the hardware designer needs specific training in fixed-point considerations in order to develop a successful implementation. The ultimate goal of our work is to provide a generalized method for implementation of fractional-order systems that can be used by control engineers without in-depth hardware training, and that exploits the unique ability of FPGAs to customize fixed-point precisions for individual computations to save hardware while preserving the quality of the implementation.

3 Approximating Fractional-Order Operators

The fractional differintegral operator s^α (for real α) is the fundamental building block of any fractional-order system. The transfer function of a fractional-order system is often approximated using two different kinds of representations: polynomial functions, which lead to FIR filter implementations and rational functions, which lead to IIR filter implementations. These are described in detail, next, and the specific approximations used for our implementations are given.

3.1 Polynomial approximations

Polynomial functions are normally generated from direct discrete-time approximations to the fractional-order system; for example, power series expansions or binomial expansions can be used. A binomial expansion based on the Grunwald–Letnikov definition is the most useful one among all direct representations (Podlubny, 1999). For our implementation, we use the Grunwald-Letnikov definition,

$${}_a D_t^\alpha f(t) \equiv \lim_{h \rightarrow 0} \frac{\nabla_h^\alpha f(t)}{h^\alpha}. \quad (1)$$

We have

$$D^\alpha \equiv \lim_{N \rightarrow \infty} T^{-\alpha} \sum_{j=0}^N (-1)^j \binom{\alpha}{j} z^{-j} \quad (2)$$

where $T = 1/h$ is the sampling time. This expression is a binomial series expansion of a backward difference, and has an infinite number of terms. FIR approximations can be derived by choosing a number of terms N to implement; the more terms an approximation uses, the more accurately it will represent the original operator.

3.2 Rational approximations

The other common representation for the transfer function of a fractional-order system is the rational function. Rational function approximations have been widely used for simulating fractional-order systems. Rational function approximations can be produced using indirect approaches based on system theory; in these approaches, s -domain approximations are obtained first and then discretized into the z -domain. Hartley et al. (1996) surveys related approximation techniques. In the 1960s, important pioneering work such as the work of Carlson and Halijak (1964) was done on the use of RC ladder networks for demonstrating and simulating fractional-order systems. Oldham (1974) also developed a set of analog approximations based on RC sections. Today, many researchers continue to extend the RC ladder concept, developing systems based on a number of first-order filters with RC-time constants broadly distributed over a very large spectral domain. Other approximation methods that result in rational functions are available, including continued fraction expansion (CFE), Pade approximation and least square (LS) approximation of the system response. A comparison of those methods

methods can be found in the work of Barbosa et al. (2004). A rational function approximation requires implementing an IIR filter.

In the frequency domain, a spectral function $H(s)$ of any slope can be fitted by a series of piecewise functions of $\pm 20\text{dB}$ slope and zero slope over a specified range (ω_l, ω_h). The places at which the fitting curve changes direction are the locations of poles and zeros. This process generates a high-order rational representation to $H(s)$:

$$H(s) \approx \frac{K \prod_i^N (s + z_i)}{\prod_i^N (s + p_i)} = \sum_{i=1}^N \frac{r_i}{s + p_i} \quad (3)$$

There are different ways to place the poles and zeros. One well-known method is presented by Oustaloup (1991). We choose to use the one developed by Charef et al. (1992) that provides more direct control over the resulting accuracy of the approximation. Poles p_i and zeros z_i are placed at regularly spaced intervals of frequency on a logarithmic scale, so as to achieve no more than a maximum allowable error ε in the magnitude of the frequency response. The first pole is placed at the desired lower frequency for the system bandwidth; here, we choose 10^{-4} rads/s. Other pole and zero locations are determined recursively, as functions of the specified error in the approximation and the order α of the fractional operator s^α :

$$z_i = 10^{\frac{\varepsilon}{10(1-\alpha)}} p_i, \quad p_{i+1} = 10^{\frac{\varepsilon}{10\alpha}} z_i \quad (4)$$

Once the locations of the poles and zeros have been determined, a partial fraction expansion can be used to derive the rational expression in (3), where the $\{r_i\}$ are the residues of the poles $\{p_i\}$. Equation (3) represents a parallel combination of first-order low-pass filters, where r_i and p_i are the coefficients for the i th section. The system must now be discretized for implementation on digital hardware; one possibility is to use the backward Euler method, resulting in the discrete-time transfer function

$$H(z) = \sum_{i=1}^N \frac{r_i T}{1 + p_i T - z^{-1}} = \sum_{i=1}^N \frac{b_i}{1 - a_i z^{-1}} \quad (5)$$

Table 1. Coefficients for the eight first-order sections of the IIR filter before fixed-point quantization

i	b_i	a_i
1	0.005038273	0.284747249
2	0.003066249	0.863193111
3	0.000872916	0.990099010
4	0.000221154	0.999369441
5	0.000055584	0.999960191
6	0.000013973	0.999997488
7	0.000003552	0.999999842
8	0.000001116	0.999999990

As an example, we implement the half-order integrator $s^{-0.5}$ over a frequency range from 10^{-4} to 10^4 radians per second. The method requires eight poles to ensure the desired accuracy. The resulting coefficients are given in Table 1.

3.3 Comparison of resulting FIR and IIR filter approximations

Figure 1 compares the frequency responses of the original half-order integrator with floating-point versions of the FIR and IIR filter approximations. Part (a) of the Figure shows the frequency response of $s^{-0.5}$, along with the frequency responses of three FIR approximations based on Eq. (2), with orders $N = 64, 128$, and $1,024$, respectively. Note that all of the FIR approximations diverge from the ideal case near DC and near the Nyquist frequency; this is to be expected. The higher-order approximations have smaller ripples, and approximate the half-order integrator more closely through a larger range of frequencies. For the chosen sampling time of $T = 0.0001$ s, the upper frequency limit in the bandwidth is about 5 KHz for all orders of approximation. The lower frequency limit is calculated as:

$$f_{\min} = \frac{1}{2\pi} \left(\frac{\Gamma(N - \alpha + 1)}{\Gamma(N + 1)} \right)^{\frac{1}{\alpha}} \left(\frac{T^{-\alpha}}{\Gamma(1 - \alpha)} \right)^{\frac{1}{\alpha}}. \quad (6)$$

The approximation with $N = 1024$ has a decade more frequency bandwidth near DC than the $N = 128$ approximation. While higher-order FIR filters yield approximations that are useful for wider frequency ranges, they also require more resources when implemented in parallel hardware, or more processing time when implemented using serial computations on a microprocessor. While tools exist for automatic synthesis for hardware of FIR filters, most commercially

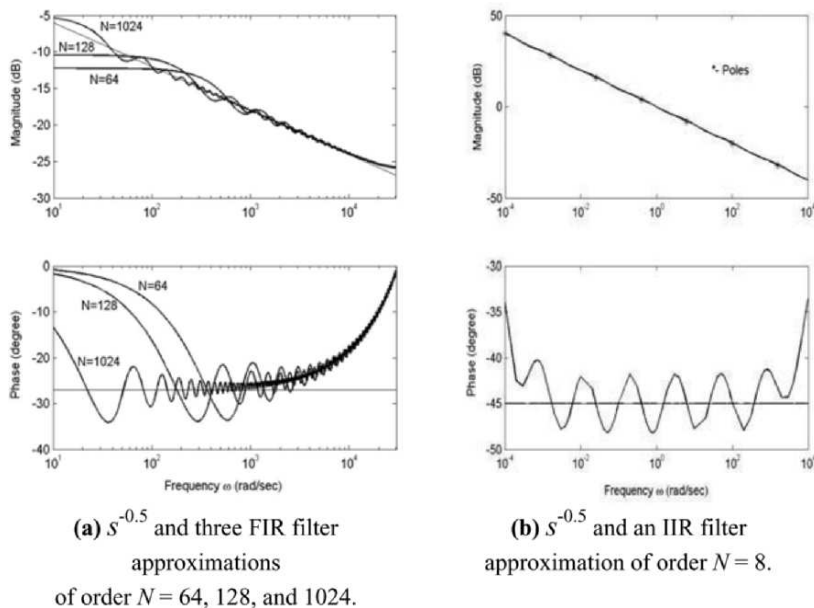


Fig. 1. The frequency responses of a half-order integrator and its digital filter approximations using floating-point mathematics.

available tools place limits on the number of taps for nonsymmetric filters. Orders higher than 512 are rarely seen in practice. In addition, the bandwidth grows very slowly with increased order beyond $N = 512$. One advantage of our particular implementation strategy, described later, is that it makes it possible to implement higher-order filters than can be produced by commercial tools.

The frequency response of a floating-point IIR filter approximation to the half-order integrator is shown in Fig. 1(b). As expected, the IIR approximation is able to achieve much high bandwidth despite its lower order. The IIR approach also achieves a predetermined degree of accuracy. However, several poles in the IIR implementation are very close to the unit circle in the z -domain, and very close to each other. Such IIR filters are very sensitive to computational errors; in fact, it is impossible to simulate this filter in a cascade or direct form. Using partial fraction expansion yields a parallel set of first-order filters, which allows us to guarantee system stability regardless of the order of the implementation; it also saves a great deal of hardware. Alternatively, we can choose to use the FIR implementation in order to avoid the stability and limit cycle oscillations issues inherent in many IIR filters; by taking advantage of an

FPGA's computational power and parallel architecture, a fast and inexpensive FIR implementation can be produced despite the high order. Whether we choose to use an FIR or IIR implementation, the result is an implementation with throughput orders of magnitude better than that achievable using traditional microprocessors or digital signal processors.

4 Implementation in Fixed-Point Mathematics

A filter is designed with infinite precision coefficients $\{b_k\}$ that must be quantized to fixed-point values $\{\bar{b}_k\}$ for hardware implementation; the error in the k th coefficient introduced by quantization is $\Delta b_k = b_k - \bar{b}_k$. The variation in a coefficient's value caused by quantization perturbs the poles and/or zeros of the filter. In traditional hardware design, all filter coefficients are represented using the same number of data bits. This may use more hardware than necessary, especially in the case of high-order systems for which the coefficients span a wide range of values. Our technique, made possible by our use of FPGAs as the computational platform for our implementation, uses a custom word length for each coefficient based on sensitivity of the filter to perturbations in the coefficient's value. We next apply the technique to FIR and IIR filters.

4.1 FIR filters

An FIR filter has the transfer function

$$H(z) = 1 + \sum_{k=1}^{N-1} b_k z^{-k} = \prod_{k=1}^N (1 - r_k z^{-1}), \quad (7)$$

where $\{r_k\}$ are the zeros of $H(z)$. Coefficient quantization causes perturbation of the zeros of the filter in a well-quantified way; if $\{\bar{r}_i\}$ are the zeros of the quantized version of the filter, then the perturbation of the i th zero $\Delta r_i = r_i - \bar{r}_i$ is related to the error in the coefficients as described in [9]:

$$\Delta r_i = \sum_{k=1}^N \frac{r_i^{N-k}}{\prod_{j=1, j \neq i}^N (r_i - r_j)} \Delta b_k. \quad (8)$$

To ensure that the frequency response of our fixed-point filter is not much different from the infinite precision original, we allow no more than a perturbation

of ε in any zero relative to its distance to the origin; for our example, we choose $\varepsilon = 0.05$, so that no zero may move more than 5%. To do this requires that we choose fixed-point values $\{\overline{b_k}\}$ such that

$$\sum_{k=1}^N r_i^{N-k} \Delta b_k \leq \varepsilon |r_i| \prod_{j=1, j \neq i}^N (r_i - r_j). \quad (9)$$

We search to find the minimum precision quantized filter coefficients that will satisfy the relationship in (9). Using as low a precision as is suitable for the application helps in two respects. First, shorter word lengths require less hardware, and thus result in a lower cost implementation. Second, shorter word lengths imply faster computations and higher throughputs.

4.2 IIR filter

We form our IIR filter by using the parallel combination of first-order sections in (5); each section takes the form shown in Fig. 2. The implementation of i th first-order section requires the following calculations using the state variable v_i :

$$\begin{aligned} v_{i,k} &= u_k + a_i * v_{i,k-1} \\ y_{i,k} &= b_i * v_{i,k} \end{aligned} \quad (10)$$

The advantage of this structure is that quantization of a coefficient affects only the independent contribution of the corresponding first-order section. Thus, we can evaluate the quantization effects one section at a time.

Implementation of a first-order section requires fixed-point quantization of coefficients a_i and b_i and of the state variable v_i . Quantization of a_i and b_i causes an error $\Delta H_i(z)$ in the response of the first-order section:

$$\Delta H_i(z) = \frac{\partial H_i(z)}{\partial a_i} \Delta a_i + \frac{\partial H_i(z)}{\partial b_i} \Delta b_i; \quad (11)$$

the maximum absolute error is experienced at dc, so that

$$\|\Delta H_i\|_{\infty} = \left[\frac{zb_i}{(z-a_i)^2} \Delta a_i + \frac{z}{z-a_i} \Delta b_i \right]_{z=1}. \quad (12)$$

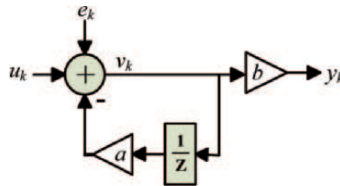


Fig. 2. A direct form II structure for a first-order filter section.

Table 2. Coefficients for the eight first-order sections of the IIR filter after fixed-point quantization, and their fixed-point formats

Quantized b_i	Format of b_i	Quantized a_i	Format of a_i	Format of state variable v_i
0.0050489	(7, -13)	0.25000000	(3, -3)	(19, -15)
0.0030518	(7, -14)	0.87500000	(6, -5)	(19, -13)
0.0008698	(7, -16)	0.99023438	(11, -10)	(21, -11)
0.0002213	(7, -18)	0.99938965	(15, -14)	(23, -9)
0.0000553	(7, -20)	0.99996948	(16, -15)	(25, -7)
0.0000141	(7, -22)	0.99999750	(26, -25)	(28, -5)
0.0000036	(7, -24)	0.99999984	(29, -28)	(30, -3)
0.0000011	(7, -25)	0.99999999	(29, -28)	(32, -1)

The maximum absolute error in the magnitude of the frequency response of the overall IIR filter $H(z)$ is conservatively bound by the sum of the maximum absolute errors in the first-order sections:

$$\|\Delta H\|_{\infty} \leq \sum_N \|\Delta H_i\|_{\infty}. \tag{13}$$

We constrain our fixed-point filter to have a maximum absolute error in the magnitude of the frequency response of 0.5 dB with respect to the floating-point version. The Δa_i and Δb_i are chosen independently for each first-order section, according to Eq. (12), such that the eight sections each contribute a maximum absolute error of no more than about 0.06 dB. Table 2 shows the resulting quantized versions of the coefficients, along with their fixed-point formats, and the chosen fixed-point formats of the state variables. A fixed-point format of (n, f) indicates that there are total of n bits in the coefficient with the least significant bit positioned with weight 2^f , i.e., with f bits to the right of the binary point.

The use of FPGA as the computational platform is what makes it possible to customize the precisions of individual computations. Filter sections with extremely sensitive poles, such as the seventh and eighth sections in our example, require wide fixed-point precisions in order to maintain system accuracy; other sections, such as the first section in our example, are not sensitive, and can be implemented more coarsely with fewer bits.

5 Hardware Implementation

The FIR and IIR filter approximations to the half-order integrator are implemented in hardware on a Spartan-3 xc3s400 FPGA device. The hardware platform also includes a 16-bit 150 Ksamples/sec serial analog-to-digital converter (ADC) (LTC1865L) which allows direct connections to analog input signals, and a 14-bit serial digital-to-analog converter with an 8 μ s conversion time (LTC1654L). State machines to control both converters are implemented on the FPGA.

The limited number of hardware multipliers on FPGA devices is a bottleneck when implementing computations with a large number of multiplications, such as high-order FIR filters. Today’s FPGAs are much cheaper and faster than those of previous generations, and are built with more dedicated hardware multipliers than ever before. However, even the newest devices do not have enough hardware multipliers to do the computations for a high-order FIR filter quickly. In our FIR implementation, the key technique for high-speed, low-cost implementations is to replace multiplication by a constant (in our case, a filter coefficient) with faster, highly pipelined shifts and additions. Only the “1”s in a

Table 3. Characteristics of the hardware implementations of FIR and IIR filter approximations to a half-order integrator

	FIR ($N = 128$)	FIR ($N = 1024$)	IIR ($N = 8$)
Frequency range (rad/s)	10^2 to 3×10^4	10 to 3×10^4	10^{-4} to 10^4
Size (occupied logic slices)	2,102	8,223	593
Throughput (Msamples/s)	50	47	36
Latency (clock cycles)	12	16	1
Maximum absolute error in magnitude of frequency response compared to ideal half-order integrator (dB)	3.11	2.36	0.853

coefficient's binary representation result in nonzero partial products in the multiplication. Therefore, no hardware is needed to process the zero partial products. The wider the range in the values of the filter coefficients, the more hardware will be saved using this technique. The details of this multiplierless filter architecture are presented in [10]. Design parameters such as the FIR coefficient values and precisions are taken as input by a C program developed by our research group; the program generates a hardware description of the filter in VHDL, an industry-standard hardware description language, and then is synthesized for the Spartan-3 xc3s400 device using the Xilinx ISE 7.1i toolset. Our IIR filter approximation has much lower order, and can be implemented in a straightforward way, using the 18-bit by 18-bit hardware multipliers embedded on the FPGA.

Hardware characteristics of the FIR and IIR filter implementations approximating the half-order integrator are given in Table 3. The FIR filter approximation processes one sample per clock cycle, and after an initial latency due to pipelining of 12 clock cycles produces one result per clock cycle; thus, its 150 MHz clock rate corresponds to a throughput of 150 Msamples/s. The IIR filter approximation uses no pipelining, and therefore has a latency of one clock cycle, but with a slower clock rate of 31 MHz.

Figure 3 shows the time domain response of the FIR and IIR filter approximations for an input step of 1V. Both plots also show the ideal response of a half-order integrator, i.e., $t^{0.5}/\Gamma(1.5)$. The IIR filter approximation tracks the ideal response well for the 10 s of time shown. The FIR filter approximation tracks well 0.0128 seconds; after that, its output saturates. This is a direct

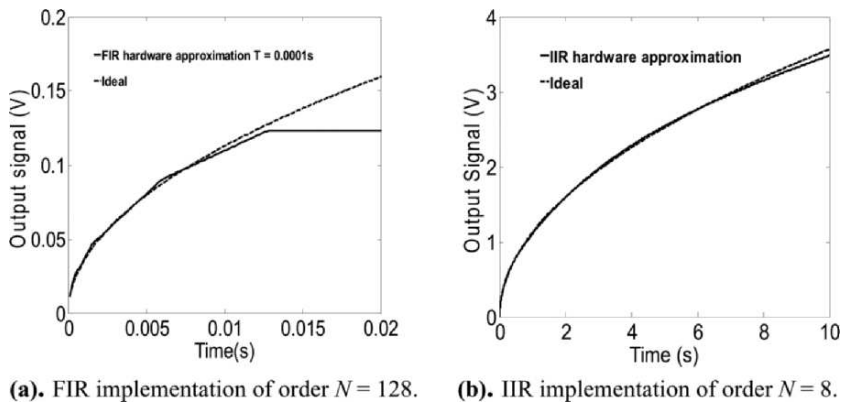


Fig. 3. The outputs of the implemented fixed-point approximations to the half-order integrator in response to an input step of 1V.

consequence of its having order $N = 128$ and a sample time of 0.0001 s; it tracks for N samples. If we want an FIR filter approximation to track longer, either a higher-order approximation or a larger sample time must be used. (For example, choosing $T = 0.1$ s allows an FIR filter with order $N = 128$ to track for 10 s.)

The results show that the IIR filter approximation has a distinct advantage in that it is able to achieve a much wider bandwidth with less hardware; it uses only 593 logic slices, while the FIR filter uses 2,102. The Table also shows that the IIR filter approximation achieves a frequency response closer to that of the ideal half-order integrator over the frequencies for which the system is designed; its maximum absolute error in the magnitude of its frequency response is 0.85 dB, while for the FIR filter approximation the error is 3.11 dB.

6 Conclusions

This paper provides two options for hardware realizations of fractional systems. Both implementations here rely on the special characteristics of FPGAs. Unlike traditional microprocessor-based designs where instructions are executed in a serial manner, an FPGA can execute many operations concurrently. Thus, FPGA-based implementations can be high in throughput. Our work shows both FIR and IIR filter approximations to fractional-order systems, with IIR filter approximations having a distinct advantage in terms of system bandwidth. Our implementations take advantage of the fact that FPGAs do not have a fixed data bus width; any data width can be used, including widths that vary from computation to computation. This allows us to customize the precision of each computation in a filter based on the sensitivity of the system to that computation.

FPGAs are effective platforms for the implementation of fractional-order systems. There are many possibilities for future work to improve the quality of FIR filter realizations of fractional-order systems, such as considerations of different filter structures such as cascade forms. Newer high-end FPGAs include digital signal processing building block that may be used for the implementation. For our IIR realizations, pipelining the current design can boost the throughput of the system.

References

1. Podlubny I (1999) *Fractional Differential Equations*. San Diego, Academic Press.
2. Hartley TT, Lorenzo CF, Qammar HK (1996) Chaos in a Fractional Order Chua System, NASA Technical Paper 3543.
3. Carlson GE, Halijak CA (1964) Approximation of fractional capacitors $(1/s)^{1/n}$ by a regular Newton process, IEEE Trans Circ Theory, CT-11:210–213.
4. Oldham KB, Spanier J (1974) *The Fractional Calculus*. San Diego, Academic Press.
5. Barbosa RS, Machado JA, Ferreira IM (2004) Least Squares Design of Digital Fractional-Order Operators, First International Federation of Automatic Control Workshop on Fractional Differentiation and Its Applications, Bordeaux, France, July, pp. 436–441.
6. Caponetto R, Fortuna L, Porto D (2004) Hardware Design of a Multi Layer Perceptron for Non Integer Order Integration, First International Federation of Automatic Control Workshop on Fractional Differentiation and Its Applications, Bordeaux, France, July, pp. 248–253.
7. Oustaloup A (1991) *La Commande CRONE: Commande Robuste d'Ordre Non Entier*, HERMES, Paris.
8. Charef A, Sun HH, Tsao YY, Onaral YB (1992) Fractal system as represented by singularity function, IEEE Trans. Auto. Control, 37(9): 1465–1470.
9. Proakis JG, Manolakis DG (1996) *Digital Signal Processing: Principles, Algorithms, and Applications*. New Jersey, Prentice-Hall.
10. Carletta JE, Rayman MD (2002) Practical considerations in the synthesis of high performance digital filters for implementation on FPGAs, *Field Programmable Logic and Applications*, in: *Lecture Notes in Computer Science*, vol. 2438, Springer, pp. 886–896.
11. Carletta JE, Veillette RJ, Krach F, Fang Z (2003) Determining Appropriate Precisions for Signals in Fixed-Point IIR Filters, Proceedings of the IEEE/ACM Design Automation Conference, Anaheim, CA, June 2–6, pp. 656–661.
12. Jiang CX, Hartley TT, Carletta JE (2005) High Performance Low Cost Implementation of FPGA-based Fractional-Order Operators, Proceedings of ASME Design Engineering Technical Conferences and Computer and Information in Engineering Conference, Long Beach, CA, September.

COMPLEX ORDER-DISTRIBUTIONS USING CONJUGATED ORDER DIFFERINTEGRALS

Jay L. Adams¹, Tom T. Hartley², and Carl F. Lorenzo³

¹ Department of Electrical and Computer Engineering, The University of Akron, Akron, OH 44325-3904; E-mail: JLA36@uakron.edu

² Department of Electrical and Computer Engineering, The University of Akron, Akron, OH 44325-3904; E-mail: TomHartley@aol.com

³ NASA Glenn Research Center, Cleveland, OH 44135;
E-mail: Carl.F.Lorenzo@grc.nasa.gov

Abstract

This paper develops the concept of the complex order-distribution. This is a continuum of fractional differintegrals of complex order. Two types of complex order-distributions are considered, uniformly distributed and Gaussian distributed. It is shown that these basis distributions can be summed to approximate other complex order-distributions. Conjugated differintegrals, introduced in this paper, are an essential analytical tool applied in this development. Conjugated-order differintegrals are fractional derivatives whose orders are complex conjugates. These conjugate-order differintegrals allow the use of complex-order differintegrals while still resulting in real time-responses and real transfer-functions. An example is presented to demonstrate the complex order-distribution concept. This work enables the generalization of fractional system identification to allow the search for complex order-derivatives that may better describe real-time behaviors.

Keywords

Fractional-order systems, fractional calculus, conjugated-order differintegrals, complex order-distributions. complex-order differintegrals.

1 Introduction

This paper uses the concept of conjugate-order differintegrals for the development of complex order-distributions. Order distributions have been introduced by Hartley and Lorenzo [1,2] as the continuum extension of collections of fractional-order operators. In that discussion, the distribution of order was required implicitly to be real, but it was able to include any real number. This concept of an order-distribution is expanded to include distributions which have non-real portions, i.e., complex order-distributions. This is done to expand on the system identification technique that used real order-distributions [1]

to include the possibility of using complex order-distributions. To ensure that only real time-responses are considered, the idea of conjugate-order differintegrals is utilized. Just as conjugate-differintegrals provide real time-responses, so do complex order-distributions which are conjugate-symmetric.

Fractional operators of non-integer, but real, order have been the focus of numerous studies. Complex, or even purely imaginary, operators have been studied by a few [2,3]. A motivation in the development of complex operators is to generalize the idea of derivatives and integrals of distributed order. Very limited work in the area of complex-order differintegrals has been done [5]. While the physical meaning of a complex function of time is still under discussion, a goal of this paper is the development of complex-order differintegrals which yield purely real time-responses. To this end, the concept of conjugate-differintegral is introduced.

Both blockwise constant and Gaussian complex order-distributions are presented in the Laplace domain. Approximate complex order-distributions with either the blockwise constant or Gaussian distributions are shown. Finally, the frequency response of a conjugate-symmetric complex order-distribution is compared to that of impulsive distributions in an example.

2 Complex Differintegrals

In general, we will consider the complex differintegral acting on a function $f(t)$ to be defined as

$$g(t) = {}_0d_t^q f(t) = {}_0d_t^{u+iv} f(t). \quad (1)$$

Following the work of Kober [3], Love [4], and Oustaloup et al. [5], this uninitialized operator will have the Laplace transform

$$L\{g(t)\} = G(s) = s^{u+iv} F(s) = s^u s^{iv} F(s) = s^u e^{iv \ln(s)} F(s). \quad (2)$$

Using Euler's identity, this can be rewritten as

$$G(s) = s^u [\cos(v \ln(s)) + i \sin(v \ln(s))] F(s). \quad (3)$$

To obtain the impulse response of this operator, the inverse Laplace transform is required. It is defined for $q < 0$ as

$$L^{-1}\{s^{-q}\} = \frac{t^{q-1}}{\Gamma(q)} \quad (4)$$

For our specific case it becomes, with an impulsive input $g(t)$,

$$f(t) = L^{-1} \left\{ s^{-(u+iv)} \right\} = \frac{t^{u+iv-1}}{\Gamma(u+iv)}, \quad (5)$$

and u and v such that the transform is defined. This can be rewritten as

$$f(t) = L^{-1} \left\{ s^{-(u+iv)} \right\} = \frac{t^{u-1}}{\Gamma(u+iv)} t^{iv} = \frac{t^{u-1}}{\Gamma(u+iv)} e^{iv \ln(t)} \quad (6)$$

or by using Euler's identity as

$$f(t) = L^{-1} \left\{ s^{-(u+iv)} \right\} = \frac{t^{u-1}}{\Gamma(u+iv)} (\cos(v \ln(t)) + i \sin(v \ln(t))). \quad (7)$$

Imaginary time responses have limited physical meaning. However, the functions $\cos(v \ln(t))$ and $\sin(v \ln(t))$ show up regularly as solutions of special time-varying differential equations known as Cauchy–Euler differential equations.

3 Conjugated-Order Differintegrals

The interpretations and inferences of individual complex-order operators are not well understood. However, we can create useful operators by considering the complex-order derivative or integral analogously to a complex eigenvalue of a dynamic system, that is, coexisting with its complex-order conjugate. We now define the uninitialized conjugated differintegral as

$$g(t) = {}_0 d_t^{q(u,v)} f(t) \equiv {}_0 d_t^q f(t) + {}_0 d_t^q \bar{f}(t) = {}_0 d_t^{u+iv} f(t) + {}_0 d_t^{u-iv} f(t). \quad (8)$$

Representing this in the Laplace domain gives

$$L\{g(t)\} = L\left\{{}_0 d_t^{q(u,v)} f(t)\right\} = (s^{u+iv} + s^{u-iv})F(s) = (s^u s^{iv} + s^u s^{-iv})F(s). \quad (9)$$

Rearranging and applying Euler's identity allows this to be written as

$$\begin{aligned} G(s) &= L\left\{{}_0 d_t^{q(u,v)} f(t)\right\} = s^u (e^{iv \ln(s)} + e^{-iv \ln(s)}) F(s) \\ &= s^u ((\cos(v \ln(s)) + i \sin(v \ln(s))) + (\cos(v \ln(s)) - i \sin(v \ln(s)))) F(s) \\ &= 2s^u \cos(v \ln(s)) F(s), \end{aligned} \quad (10)$$

which is a purely real operator. Likewise, the complementary conjugated differintegral is defined as

$$\bar{g}(t) = {}_0 d_t^{q(\bar{u},\bar{v})} f(t) \equiv {}_0 d_t^q f(t) - {}_0 d_t^{\bar{q}} f(t) = {}_0 d_t^{u+iv} f(t) - {}_0 d_t^{u-iv} f(t). \quad (11)$$

Representing this in the Laplace domain gives

$$L\{\bar{g}(t)\} = L\left\{{}_0d_t^{q(u,\bar{v})}f(t)\right\} = (s^{u+iv} - s^{u-iv})F(s) = (s^u s^{iv} - s^u s^{-iv})F(s). \quad (12)$$

Rearranging and using the Euler identity allows us to write

$$\begin{aligned} \bar{G}(s) &= L\left\{{}_0d_t^{q(u,v)}f(t)\right\} = (s^u e^{iv\ln(s)} - s^u e^{-iv\ln(s)})F(s) \\ &= (s^u (\cos(v\ln(s)) + i\sin(v\ln(s))) - s^u (\cos(v\ln(s)) - i\sin(v\ln(s))))F(s) \\ &= 2is^u \sin(v\ln(s))F(s), \end{aligned} \quad (13)$$

which is a purely imaginary operator.

It should be noted that a multiplicative operation returns a real operator,

$$s^{u+iv}s^{u-iv}F(s) = s^{2u}F(s), \quad (14)$$

while a division will yield the imaginary operator $s^{2iv}F(s)$. We note that a real differintegral can always be broken into the product of two complex conjugate derivatives.

The conjugated-order fractional integral may be expressed for negative real order as

$$g(t) = {}_0d_t^{-q(u,v)}f(t) = {}_0d_t^{-(u+iv)}f(t) + {}_0d_t^{-(u-iv)}f(t), \quad (15)$$

with Laplace transform given by

$$L\{g(t)\} = L\left\{{}_0d_t^{-q(u,v)}f(t)\right\} = (s^{-(u+iv)} + s^{-(u-iv)})F(s) = s^{-u}(s^{-iv} + s^{iv})F(s). \quad (16)$$

For $f(t)$ a unit impulse, the inverse Laplace transform of the conjugated integral can also be obtained using the operator inverse of Eq. (5),

$$g(t) = L^{-1}\{s^{-(u+iv)} + s^{-(u-iv)}\} = \frac{t^{u+iv-1}}{\Gamma(u+iv)} + \frac{t^{u-iv-1}}{\Gamma(u-iv)} \quad (17)$$

The presence of the gamma function of complex argument is somewhat problematic, and to move forward we note that the reciprocal gamma function has symmetry about the real axis [6]. Thus we can write

$$\frac{1}{\Gamma(u+iv)} = \operatorname{Re}\left[\frac{1}{\Gamma(u+iv)}\right] + i\operatorname{Im}\left[\frac{1}{\Gamma(u+iv)}\right] \quad (18)$$

and

$$\frac{1}{\Gamma(u-iv)} = \operatorname{Re}\left[\frac{1}{\Gamma(u+iv)}\right] - i\operatorname{Im}\left[\frac{1}{\Gamma(u+iv)}\right]. \quad (19)$$

The desired inverse Laplace transform can then be written

$$\begin{aligned}
g(t) &= t^{u-1} \left[\begin{aligned} &\text{Re} \left[\frac{1}{\Gamma(u+iv)} \right] t^{-iv} + i \text{Im} \left[\frac{1}{\Gamma(u+iv)} \right] t^{-iv} + \\ &\text{Re} \left[\frac{1}{\Gamma(u+iv)} \right] t^{iv} - i \text{Im} \left[\frac{1}{\Gamma(u+iv)} \right] t^{iv} \end{aligned} \right] \\
&= t^{u-1} \left[\text{Re} \left[\frac{1}{\Gamma(u+iv)} \right] (t^{-iv} + t^{iv}) + i \text{Im} \left[\frac{1}{\Gamma(u+iv)} \right] (t^{-iv} - t^{iv}) \right] \quad (20)
\end{aligned}$$

We can now write $t^{\pm iv} = e^{\pm iv \ln(t)}$ and use Euler's identity to give

$$\begin{aligned}
g(t) &= t^{u-1} \text{Re} \left[\frac{1}{\Gamma(u+iv)} \right] ((\cos(v \ln(t)) - i \sin(v \ln(t))) + (\cos(v \ln(t)) + i \sin(v \ln(t)))) \\
&\quad + i t^{u-1} \text{Im} \left[\frac{1}{\Gamma(u+iv)} \right] ((\cos(v \ln(t)) - i \sin(v \ln(t))) - (\cos(v \ln(t)) + i \sin(v \ln(t))))
\end{aligned}$$

Thus

$$\begin{aligned}
g(t) &= L^{-1} \{ s^{-(u+iv)} + s^{-(u-iv)} \} = L^{-1} \{ 2s^{-u} \cos(v \ln(s)) \} \\
&= 2t^{u-1} \left[\text{Re} \left[\frac{1}{\Gamma(u+iv)} \right] \cos(v \ln(t)) + \text{Im} \left[\frac{1}{\Gamma(u+iv)} \right] \sin(v \ln(t)) \right]. \quad (21)
\end{aligned}$$

When $f(t)$ is not a unit impulse, the time response is given by the convolution of $g(t)$ with $f(t)$. It should be noted then that the conjugated differintegral has a purely real time response.

Similarly, the inverse transform of the complementary conjugated-order derivative of a unit impulse can be found as

$$\begin{aligned}
\bar{g}(t) &= L^{-1} \{ s^{-(u+iv)} - s^{-(u-iv)} \} = L^{-1} \{ i 2s^{-u} \sin(v \ln(s)) \} \\
&= i 2t^{u-1} \left[\text{Re} \left[\frac{1}{\Gamma(u+iv)} \right] \sin(v \ln(t)) + \text{Im} \left[\frac{1}{\Gamma(u+iv)} \right] \cos(v \ln(t)) \right], \quad (22)
\end{aligned}$$

a purely imaginary time response.

The frequency response of a particular conjugated integral is shown in Fig. 1. The magnitude frequency response rolls off (or up) at a mean rate set by u . It has superimposed on it a variation that is periodic in $\log(w)$, the period of which is determined by v . The phase-frequency response also rolls off (or up) at an average linear rate, similar to a delay. It also has superimposed on it a variation that is periodic in $\log(w)$. Frequency responses of this form are said to have scale-invariant frequency responses [7], which are fractal in the frequency domain. The Nyquist plane representation is given in the Fig. 2a. It can be seen to have a spiral form. Finally the Nichols plane representation is given in Fig. 2b. Here the plot is a roughly straight line, having the angle from the horizontal determined by v . Frequency domain functions of this form can be approximated to any accuracy using rational transfer functions over any desired range of frequencies [8]. A frequency response of this form is of great use for

control-system design as it is roughly a straight line in the Nichol's plane, with the angle and roll-off rate easily defined by $u + iv$, respectively. The CRONE (controller) design [5], contains terms similar to those seen here, however, they are not recognized as being related to conjugated-order differintegrals.

3.1 Special conjugate derivative forms

In the introduction of conjugated derivatives the weightings of the complex derivatives were real and unity. However, complex derivatives can also have complex weightings. Such complex coefficients may lead to real time-responses, so it is important to determine the effects of different combinations. The determination of effects is presented here, with the purely real time-responses boxed. There is also a corresponding impulse response.

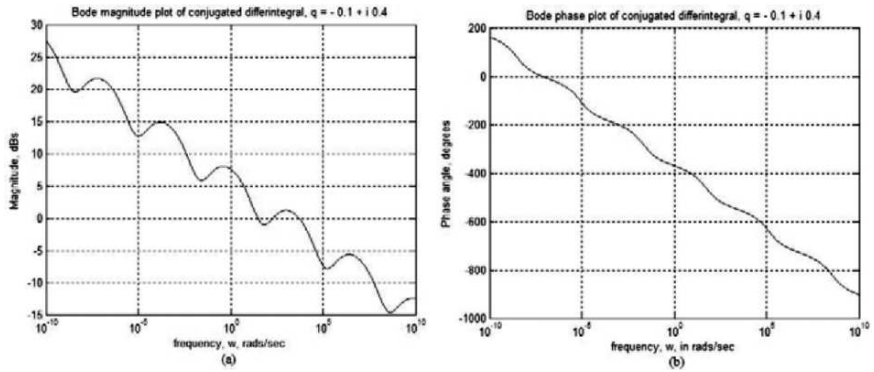


Fig. 1. Bode (a) magnitude and (b) phase plots for $s^{-0.1+i0.4} + s^{-0.1-i0.4}$.

Real coefficients:

$$\begin{aligned} G(s) &= ks^{u+iv} + ks^{u-iv} = ks^u (s^{iv} + s^{-iv}) = ks^u (e^{iv \ln(s)} + s^{-iv \ln(s)}) \\ &= 2ks^u \cos(v \ln(s)) \end{aligned} \quad (23)$$

$$\begin{aligned} G(s) &= ks^{u+iv} - ks^{u-iv} = ks^u (s^{iv} - s^{-iv}) = ks^u (e^{iv \ln(s)} - s^{-iv \ln(s)}) \\ &= i2ks^u \sin(v \ln(s)) \end{aligned} \quad (24)$$

Imaginary coefficients:

$$\begin{aligned} G(s) &= iks^{u+jv} + iks^{u-jv} = iks^u (s^{jv} + s^{-jv}) = iks^u (e^{jv \ln(s)} + s^{-jv \ln(s)}) \\ &= i2ks^u \cos(v \ln(s)) \end{aligned} \quad (25)$$

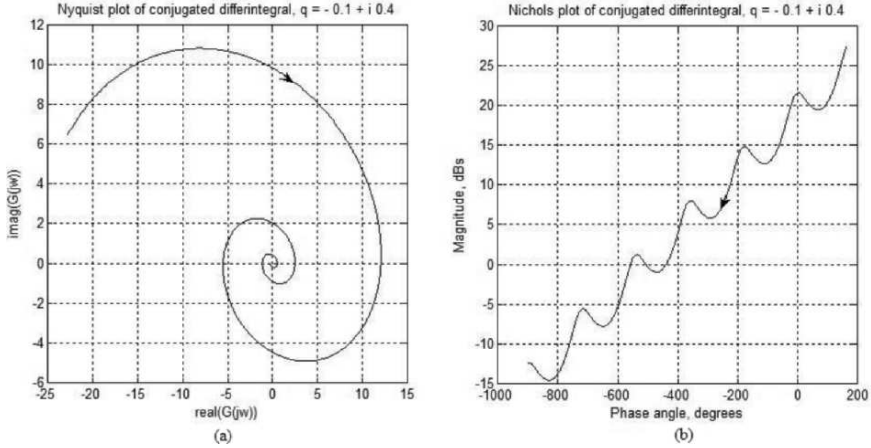


Fig. 2. (a) Nyquist and (b) Nichols Pplots for $s^{-0.1+i0.4} + s^{-0.1-i0.4}$.

$$G(s) = iks^{u+iv} - iks^{u-iv} = iks^u (s^{iv} - s^{-iv}) = iks^u (e^{iv \ln(s)} - s^{-iv(\ln(s))}) = -2ks^u \sin(v \ln(s)) \quad (26)$$

Complex coefficients (4 of the 16 possible):

$$\begin{aligned} G(s) &= (a+ib)s^{u+iv} + (a+ib)s^{u-iv} = a(s^{u+iv} + s^{u-iv}) + ib(s^{u+iv} + s^{u-iv}) \\ &= as^u (e^{iv \ln(s)} + e^{-iv(\ln(s))}) + ibs^u (e^{iv \ln(s)} + e^{-iv(\ln(s))}) \\ &= 2as^u \cos(v \ln(s)) + i2bs^u \cos(v \ln(s)) \end{aligned} \quad (27)$$

$$\begin{aligned} G(s) &= (a+ib)s^{u+iv} + (a-ib)s^{u-iv} = a(s^{u+iv} + s^{u-iv}) + ib(s^{u+iv} - s^{u-iv}) \\ &= as^u (e^{iv \ln(s)} + e^{-iv(\ln(s))}) + ibs^u (e^{iv \ln(s)} - e^{-iv(\ln(s))}) \\ &= 2as^u \cos(v \ln(s)) - 2bs^u \sin(v \ln(s)) \end{aligned} \quad (28)$$

$$\begin{aligned} G(s) &= (a+ib)s^{u+iv} + (-a-ib)s^{u-iv} = a(s^{u+iv} - s^{u-iv}) + ib(s^{u+iv} - s^{u-iv}) \\ &= as^u (e^{iv \ln(s)} - e^{-iv(\ln(s))}) + ibs^u (e^{iv \ln(s)} - e^{-iv(\ln(s))}) \\ &= i2as^u \sin(v \ln(s)) - 2bs^u \sin(v \ln(s)) \end{aligned} \quad (29)$$

$$\begin{aligned}
G(s) &= (a - ib)s^{u+iv} + (-a + ib)s^{u-iv} = a(s^{u+iv} - s^{u-iv}) + ib(-s^{u+iv} + s^{u-iv}) \\
&= as^u(e^{iv \ln(s)} - e^{-iv \ln(s)}) + ibs^u(-e^{iv \ln(s)} + e^{-iv \ln(s)}) \\
&= i2as^u \sin(v \ln(s)) + 2bs^u \sin(v \ln(s))
\end{aligned} \tag{30}$$

4 Complex Order-Distribution Definition

The conjugated derivative will now be applied to the development of complex order-distributions. In previous studies [1,7] the real order-distribution was defined as

$$h(t) \equiv \int_a^b k(q) \left\{ {}_0 d_t^q f(t) \right\} dq, \tag{31}$$

for q real. We will define the complex order-distribution as

$$h(t) \equiv \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} k(u, v) \left\{ {}_0 d_t^{u+iv} f(t) \right\} du dv. \tag{32}$$

This equation can be Laplace transformed as

$$H(s) \equiv \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} k(u, v) s^{u+iv} F(s) du dv. \tag{33}$$

We now must consider two complex planes as in [5]. One is the standard Laplace s -plane, and the other is the complex order-plane, or q -plane, where $q = u + iv$. It is understood that the order of a given operator is not necessarily an impulse in the q -plane as is usually the case for fractional-order differential equations, $k(q) = \delta(q)$. The order will now be considered to be a continuum or distribution in the complex order-plane, a complex generalization of [1]. When the weighting function $k(u, v)$ is complex and it has symmetry about the real order-axis, then the corresponding time response is real.

4.1 Blockwise constant complex order-distribution

We now consider complex order-distributions that are constant intensity, k , symmetric about the real axis from $\bar{u} - \delta u$ to $\bar{u} + \delta u$, and from $-i\delta v$ to $+i\delta v$. A detailed derivation is given by Hartley et al. [9] but the idea and results are presented here.

$$H(s) = \int_{-\delta v}^{\delta v} \int_{\bar{u}-\delta u}^{\bar{u}+\delta u} ks^{u+iv} dudv = \int_{-\delta v}^{\delta v} \int_{-\delta u}^{\delta u} ks^{w+\bar{u}+iv} dw dv, \text{ where } u = w - \bar{u}, du = dw,$$

$$\begin{aligned}
&= ks^{\bar{u}} \int_{-\delta\bar{v}}^{\delta\bar{v}} \int_{-\delta\bar{u}}^{\delta\bar{u}} s^w s^{iv} dw dv = ks^{\bar{u}} \left(\int_{-\delta\bar{u}}^{\delta\bar{u}} e^{w \ln(s)} dw \right) \left(\int_{-\delta\bar{v}}^{\delta\bar{v}} e^{iv \ln(s)} dv \right) \\
&= 2ks^{\bar{u}} \left(\frac{e^{w \ln(s)}}{\ln(s)} \right) \Bigg|_{w=-\delta\bar{u}}^{\delta\bar{u}} \left(\frac{\sin(v \ln(s))}{\ln(s)} \right) \Bigg|_{v=0}^{\delta\bar{v}} = 4ks^{\bar{u}} \frac{\sinh(\delta\bar{u} \ln(s))}{\ln(s)} \frac{\sin(\delta\bar{v} \ln(s))}{\ln(s)}. \quad (34)
\end{aligned}$$

For constant block order-distributions of intensity k which are centered at $q = \bar{u} + i\bar{v}$ (off the real-axis), then, as shown by Hartley et al. [9], is

$$\begin{aligned}
H_1(s) &= \int_{\bar{v}-\delta\bar{v}}^{\bar{v}+\delta\bar{v}} \int_{\bar{u}-\delta\bar{u}}^{\bar{u}+\delta\bar{u}} ks^{u+iv} du dv \\
&= \int_{\bar{u}-\delta\bar{u}}^{\bar{u}+\delta\bar{u}} ks^u du \int_{\bar{v}-\delta\bar{v}}^{\bar{v}+\delta\bar{v}} s^{iv} dv \\
&= \int_{-\delta\bar{u}}^{\delta\bar{u}} ks^{w+\bar{u}} dw \int_{-\delta\bar{v}}^{\delta\bar{v}} s^{i(r+\bar{v})} dr \\
&= ks^{\bar{u}+i\bar{v}} \int_{-\delta\bar{u}}^{\delta\bar{u}} s^w dw \int_{-\delta\bar{v}}^{\delta\bar{v}} s^{ir} dr \\
&= 2ks^{\bar{u}+i\bar{v}} \left(\frac{e^{w \ln(s)}}{\ln(s)} \right) \Bigg|_{w=-\delta\bar{u}}^{\delta\bar{u}} \left(\frac{\sin(r \ln(s))}{\ln(s)} \right) \Bigg|_{r=0}^{\delta\bar{v}} \\
&= \frac{4ks^{\bar{u}+i\bar{v}}}{\ln^2(s)} (\sinh(\delta\bar{u} \ln(s))) (\sin(\delta\bar{v} \ln(s))). \quad (35)
\end{aligned}$$

Similarly, for constant block order-distributions of intensity k which are centered at $q = \bar{u} - i\bar{v}$ [9]

$$\begin{aligned}
\bar{H}_1(s) &= \int_{-\bar{v}-\delta\bar{v}}^{-\bar{v}+\delta\bar{v}} \int_{\bar{u}-\delta\bar{u}}^{\bar{u}+\delta\bar{u}} ks^{u+iv} du dv \\
&= \frac{4ks^{\bar{u}-i\bar{v}}}{\ln^2(s)} (\sinh(\delta\bar{u} \ln(s))) (\sin(\delta\bar{v} \ln(s))). \quad (36)
\end{aligned}$$

Combining these two complex results, Eqs. (35) and (36), give the real conjugated block differintegral as shown below [9].

$$\begin{aligned}
H(s) &= H_1(s) + \bar{H}_1(s) = \frac{4k}{\ln^2(s)} (\sinh(\delta\bar{u} \ln(s))) (\sin(\delta\bar{v} \ln(s))) (s^{\bar{u}+i\bar{v}} + s^{\bar{u}-i\bar{v}}) \\
&= \frac{4k}{\ln^2(s)} (\sinh(\delta\bar{u} \ln(s))) (\sin(\delta\bar{v} \ln(s))) (s^{\bar{u}} s^{i\bar{v}} + s^{\bar{u}} s^{-i\bar{v}}) \\
&= \frac{8ke^{\bar{u} \ln(s)}}{\ln^2(s)} (\sinh(\delta\bar{u} \ln(s))) (\sin(\delta\bar{v} \ln(s))) (\cos(\bar{v} \ln(s))). \quad (37)
\end{aligned}$$

Sums of these order-distributions can be used to approximate complex order-distributions that are symmetric with respect to the real-order axis as follows. Assuming the widths of each block are the same and the intensities are $k_{m,n}$,

$$\begin{aligned}
 H(s) &= \sum_{m=-\infty}^{\infty} \sum_{n=1}^{\infty} \int_{\bar{v}_m - \delta v}^{\bar{v}_m + \delta v} \int_{\bar{u}_n - \delta u}^{\bar{u}_n + \delta u} k_{n,m} s^{u+iv} du dv \\
 &= \sum_{m=-\infty}^{\infty} \sum_{n=1}^{\infty} k_{n,m} s^{\bar{u}_n + i\bar{v}_m} \left(\frac{\sinh(\delta u \ln(s))}{\ln(s)} \right) \left(\frac{\sin(\delta v \ln(s))}{\ln(s)} \right) \\
 &= 4 \left(\frac{\sinh(\delta u \ln(s))}{\ln(s)} \right) \left(\frac{\sin(\delta v \ln(s))}{\ln(s)} \right) \sum_{m=-\infty}^{\infty} \sum_{n=1}^{\infty} k_{n,m} s^{\bar{u}_n + i\bar{v}_m}. \quad (38)
 \end{aligned}$$

4.2 Gaussian complex order-distribution

Finally, we consider complex order-distributions that have the form of Gaussians of intensity k centered on, and symmetric about, the real order-axis [9],

$$\begin{aligned}
 H(s) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} k e^{-\frac{(u-\bar{u})^2}{\sigma_u^2} - \frac{v^2}{\sigma_v^2}} s^{u+iv} du dv \\
 &= \int_{-\infty}^{\infty} k e^{-\frac{(u-\bar{u})^2}{\sigma_u^2}} s^u du \int_{-\infty}^{\infty} e^{-\frac{v^2}{\sigma_v^2}} s^{iv} dv \\
 &= k s^{\bar{u}} \int_{-\infty}^{\infty} e^{-\frac{w^2}{\sigma_u^2}} s^w dw \int_0^{\infty} e^{-\frac{v^2}{\sigma_v^2}} (s^{iv} + s^{-jv}) dv \\
 &= k s^{\bar{u}} \left(-\frac{\sigma_u \sqrt{\pi}}{2} e^{\frac{1}{4}\sigma_u^2 \ln^2(s)} \operatorname{Erf} \left(\frac{\sigma_u^2 \ln(s) - 2u}{s \sigma_u} \right) \right) \Bigg|_{u=-\infty}^{\infty} \cdot \\
 &\quad \left(-\frac{i \sigma_v \sqrt{\pi}}{2} e^{-\frac{1}{4}\sigma_v^2 \ln^2(s)} \operatorname{Erfi} \left(\frac{\sigma_v^2 \ln(s) + i2u}{s \sigma_v} \right) \right) \Bigg|_{v=-\infty}^{\infty} \\
 &= k s^{\bar{u}} \left(\sigma_u \sqrt{\pi} e^{\frac{1}{4}\sigma_u^2 \ln^2(s)} \right) \left(\sigma_v \sqrt{\pi} e^{-\frac{1}{4}\sigma_v^2 \ln^2(s)} \right) \\
 &= k \pi \sigma_u \sigma_v \left(e^{\frac{1}{4}(\sigma_u^2 - \sigma_v^2) \ln^2(s)} \right) s^{\bar{u}}, \quad (39)
 \end{aligned}$$

a real operator. If $\sigma_u = \sigma_v$, this reduces to

$$\begin{aligned} H(s) &= ks^{\bar{u}} \sigma_u \sigma_v \pi \\ &= k\pi \sigma_u^2 s^{\bar{u}}. \end{aligned} \quad (40)$$

For Gaussian complex order-distributions centered off the real axis, these results generalize to

$$H(s) = k\pi \sigma_u \sigma_v \left(e^{\frac{1}{4}(\sigma_u^2 - \sigma_v^2) \ln^2(s)} \right) s^{\bar{u} + i\bar{v}} \quad (41)$$

and when $\sigma_u = \sigma_v$,

$$H(s) = k\pi \sigma_u^2 s^{\bar{u} + i\bar{v}}. \quad (42)$$

This operator is complex, however when summed with its conjugate $\bar{H}(s) = k\pi \sigma_u^2 s^{\bar{u} - i\bar{v}}$ a real operator is obtained.

Sums of these Gaussian order-distributions can be combined to approximate continuous order-distributions that are symmetric in the complex order-plane as follows. For $\sigma_u = \sigma_v$,

$$\begin{aligned} H(s) &= \sum_{m=-\infty}^{\infty} \sum_{n=1}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} k_{n,m} e^{-\frac{(u-\bar{u})^2}{\sigma_u^2} - \frac{(v-\bar{v})^2}{\sigma_v^2}} s^{u+iv} du dv. \\ &= \pi \sum_{m=-\infty}^{\infty} \sum_{n=1}^{\infty} k_{n,m} \sigma_{u_n}^2 s^{\bar{u}_n + i\bar{v}_m} \end{aligned} \quad (43)$$

If $\sigma_u \neq \sigma_v$,

$$\begin{aligned} H(s) &= \sum_{m=-\infty}^{\infty} \sum_{n=1}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} k_{n,m} e^{-\frac{(u-\bar{u})^2}{\sigma_u^2} - \frac{(v-\bar{v})^2}{\sigma_v^2}} s^{u+iv} du dv. \\ &= \pi \sum_{m=-\infty}^{\infty} \sum_{n=1}^{\infty} k_{n,m} \sigma_{u_n} \sigma_{v_m} \left(e^{\frac{1}{4}(\sigma_u^2 - \sigma_v^2) \ln^2(s)} \right) s^{\bar{u}_n + i\bar{v}_m} \end{aligned} \quad (44)$$

4.3 Example

Figure 3a shows a complex order-distribution with four Gaussians summed in the transfer function denominator, each with variance 0.5; one centered at $q = 0$ with weighting $4/\sqrt{2\pi}$, one centered at $q = 1.5$ with weighting $6\sqrt{2\pi}$, one

centered at $q = 1 + 0.5i$ with weighting $-1/\sqrt{2\pi}$, and one centered at $q = 1 - 0.5i$ with weighting $-1/\sqrt{2\pi}$, that is

$$H(s) = \frac{1}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{6}{\sqrt{2\pi}} e^{-\frac{(u-1.5)^2}{0.5} - \frac{v^2}{0.5_v}} s^{u+iv} dudv - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(u-1)^2}{0.5} - \frac{(v-0.5)^2}{0.5_v}} s^{u+iv} dudv - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(u-1)^2}{0.5} - \frac{(v+0.5)^2}{0.5_v}} s^{u+iv} dudv + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{4}{\sqrt{2\pi}} e^{-\frac{(u)^2}{0.5} - \frac{(v)^2}{0.5_v}} s^{u+iv} dudv} \quad (45)$$

Using the results of Eqs. (39) and (41), this can be simplified to the transfer function

$$H(s) = \frac{\sqrt{\pi/2}}{6s^{1.5} - (s^{1+0.5i} + s^{1-0.5i}) + 4}, \quad (46)$$

or, using the results of section 2, this becomes

$$H(s) = \frac{\sqrt{\pi/2}}{6s^{1.5} - 2s(\cos(0.5 \ln(s)) + 4)}. \quad (47)$$

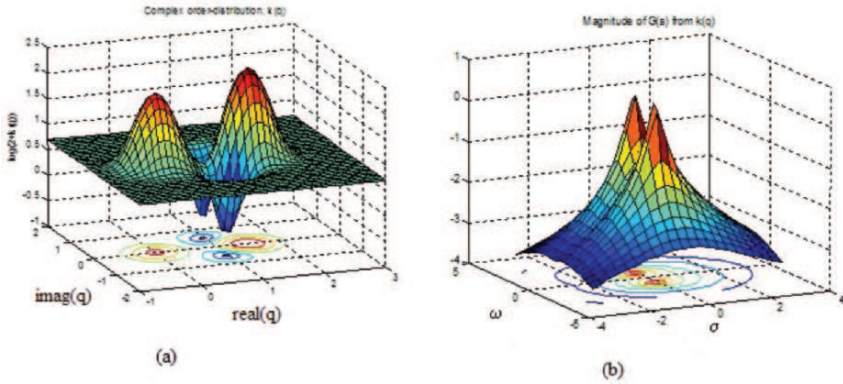


Fig. 3. (a) Complex order-distribution used in the example, Eq. (45). (b) Magnitude of the example system, Eq. (45), for all s .

The negative weighting on the two “damping terms” allows some resonance in the system. The magnitude plot of this complex order-distribution as a function of the Laplace variable s is shown in Fig. 3b. Two s -plane singularities can be clearly seen which lead to the resonances. Figure 4 shows the Bode magnitude and phase responses. The Bode plots were obtained from each of these transfer functions (Eqs. (45) and (46)), and they were visibly identical. The transfer function of Eq. (45) required the computation of the double integrals via

Euler integration which yielded a double summation, for each frequency. The transfer function of Eq. (46) was easily evaluated for each frequency. It is interesting to observe that even when the center terms of Eq. (46) have real exponents representing real derivatives, or $\bar{\nu} = 0$ in Eq. (42), $H(s)$ still has (symmetric) complex content. That is, complex Gaussian order-distributions are indistinguishable from individual isolated differintegrals (delta-function distributions). This is an interesting property of the Gaussian order-distributions not seen in the block complex order-distributions. This seems to have important implication to physical processes and requires further study.

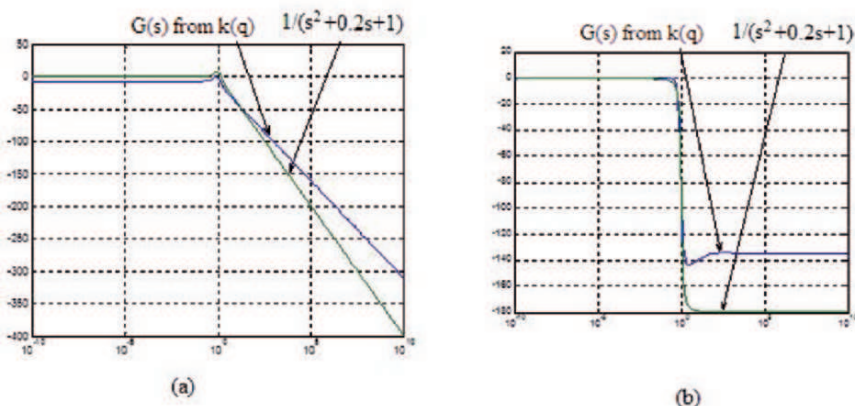


Fig. 4. Bode (a) magnitude and (b) phase plots for the example given in Eq. 45 with the Bode plots for $\frac{1}{s^2 + 0.2s + 1}$ for reference.

5 Conclusions and Practical Implications

Conjugated-order differintegrals have been defined in the time-domain, and their Laplace transforms have been determined. The use of conjugated-order differintegrals allows the use of complex-order operators while retaining real time-responses. Complex-weighted conjugated differintegrals have been investigated, showing that particular weightings have real time-responses.

Expanding collections of conjugated differintegrals to a continuum, complex order-distributions have been introduced. Both blockwise constant and Gaussian complex order-distributions were presented in the Laplace domain. Results which show how to use the blockwise constant or Gaussian order-distributions for approximating any complex order-distribution have been given. Further, it was shown that Gaussian distributions with circular symmetry have Laplace transforms proportional to that of an impulsive order-distribution, although

scaled by the width of the Gaussian. The conjecture that the frequency responses of impulsive distributions are indistinguishable from those of Gaussian distributions centered at the same locations is still under study.

With the analysis completed here, extension of the fractional identification procedure of [1] from real order-distributions to complex order-distributions is possible. Thus, complex order-distributions may be identified using real physical data. From this study we speculate that it may be possible to better describe the behavior of some real dynamic systems with complex-order distributions than with conventional methods. This may allow new understanding and modeling of fractional physical systems.

Acknowledgment

The authors gratefully acknowledge the support of the NASA Glenn Research Center.

References

1. Hartley TT, Lorenzo CF (2003) Fractional system identification based continuous order-distributions, *Signal Processing*, 83:2287–2300.
2. Lorenzo CF, Hartley TT (2002) Variable order and distributed order fractional operators, *J. Nonlinear Dyn., Spec. J. Fract. Calc.* 29(1–4):201–233.
3. Kober H (1941) On a theorem of shur and on fractional integrals of purely imaginary order, *J. Am. Math. Soc.* 50.
4. Love ER (1971) Fractional derivative of imaginary order, *J. Lond. Math. Soc.* 2(3):241–259.
5. Oustaloup A, Levron F, Mathieu B, Nanot FM (2000) Frequency-band complex noninteger differentiator: characterization and synthesis, *IEEE Trans. Circ. Syst. I*, 47(1):25–39.
6. Abramowitz M, Stegun IA (1964) *Handbook of Mathematical Functions*, Dover, New York.
7. Maskarinec GJ, Onaral B (1994) A class of rational systems with scale-invariant frequency response, *IEEE Trans. Circ. Syst. I*, 41(1).
8. Charef A, Sun HH, Tsao YY, Onaral B (1992) Fractal System as Represented by Singularity Function, *IEEE Trans. Auto. Control*, 32(9).
9. Hartley TT, Adams JL, Lorenzo CF (2005) Complex Order Distributions, *Proceedings of 2005 ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Long Beach, CA, September 24–28.

Part 6

Viscoelastic and Disordered Media

FRACTIONAL DERIVATIVE CONSIDERATION ON NONLINEAR VISCOELASTIC STATICAL AND DYNAMICAL BEHAVIOR UNDER LARGE PRE-DISPLACEMENT

Hiroshi Nasuno¹, Nobuyuki Shimizu², and Masataka Fukunaga³

¹ Iwaki Meisei University, Japan; E-mail: db0201@iwakimu.ac.jp

² Iwaki Meisei University, 5-5-1 Chuodai Iino, Iwaki, Japan; E-mail: nshim@iwakimu.ac.jp

³ Nihon University, Japan; E-mail: fukunaga@apple.ifnet.or.jp

Abstract

The nonlinear force-displacement relations of a viscoelastic cylindrical column under uniaxial monotonic slow compressive displacement with a constant speed, and under uniaxial rapid sinusoidal displacement with a constant compressive pre-displacement were experimentally and theoretically investigated to describe fractional derivative models for these relations. They were separately extracted from the slow compressive and the rapid sinusoidal experiments. These fractional derivative models were combined to construct a unified nonlinear viscoelastic model to cover from slow to rapid phenomenon appeared in the test specimen. This model successfully reproduced the slow and the rapid phenomena in the experiment.

Keywords

Viscoelastic cylindrical column, slow quasi-static phenomenon, rapid dynamic phenomenon, nonlinear fractional derivative model.

1 Introduction

Fractional calculus is known as a fundamental tool to describe the behavior of weak frequency dependence of viscoelastic materials in a broad frequency range. Fractional derivative constitutive models offer many successes in engineering fields to analyze linear viscoelastic problems (Rabotnov, 1980; Koeller, 1984; Bagley et al., 1983a, b). However, there are few experimental studies on nonlinear fractional derivative models that describe nonlinear force-displacement relations of viscoelastic bodies. Recently, Sjöberg et al. (2003) proposed a nonlinear dynamic fractional derivative model which consists of a linear fractional derivative element, a nonlinear elastic element, and a friction element for a rubber vibration isolator under harmonic displacement

excitation with static pre-compression. Deng et al. (2004) presented a fractional derivative model with nonlinear elastic element to describe quasi-static viscoelastic compression responses.

The authors have been trying to construct a fractional viscoelastic model of a viscoelastic body by experimental ways and theoretical ways (Nasuno et al., 2004, 2005). In this paper, we summarize the experimental results and give some considerations on the individual nonlinear model for slow and rapid phenomena to construct a unified model which can describe these phenomena in a whole frequency region. In the quasi-static experiments, the column specimen was compressed slowly to a target displacement x_0 with a constant speed, which is referred to as the ramp stage. In the dynamical experiments, the test specimen is first compressed slowly to the displacement x_0 . Then it was forced to oscillate sinusoidally around x_0 , which is referred to as the oscillatory stage. In Chapter 2, fundamental properties of nonlinear analytical responses for a type of element $x^\nu D^q x(t)$ are investigated analytically. In Chapter 3, the experiments are summarized briefly to extract the nonlinear fractal derivative models for the slow and the rapid phenomena. In Chapters 4, the following type of nonlinear fractional derivative model

$$c(x)D^q x(t) = F(t), \quad (1)$$

is proposed for both in the slowly compressed process and in the rapidly oscillatory process. In Eq. (1), $c(x)$ is the function of the input $x(t)$ of the test specimen, $F(t)$ is the reaction force, and D^q is the Riemann–Liouville’s fractional derivative defined by (Miller et al., 1993)

$${}_a D_t^q x(t) = \left(\frac{d}{dt}\right)^n \int_a^t \frac{(t-\tau)^{n-q-1}}{\Gamma(n-q)} x(\tau) d\tau, \quad (2)$$

where n is a integer number satisfies $n-1 \leq q < n$, and $\Gamma(\cdot)$ is the gamma function.

In Chapter 5, a unified nonlinear fractional derivative model is proposed to explain the results of the quasi-static experiments and the dynamical experiments. The model consists of two terms that represent the rapid process and the slow process.

2 Fundamental Properties of Nonlinear Response

2.1 Separation of variables

For the analysis of properties of the ramp stage and the oscillatory stage it is convenient to separate the variable $x(t)$ into the slowly varying part $x_g(t)$ and the rapidly oscillating part $y(t)$ as (Fukunaga et al., 2005)

$$x(t) = x_g(t) + y(t). \quad (3)$$

It is assumed that $x_g(t)$ is constant in the oscillatory stage, $t \geq 0$. The right-hand side of Eq. (1) is divided into

$$F(t) = F_g(t) + F_p(t), \quad (4)$$

where

$$\begin{aligned} F_g(t) &= c(x_g) D^q x_g(t), \\ F_p(t) &= [c(x) - c(x_g)] D^q x_g(t) + c(x) D^q y(t). \end{aligned} \quad (5)$$

If the order q in Eq. (5) is integer, and if $x_g(t)$ is constant in $t > 0$, one has only to solve the equation for $t > 0$. In the present case, however, the fractional derivative $D^q x_g(t)$ does not vanish unless $x_g(\tau)$ vanishes identically for both in $\tau > 0$ and in $\tau \leq 0$. Therefore one has to solve whole of Eqs. (4) and (5).

2.2 Response of nonlinear elements

Analytical solutions of

$$F_\nu(t) = x^\nu D^q x(t). \quad (6)$$

is obtained for the displacement given by Eq. (3) with

$$x_g(t) = \begin{cases} 0, & t \leq -t_0, \\ x_0(1 + t/t_0) & -t_0 < t \leq 0, \\ x_0 & t > 0, \end{cases} \quad (7)$$

and

$$y(t) = y_0 \sin(\omega t) = y_0 \operatorname{Re}[\exp(i\omega t)], \quad (8)$$

where $\operatorname{Re}[\cdot]$ denotes the real part of a complex number.

For each integer ν , the response F_ν is divided into

$$F_\nu(t) = F_{g,\nu}(t) + F_{p,\nu}(t), \quad (9)$$

where $F_{g,\nu}(t)$ and $F_{p,\nu}(t)$ are given by

$$\begin{aligned} F_{g,\nu}(t) &= x_g(t)^\nu D^q x_g(t), \\ F_{p,\nu}(t) &= [x(t)^\nu - x_g(t)^\nu] D^q x_g(t) + x(t)^\nu D^q y(t). \end{aligned} \quad (10)$$

The solution for $\nu = 1$ and 2 are given in Fukunaga et al. (2005).

In the experiments given by Nasuno et al. (2004, 2005), the sinusoidal input is imposed to the specimen after the ramp stage. Thus, the solution to the input given by Eqs. (3), (7), and

$$y(t) = \begin{cases} 0, & t \leq 0, \\ y_0 \sin(\omega t), & t > 0. \end{cases} \quad (11)$$

is also examined. The solutions tend to approach the solutions to the input given by Eq. (8) in a few periods of oscillation. An example is given in Fig. 1. In Fig. 1(a), the response $F(t)$ to the displacement given by Eqs. (3), (7), and (11) is plotted (*the solid line*) for the parameters, $\nu = 2$, $q = 1/2$, $t_0 = 540$, $x_0 = -1$, $y_0 = 0.2$, and $\omega = 2\pi$. The center of oscillation shifts to negative F because of negative pre-displacement x_0 . The *dotted line* is the input $y(t)$. The advanced phase shift of $F(t)$ relative to $x(t)$ is due to the fractional derivative of $x(t)$ of order $1/2$. In the early stage, the response is not periodic, since $y(t) = 0$ in $t \leq 0$. However, it tends to be periodic in a few period of oscillation. This can be seen clearly in Fig. 1(b) in which the oscillatory part of the response, $F_p = F(t) - F_g(t)$ (*the solid line*), is compared with the analytic solution $F_{p,2}(t)$ (*the dotted line*) to the input given by Eqs. (3), (7), and (8).

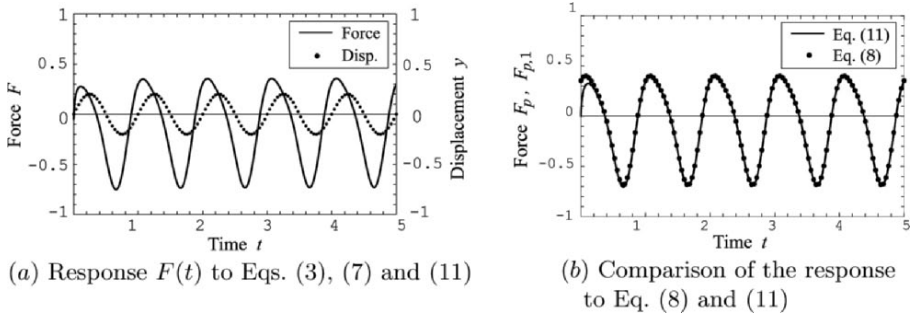


Fig. 1. The response of Eqs. (3), (7), and (11) with $q = 1/2$, $t_0 = 540$, $x_0 = -1$, $y_0 = 0.2$.

3 Nonlinear Experiment

3.1 Procedure of experiment

The authors conducted the experiments to investigate the nonlinear viscoelastic behavior of a cylindrical column subjected to a uniaxial displacement. Two types of experiment have been conducted (Nasuno et al., 2005) for the acrylic laminated viscoelastic cylindrical column (material; SD112 of Sumitomo 3M Co. Ltd.) with diameter $\phi = 60$ mm and height $h = 27$ mm ($1 \text{ mm} \times 27$ layers).

Type 1 experiment

The test specimen is slowly compressed uniaxially by a constant speed α_v up to the target displacement x_0 . This stage is referred to as the ramp stage ($x_g(t)$ in Eq. (7)). After the final point of the ramp stage is reached, the displacement

is held constant at $x = x_0$, which is referred to as the constant stage. The combinations for three different α_v [m/s] ($\alpha_v \times 10^6 = -3.2, -7.0, -15.0$) and five different x_0 [m] ($x_0 \times 10^3 = -2.7, -5.4, -8.1, -10.8, -13.5$) are employed as the experimental condition. Figure 2 shows a hydraulic servo-type actuator. The deformation of the specimen in axial direction was measured by a laser displacement meter, and the reaction force in axial direction by a load cell. The definition of the symbols used in measurement and analysis is shown in Fig. 3.

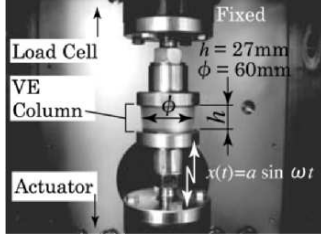


Fig. 2. Experimental set up of hydraulic actuator and VE column specimen.

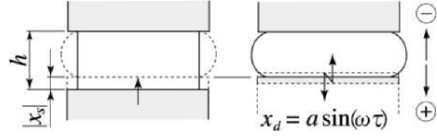


Fig. 3. Definition of displacement symbols used in experiment.

Type 2 experiment

First, the specimen is compressed uniaxially as in the ramp stage of Type 1 experiment. Then, the specimen is forced to oscillate sinusoidally with an amplitude y_0 around the center x_0 . This stage is referred to as oscillatory stage ($y(t)$ in Eq. (11)). The combinations for α_v [m/s], y_0 [m] ($\alpha_v \times 10^6 = -3.0, y_0 \times 10^3 = 1.5$ and $\alpha_v \times 10^3 = -15.0, y_0 = 1.0$ mm) and six different f [Hz] ($f = 0.1, 0.3, 1, 3, 5, 10$) are employed as the experimental condition.

In Type 1 experiment, the applied displacement in the ramp stage induces viscoelastic slow process. When the $|x_0|$ is large, the response of the specimen shows a nonlinear strain-dependent behavior. In Type 2 experiment, the applied displacement in the oscillatory stage induces viscoelastic rapid process. When x_0 is large, the response of the specimen also shows a nonlinear strain-dependent behavior, even if the amplitude of oscillation is small compared with the value x_0 . This is due to the aftereffect of the slow process.

3.2 Result of experiment

3.2.1 Result of Slow Process

The curves in Fig. 4 show the applied displacements and the corresponding measured forces of the test specimen for different target displacement x_0 and

for different constant applied speed α_v of the displacement for Type 1 experiment. Larger values of $|\alpha_v|$ and larger values of $|x_0|$ give higher values of the reaction force $|F|$. The time response of the force do not follow the linear fractional derivative model for larger $|x_0|$.

The curves in Fig. 5 show the displacement-force relations extracted from Fig. 4. These curves also show that the force and the displacement relations do not follow the linear fractional derivative model

$$F_l(t) = c_l D^{q_l} x(t). \quad (12)$$

except for the small displacement-force region (*white dashed lines* in Fig. 5). Where $F_l(t)$ is the force, $x(t)$ is the displacement, c_l and q_l are the viscoelastic coefficient and fractional order for the slow process, respectively (Subscript l refers to the linear viscoelastic term).

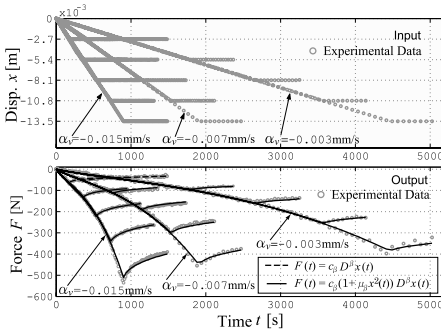


Fig. 4. Time histories of applied displacements (*upper curves*) and of measured forces (*lower curves*).

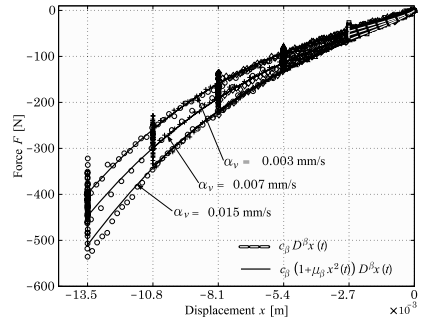


Fig. 5. Displacement-force relations of slow process for different applied speed.

3.2.2 Result of Rapid Process

Figure 6 shows a typical example of the applied sinusoidal displacement and the corresponding measured force of the test specimen for Type 2 experiment. The force curve does not follow the sinusoidal response. The amplitude of the response in compression side and in tension side are different, which shows a nonlinear response.

Figure 7 shows the displacement-force relation, the hysteresis loop drawn by the data in Fig. 6. The curve does not follow a perfect oval hysteresis loop of the linear fractional derivative model as in Eq. (12). Thus the loop shows the nonlinear behavior of the response.

Figure 8 shows hysteresis loops for all experimental conditions conducted in this experiment. The amplitude of the force $|F_0|$ increases with the increase of the target displacement $|x_0|$ under the same excitation frequency f . The area of the hysteresis loop increases with $|x_0|$ and with the frequency of oscillation.

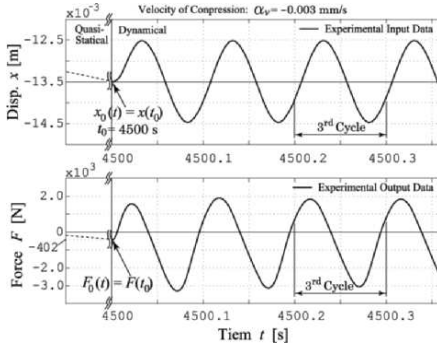


Fig. 6. Time histories of applied sinusoidal displacement (*upper curve*) and of measured force (*lower curve*).

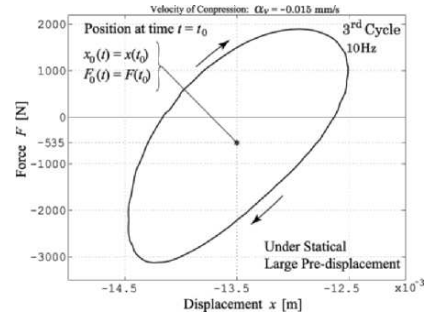


Fig. 7. Hysteresis loop of displacement-force relation of rapid process ($f = 1.0$ Hz, $x_0 = -13.5$ mm, $F_0 = 535$ N).

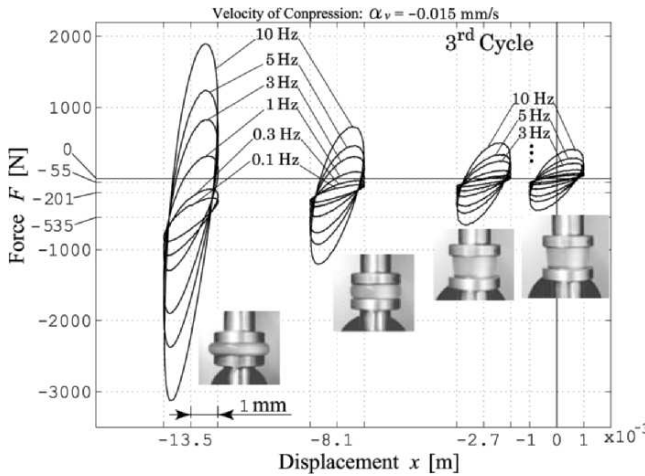


Fig. 8. Hysteresis loops of rapid process for all experimental conditions.

4 Two Fractional Derivative Models for Slow and Rapid Processes

4.1 Nonlinear slow process

In the experiment, the response $F(t)$ of the specimen is measured against $x(t)$ given by Eqs. (3), (7), and (11) from uncompressed equilibrium position. The authors assume that the time response in the early part of the ramp stage follows a linear fractional derivative model. Thus, the analysis is carried out based on the linear fractional derivative model written by Eq. (12).

The force response $F(t)$ was identified with the type of the model

$$F(t) = c_\beta(x)F_i(t), \quad (13)$$

$$c_\beta(t) = c_l (1 + \mu_\beta x^2(t)). \quad (14)$$

The displacement is normalized by the total height $h = 27$ mm of the specimen as

$$\begin{cases} \epsilon(t) = x(t)/h, & \eta(t) = y(t)/h, & \epsilon_g(t) = x_g(t)/h, \\ \epsilon_0 = x_0/h, & \eta_0 = y_0/h, & \epsilon(t) = \epsilon_g(t) + \eta(t). \end{cases} \quad (15)$$

for later convenience.

As is shown in Fig. 5, the specimen shows nonlinear force response $F(t)$ to the input $x(t)$ in the ramp stage and in the constant stage. The expression of the type of Eq. (1) i.e., Eqs. (13) and (14) for these stages is given as (Nasuno et al., 2004, 2005)

$$F(t) = F_{\beta,exp}(t) = c_{\beta,exp} D^{\beta_{exp}} \epsilon(t), \quad (16)$$

where

$$\begin{aligned} \beta_{exp} &= 0.20 [-], \\ c_{\beta,exp} &= b_{0,exp}(1 + 5.17\epsilon^2) \quad [\text{Ns}^{\beta_{exp}}], \\ b_{0,exp} &= 1.66 \times 10^3 \quad [\text{Ns}^{\beta_{exp}}] \end{aligned} \quad (17)$$

for $\epsilon_0 \leq 0$ and for the duration of the ramp stage is $540 \text{ s} \leq t_0 \leq 4500 \text{ s}$. The negative sign of $x, (\epsilon_0)$ means compression. The subscript *exp* indicates the experimental value.

4.2 Nonlinear rapid process

The energy dissipation of a viscoelastic body or a damping device per unit cycle is called damping capacity, and is defined by

$$W = \oint c_\alpha D^\alpha x(t) dx, \quad (18)$$

where the suffix α indicates the rapid process. Substituting $x = x_0 + y_0 \sin(\omega t)$ into Eq. (18), and neglecting the contribution from higher-order terms, one obtains

$$W = \pi c_\alpha y_0^2 \omega^\alpha \sin\left(\frac{\pi\alpha}{2}\right) = W_0 \left(\frac{f}{f_0}\right)^\alpha, \quad (19)$$

where W_0 is written as

$$\begin{aligned} W_0 &= c_\alpha(x_0) \pi y_0^2 (2\pi)^\alpha \sin\left(\frac{\pi\alpha}{2}\right) = c_\alpha(x_0) y_0^2 w_0, \\ w_0 &= \pi (2\pi)^\alpha \sin(\pi\alpha/2). \end{aligned} \quad (20)$$

The fractional order α and the nonlinear coefficient $c_\alpha(x_0)$ can be estimated by the frequency dependence and x_0 -, y_0 - dependence of the damping capacity in Eqs. (19) and (20).

The damping capacity of the specimen was experimentally estimated from the area of the third cycle of the hysteresis loop. Figure 9 shows the (f/f_0) - W relation in the log-log scale for $y_0 = 1.0$ mm under the standard frequency of $f_0 = 1.0$ Hz. Figure 10 shows the relation of $(|x|/h)$ - W in the linear-log scale (The marks indicate the experimental values for $x = |x_0|$). From these figures, $\log_{10} W$ is proportional to $|x|/h$ with a constant slope. Thus, W obeys the exponential law of $|x|/h$. Further, it was separately confirmed that W is being proportional to y_0^2 by the direct data analysis. From the above results under constant temperature, the general description of W can be written as

$$\log_{10} W = a_1 + a_2 \left(\frac{|x|}{h} \right) + 2 \log_{10} y_0 + \alpha \log_{10} \left(\frac{f}{f_0} \right). \quad (21)$$

The parameters α , a_1 and a_2 estimated from Figs. 9 and 10 are $\alpha = 0.54$, $a_1 = 5.46$ and $a_2 = 1.29$, respectively. From Eqs. (19), (20), and (21), the expression Eq. (1), for the oscillatory stage is obtained as (Nasuno et al., 2004, 2005)

$$F(t) = F_{\alpha,exp}(t) = c_{\alpha,exp} D^{\alpha_{exp}} \epsilon(t), \quad (22)$$

with

$$\begin{aligned} c_{\alpha,exp}(x) &= c_{1,exp} \exp(\mu_{1,exp} \epsilon) \\ &\simeq 1.21 \times 10^3 (1 - 2.6\epsilon + 6.0\epsilon^2) \quad [\text{Ns}^{\alpha_{exp}}], \\ c_{1,exp} &= 4.57 \times 10^4 \quad [\text{Ns}^{\alpha_{exp}}], \\ \mu_{1,exp} &= 1.10 \times 10^2 \quad [-] \end{aligned} \quad (23)$$

for $-0.5 \leq \epsilon_0 \leq 0$ ($-13.5 \text{ mm} \leq x_0 \leq 0 \text{ mm}$), $\eta_0 = 0.037$ ($y_0 = 1 \text{ mm}$) and $0.1 \text{ Hz} \leq f = \omega/2\pi \leq 10 \text{ Hz}$, respectively.

It should be noted that the numerical values of the damping coefficients $c_{\beta,exp}$ for the slow process and $c_{\alpha,exp}$ for the rapid process are similar in spite of the difference in the dimensionality of these coefficients. The reason why the fractional order $\alpha \simeq 0.5$ for rapid process and $\beta \simeq 0.2$ for slow process are different may be explained by the difference of the values of complex elastic modulus in the transition region and in the rubbery region (Sato et al., 2004).

The expressions $F_{\beta,exp}$ and $F_{\alpha,exp}$ were obtained by assuming that a single term acts in the ramp stage and in the oscillatory stage, respectively. This assumption is validated by the fact that in the oscillatory stage, the damping capacity can be fitted by a single $W \propto f^{\alpha_{exp}}$ curve given by Eq. (22) over the observed frequency range, $0.1 \text{ Hz} \leq f \leq 10 \text{ Hz}$, and in the ramp stage $F(t)$ can be fitted by a single curve given by Eq. (16).

Here, a simple question emerges. When and how do the two terms change their appearance?

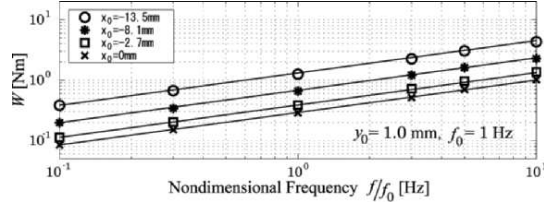


Fig. 9. Damping capacity with respect to nondimensional frequency.

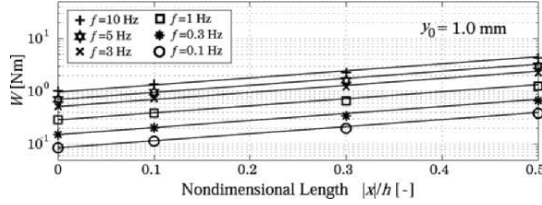


Fig. 10. Damping capacity with respect to nondimensional height.

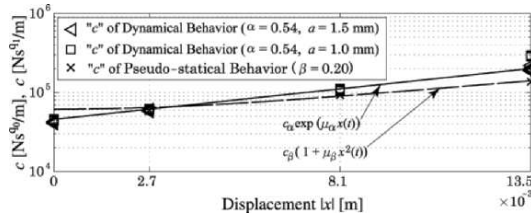


Fig. 11. Damping coefficient $c_\beta(x)$ for slow process and $c_\alpha(x)$ for rapid process.

5 Consideration on Unified Fractional Model for Two Processes

5.1 Variable coefficient model

There would be many possible models that explain the experimental results (Nasuno et al., 2004, 2005) in unified way. One of them is a simple sum of Eq. (16) with Eq. (17) and Eq. (22) with Eq. (23). However, it was found that this model failed from the comparison between the amplitude of response force at the oscillatory stage and the response at the end of the ramp stage (Fukunaga et al., 2005). Here, the authors consider a model consisting of two terms that characterize the fractional orders observed in the ramp stage and in the oscillatory stage. It is assumed that the coefficients of the both terms vary with the frequency or the parameters that describes the speed of change in deformation. Transition between the two terms is given in terms

of frequency-dependent coefficients (or other parameters) in an explicit way. The model equation is written as (Fukunaga et al., 2005)

$$c_\alpha(\epsilon; v)D^\alpha \epsilon(t) + c_\beta(\epsilon; v)D^\beta \epsilon(t) = F(t), \quad (24)$$

where the fractional orders are $\alpha \simeq 0.5$ and $\beta \simeq 0.2$. The deformation velocity defined as

$$v = d\epsilon/dt \quad (25)$$

is used tentatively for the index of rate of change. Note that v is not the definite parameter for the index of the rate of change.

The deformation velocity falls in the interval $1.1 \cdot 10^{-4}/\text{s} \leq v \leq 5.6 \cdot 10^{-4}/\text{s}$ for the ramp stage and in the interval $2.3 \cdot 10^{-2} \leq v \leq 2.3/\text{s}$ for the oscillate stage. In the oscillate stage, v varies as $v = \omega\eta_0 = 0.23f$ for $\eta_0 = 0.037$. Thus, the ramp stage is characterized by small v , whereas the oscillate stage is characterized by large v .

Note that individual term in Eq. (24) is essentially different from the experimentally obtained terms, Eq. (16) or Eq. (22). Equation (24) as a whole tends to Eq. (16) for small v , while it tends to Eq. (22) for large v . In this model, one or both coefficients vary with v .

5.2 Numerical consideration of variable coefficient model

In this section, the variable two-term model is discussed. As the first step of the analysis, the authors fixed the parameters to be constant as many as possible. Once dependence on v is proved to be established there are some parameters which were fixed may be relaxed to vary with v . The coefficient c_α of the α term is assumed to be constant.

Further it is assumed that the functional forms of non-linearity of c_α and c_β are fixed to those obtained by the experiment for the oscillatory stage and the ramp stage, respectively. The fractional order of the β term is fixed to that obtained by the experiment for the ramp stage. Thus, the coefficients of the model are written as

$$c_\alpha = a_0(1 - 2.6\epsilon + 6.0\epsilon^2) \quad [\text{Ns}^\alpha], \quad (26)$$

and

$$\begin{aligned} \beta &= 0.20, \\ c_\beta &= b_0(v)(1 + 5.17\epsilon^2) \quad [\text{Ns}^\beta]. \end{aligned} \quad (27)$$

The remaining parameter to be determined is a_0 , $b_0(v)$, and α .

First it will be shown that observed response given in Fig. 4 can be explained by Eq. (24). The response $F(t) = F_g(t)$ in the ramp stage is given by

$$F_g(t) = c_\alpha(\epsilon) \frac{\epsilon_0(1+t/t_0)^{1-\alpha}}{t_0^\alpha \Gamma(2-\alpha)} + c_\beta(\epsilon) \frac{\epsilon_0(1+t/t_0)^{1-\beta}}{t_0^\beta \Gamma(2-\beta)}. \quad (28)$$

The *dotted line* in Fig. 12 shows $F_g(t)$ with $a_0 = 0$, $b_0 = 1.66 \times 10^3 [\text{Ns}^\beta]$, $\epsilon_0 = -0.3$, and $t_0 = 540$ [s], which is equivalent to Eqs. (16) and (17) for the experiment. The maximum contribution from the α term is expected for the experimentally obtained values in the oscillatory stage

$$\begin{aligned} \alpha &= 0.54, \\ a_0 &= 1.21 \cdot 10^3 \quad [\text{Ns}^\alpha]. \end{aligned} \quad (29)$$

The value b_0 of the β term is estimated from Eqs. (28), (29), and the observed value of $F(0)$ at the end of the ramp stage as

$$b_0(\text{low } v) = 1.42 \times 10^3 \quad [\text{Ns}^\beta]. \quad (30)$$

The *soild line* in Fig. 12 shows $F_g(t)$ with Eqs. (29) and (30). The coincidence of the two curves shows the validity of the model in the ramp stage.

The values of α , a_0 , and $b_0(v)$ in the oscillatory stage are derived from the damping capacity and the amplitude of the response in the oscillatory stage and those of the experiment given in Figs. 8 and 9. The response $F_p(t)$ of the model in the oscillatory stage is given by Fukunaga et al. (2005). The damping capacity of the model is given by

$$W = \pi y_0^2 [c_\alpha(x_0) \omega^\alpha \sin \frac{\pi\alpha}{2} + c_\beta(x_0) \omega^\beta \sin \frac{\pi\beta}{2}]. \quad (31)$$

We seek the values of α , a_0 , and $b_0(v)$ that satisfy the relation

$$W \propto f^{q_{\text{single}}}, \quad (32)$$

where q_{single} is the order of fractional derivative for Eq. (31) when the two terms are combined to a single term. If Eq. (29) is adopted for the α term, $b_0(v)$ that satisfies Eq. (32) is obtained from the amplitude of oscillation. As for $\epsilon_0 = -0.3$ and $t_0 = 540$ we obtain $b_0(2.3/\text{s}) = b_0(10 \text{ Hz}) = 35 [\text{Ns}^\beta]$, $b_0(1 \text{ Hz}) = 470 [\text{Ns}^\beta]$, $b_0(0.1 \text{ Hz}) = 833 [\text{Ns}^\beta]$, etc. However, the exponent is $q_{\text{single}} = 0.51$ which is significantly less than $\alpha_{\text{exp}} = 0.54$.

The expected fractional order of the α term that satisfies $\alpha = \alpha_{\text{exp}}$ should be $\alpha > 0.54$. This is very important to understand the unified model by the separate models in the slow process and in the rapid process. Figure 13 shows the reproduction of Eq. (32) with $q_{\text{single}} = \alpha_{\text{exp}}$. The fractional order of the α term is $\alpha = 0.58$. The value of b_0 varies with the frequency. In Fig. 14, the hysteresis loop of the model (*soild line*) is compared with the experimentally obtained loop (*dots*) of $\epsilon_0 = -0.3$, $t_0 = 540$ s, and $f = 1$ Hz. The two curves are in good agreement, which shows again the validity of the model.

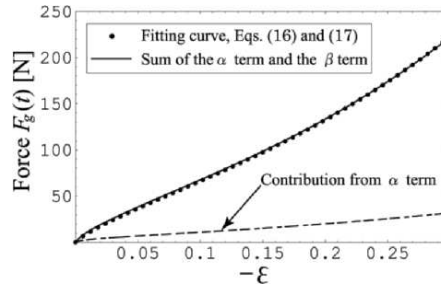


Fig. 12. The response $F_g(t)$ defined by Eq. (28) in the ramp stage.

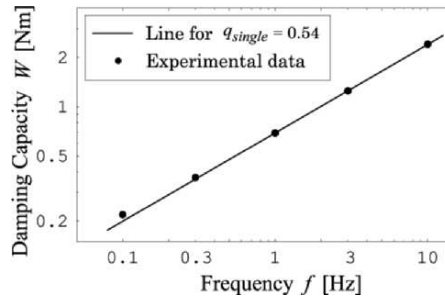


Fig. 13. The damping capacity of the two term model for $q_{single} = 0.54$.

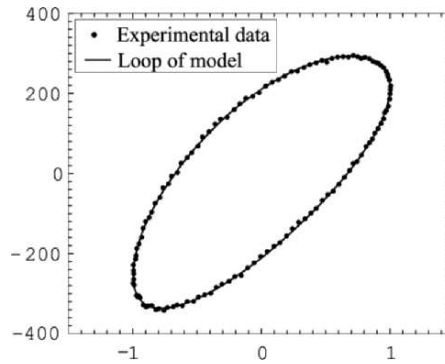


Fig. 14. The hysteresis loop of the model for $\alpha = 0.58$.

6 Conclusion

Nonlinear fractional derivative behaviors for the slow process and for the rapid process of the test specimen caused by geometrical nonlinearity have been modeled from the experimental data for both processes. It is found that the models for the slow process and the rapid process can be approximately expressed by the nonlinear fractional derivative term based on Eq. (1). The coefficients in the nonlinear fractional derivative models are the functions of

a compressive speed and a target depth. The values of viscoelastic coefficients in Eqs. (17) and in (23) are close each other as shown in Fig. 11.

A unified model in the form of Eq. (24) with velocity-dependent viscoelastic damping coefficients for the observed frequency range is constructed by considering the two separate models for the slow process and for the rapid process. The unified model could successfully reproduce the slow process phenomenon and the rapid process phenomenon.

7 Acknowledgment

The authors are thankful to Mr. Kiyoshi Okuma and Ken Tokoro of Sumitomo 3M Co. Ltd. who supplied the acrylic viscoelastic material, and Dr. Takuya Yasuno, an associate professor of Iwaki Meisei University who gave the authors valuable advices during the experiment.

This paper is a modified version of a paper published in proceedings of IDETC/CIE 2005, September 24–28, 2005, Long Beach, California, USA. The authors would like to thank the ASME for allowing them to republish this modification in this book.

References

1. Koeller RC (1984) *J. App. Mech.* 51:299–307.
2. Bagley RL, Torvik PJ (1983a) *J. Rheology* 27:201–210.
3. Bagley RL, Torvik PJ (1983b) *AIAA J.* 21:741–748.
4. Nasuno H, Shimizu N (2004) Proceedings of the 1st IFAC Workshop on FDA 2004. pp. 620–625.
5. Nasuno H, Shimizu N (2005) IDETC on MSNDC-12 Fractional Derivatives and Their Applications (Contribution number “DETC2005-84336” on CD-ROM).
6. Miller KS, Ross B (1993) *An Introduction to the Fractional Calculus and Fractional Differential Equations*. Wiley, New York.
7. Fukunaga M, Shimizu N, Nasuno H (2005) IDETC on MSNDC-12 Fractional Derivatives and Their Applications (Contribution number “DETC2005-84452” on CD-ROM).
8. Sato Y, Shimizu N, Yokomura T (2004) Proceedings of the 1st IFAC Workshop on FDA 2004: pp. 609–614.

QUASI-FRACTALS: NEW POSSIBILITIES IN DESCRIPTION OF DISORDERED MEDIA

R.R. Nigmatullin and A.P. Alekhin

Kazan State University, Kremlevskaya St., 18, 420008, Kazan,
Tatarstan, Russian Federation; E-mail: nigmat@knet.ru

Abstract

New generalization of fractals named as *quasi-fractals* (QF) is introduced for description of wide class of disordered media. The numerical calculations show that new fractal objects have wide region of applicability and can be used for description of fractals obtained by the diffusion-limited aggregation (DLA) procedure and for distorted lattices. These new facts found give new possibilities to apply the methods of the mathematics of the fractional calculus for description of relaxation and transport phenomena in disordered and heterogeneous systems.

Keywords

Quasi-fractals, disordered media, fractional calculus.

1 Introduction

As it is known [1] that description of a self-similar structure with the help of fractal dimension represents an effective tool for understanding of the scaling properties of disordered media. In paper [2] we suggested a new type of fractals (they were defined as *quasi-fractals* (QF) with logarithmic asymptotics) that can be suitable for description of a wide class of clusters formed by random fractals. As independent parameter we chose a *number* of coordination sphere j ($j = 1, 2, \dots$). With respect to this parameter the radius of coordination sphere can be expressed as $R(j) = R_0 j^\alpha$ and number of particles located inside of the sphere $R(j)$ is expressed by another power-law function $N(j) = N_0 j^\beta$. In the first time this new power-law dependence was confirmed in the model of coordination spheres [3] when the share of the volume formed by atoms of a *regular* lattice with respect to free volume of a crystal having certain symmetry was calculated. So, the first problem is to confirm the existence of QF that can be formed in the process of growth of random fractals. If these QF having slow logarithmic asymptotic can be applicable for description of different random clusters, including clusters having near-neighboring order then with their usage one can describe

describe wide class of disordered media. From another side, it could increase possibilities of the application of the fractional calculus for description of relaxation and transport properties in such kind of media, where the non-integer operators of differentiation and integration are appeared in the result of averaging procedure of a smooth function over fractal media. This statement has been confirmed in paper [4] and a “universal” decoupling procedure leading to kinetic equations, containing non-integer operators has been recently suggested in [5]

In this paper we obtained different types of QF modifying and generalizing the conventional diffusion-limited aggregation (DLA) procedure and procedure and considered the distorted lattices also. In all cases considered we confirmed the existence of QF having logarithmic asymptotic. This fact gives a possibility to increase of applications of the fractional calculus for description of different types of disordered media. From another side, one can expect that many properties of disordered media as relaxation and transport phenomena will be described by differential equations containing non-integer integrals and derivatives. So, the investigation of properties of QF will help essentially to many researches to reconsider the properties of disordered media from “fractal” point of view in order to increase the limits of applicability of the mathematics of the fractional calculus in physics and modern technology.

2 Description of the General Procedure

We decided to verify the relationships

$$R(j) = R_0 j^\alpha, \quad N(j) = N_0 j^\beta \quad (1)$$

on random fractals obtained numerically with the help of the conventional DLA procedure described in [6]. For calculation of the desired dependence it is necessary to choose the value of $R(j)$. Let us imagine that previous radius of sphere $R(j-1)$ has been found. The following algorithm is accepted:

1. It is chosen the nearest particle located on the distance d_0 with respect to the radius $R(j-1)$.
2. Then the distance d_0 between the center of the nearest particle and $R(j-1)$ is doubled and the next $R(j)$ is determined $R(j) = R(j-1) + 2d_0$.
3. In the spherical layer $2d_0$ the number of particles is calculated taking into account all particles having their centers located in the given layer.

Then the steps (1–3) are repeated and the power-law dependencies $R(j) = R_0 j^\alpha$ and $N(j) = N_0 j^\beta$ are verified in double-log scale. It is easy to show that the fractal dimension D_f is determined from the relationship

$$N = N_0 \left(\frac{R}{R_0} \right)^{\beta/\alpha} \equiv N_0 \left(\frac{R}{R_0} \right)^{D_f} \quad (2)$$

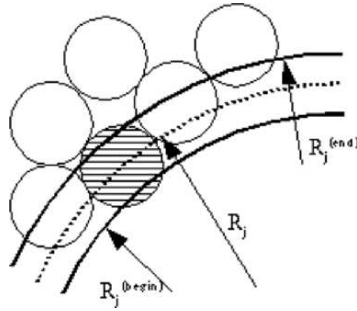


Fig. 1. Explains the formation of the layer $2d_0$ marked by two solid lines. The nearest particle is shaded. $R(j-1)$ coincides with the radius $R_j^{(begin)}$, $R(j)$ coincides with the radius $R_j^{(end)}$. The difference $R_j - R_j^{(begin)} = d_0$ gives the distance for the nearest particle.

One can suggest also the following modifications. One can put $\alpha=1$ and thereby to require the linear law for number of particles. In this case the radius of the following coordination sphere is determined as a distance up to the center of the following nearest particle, which is located out of the previous sphere. In the result of application of this modification one can obtain a possible dependence $R = R_0 j^\beta$, which helps to determine the value of β .

In the second modification one can put $\beta=1$ and to require the linear dependence for R . The calculation of number of particles located in the desired coordination sphere gives the value of α . The model calculations based on the basic procedure and modifications show that the first procedure is an “average” and modifications suggested can be considered as the limiting cases. The fractal dimensions calculated with the usage of these three methods are very close to each other and presumably follow to Eq. (1).

If in Eq. (1) we replace $\alpha \rightarrow \{\alpha_i\}$, $\beta \rightarrow \{\beta_i\}$, $j \rightarrow \varphi(j)$ then it is easy to obtain a possible generalization for more complex dependencies that might expect from analysis of real data.

3 The Basic Models Leading to QF

3.1 Description of the diffusion-limited aggregation (DLA) procedure

We introduce on the given plane the Cartesian coordinate system (XOY) and the center of this system is occupied by an initial “seed”-particle. (the particle has a size $r_0 = 0.1$). Then we form two circles with centers located in the point (0,0) and having the radiuses $r_{init} = 100r_0$ and $r_{out} = 300r_0$, respectively. The first circle serves as a source of the generated particles, the second one is used as a particles sink.

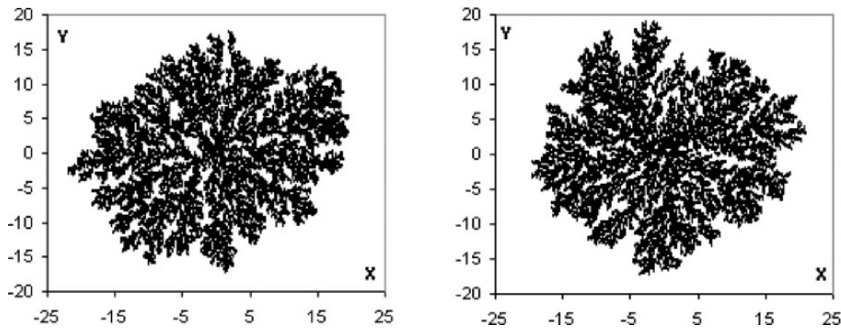
Let us consider the process of the particle movement in detail. A particle having the unit mass $m = 1$ is generated on the circle r_{init} and starts its moving with the value of velocity $V = 1$ in randomly taken direction. In the process of movement two forces acts: the random force having the constant value and uniform angular distribution and the second force is associated with viscosity friction force, which is proportional to the first degree of the given velocity. The first force gives the randomness factor and the friction force limits the value of velocity. “Quanta” of time dt is chosen from the condition that during of this period dt the chosen particle could move a path equaled approximately $(0.01 \div 0.02)r_0$. Starting its movement on the first circle we have two events: (1) “sedimentation” on the cluster or (2) the leaving of the system. The influence of the boundary conditions for this algorithm is negligible because the value of $r_{init} = 100r_0$ and the difference $r_{out} - r_{init}$ remain the constant during the whole growing process. In such way we “planted” 10 clusters having approximately 10^4 particles in each cluster.

3.2 Random rain model

This method is obtained from the first model if we switched off respectively the random and friction forces. The particles in the process of growing are moving in the arbitrary direction with constant value of the given velocity. Other peculiarities are remained the same as it was accepted for the DLA process.

Below we are giving the figures of typical clusters planted by two methods described above. The clusters obtained by the DLA and random rain models are shown by Figs. 2 and 3 respectively. The corresponding fractal dimensions and other parameters characterizing these fractals are given in Table 1. So, we can

prove that in the case of the fractal growth we obtain QF which gives the same fractal dimension and thereby cannot be detected by conventional methods.



Figs. 2 and 3 demonstrate two typical fractals obtained by the conventional DLA (on the left) and random rain model (on the right).

3.3 The DLA lattice model

In the lattice model a movement of the chosen particle is reduced to a floating random walk over knots (sites) of the chosen plane lattice. The seed-particle is placed on the site (0,0). Two source- and sink- circles are created as in the previous cases. The particle is walking randomly over the free sites of the lattice up to the moment when the neighboring particle belongs to the cluster. Then the fixed particle is stopped and the cluster is occupying this site of the lattice. After this “event” the following particle from the source-circle is initialized and the process is repeated.

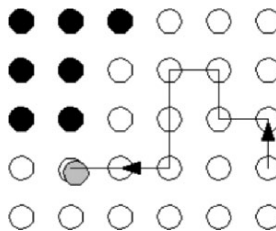


Fig. 4. Presents schematically a pass of a particle moving over the sites of the lattice. The next “hop” is possible only to some neighboring sites, which are chosen randomly. The growth process is stopped when the cluster achieves the given size. The size of the lattice (200×200) remains constant.

Using this method we obtained 10 “lattice” clusters. Some of them are depicted below.

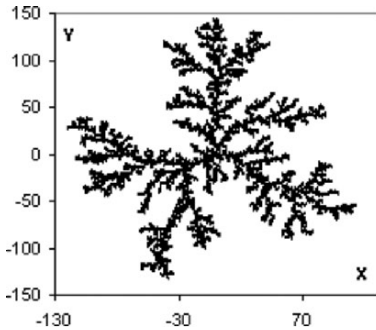


Fig. 5. Depicts the lattice cluster obtained on the honeycomb lattice. On the right-hand side.

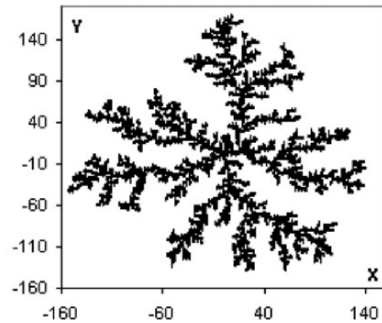


Fig. 6. We show the fractal obtained on the square lattice.

This lattice model was modified by introduction of random exclusion and permission. In the model with random exclusion the “overlooking” of a part of sites that can be occupied is not allowed. In the result of this exclusion the moving particle “does not know” about the state (free/occupied) of the banned site. In the model with limited permission to the particle moved is allowed only the “overlooking” only one neighboring site. Other conditions of the movement are remained the same.

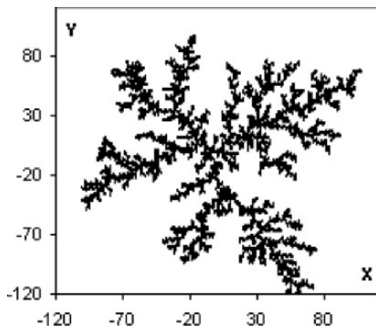


Fig. 7. (on the left) depicts the lattice cluster obtained on the honeycomb lattice with the random exclusion.

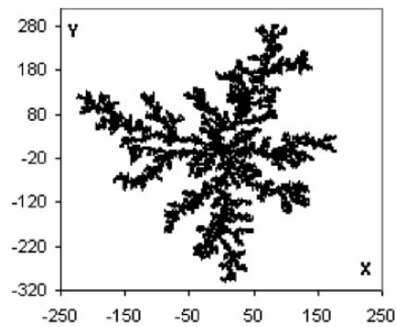


Fig. 8. (given on the right) shows the cluster obtained on the square lattice with the random permission.

In such a way we “planted” again 10 clusters with random modifications. All of them are turned out QF. The complete results of analysis of different fractal clusters are given in Table 1 of Appendix A. In saving place we present only two plots demonstrating the dependence of the type (1) in double-log scale.

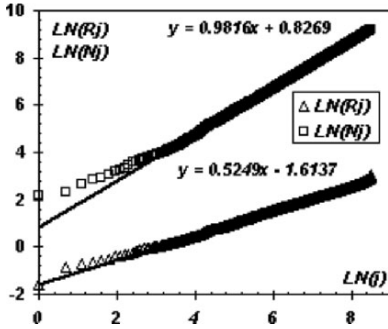


Fig. 9. Demonstrates the dependence (1) taken in the double-log for the conventional DLA.

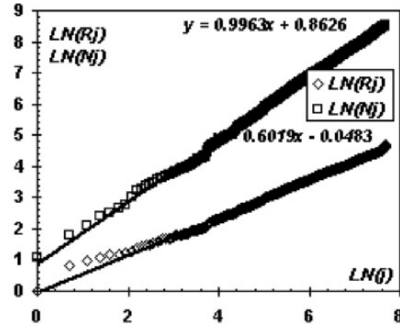


Fig. 10. The dependence (1) obtained for triangle lattice.

4 QF Found in Distorted Lattices

Let us suppose that we have plane square lattice with sizes $N_x \times N_y$ (in the program used we have the square 200×200), the lattice constant equals one. Each knot has two states. The first state corresponds to a regular structure; the second state corresponds to the distorted structure. Then we start to distort the lattice. The distorted state can be obtained by two methods: (1) to add a random value to the values of coordinates describing the given knot or (2) to delete a part of knots replacing them by voids. Let us consider each distortion in detail.

The value of the error for coordinates of the knot one can write as

$$\begin{cases} \delta x = \delta r \cos \varphi \\ \delta y = \delta r \sin \varphi \end{cases}$$

where δr is the value of the radial error. It is chosen randomly from the interval $[0; \delta r_{\max}]$. $\delta r_{\max}$ - is the maximal value of the deviation chosen as $\delta r_{\max} = 0.1$ from the value of the lattice constant (equaled unit value). The angle φ is also randomly chosen from the interval $[0; 2\pi]$. Both values $(\delta r, \varphi)$ are considered to be uniformly distributed.

As for the deleted knots of the lattice, they are chosen also randomly. The share of the deleted knots we put 0.1% from the total value. It is obvious that each knot is subjected by the influence of its neighbors, which are trying to keep the knot on the center distance equaled the lattice constant. In order to take into account this fact we realized the following procedure. Some knot is chosen and all possible distorted knots are counted. If some neighboring knots are in the second (distorted) state then the coordinates of the distorted knot are obtained as summation of coordinates belonging to the regular structure which are subjected by the error of the first kind. The same procedure is repeated for each distorted knot.

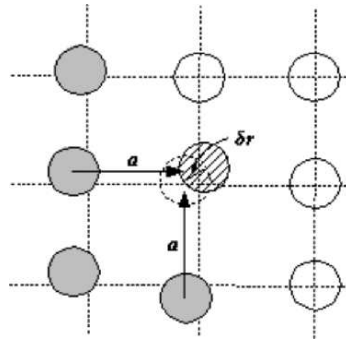


Fig. 11. explains pictorially the basic step of this algorithm. Summation of distortions coming from neighbors (*shown by arrows*) is shown for the central knot by the *dotted line*. Further summation including distortion from the central knot is shown by the shaded area.

After this procedure due to the algorithm the knot being located in the second state generates by the neighboring knots, which are in the same state. In this case we have gradual accumulation of the error which increases with the increasing of the distance accounted from the center of the lattice. In practical realization of this algorithm as a seed particle we used a group of the particles, which were located in the center circle ($r = 5$). This *shaded circle* is shown on Fig. 12. In this case the distortion “wave” has symmetry close to the spherical one. After transition of the whole lattice to the second state we deleted a certain number of knots (0.1%) in the final stage. Below we are giving an example of the distorted lattice realized with the usage of this algorithm.

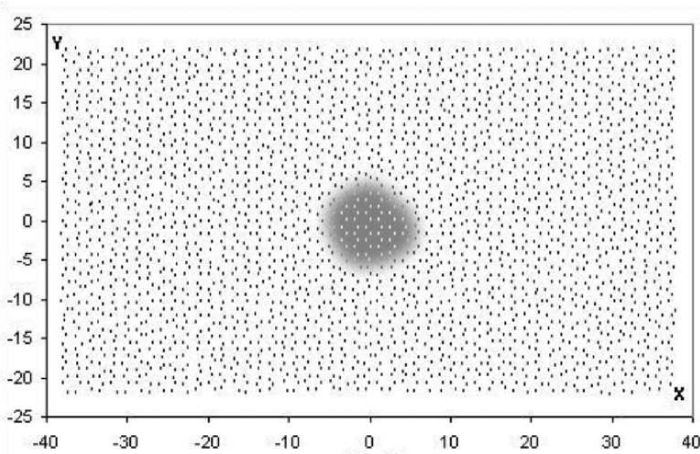


Fig. 12. This figure depicts an example of the honeycomb distorted lattice. The seed cluster forming the regular lattice is shaded.

5 Results and Discussions

In this paper we considered different models of disordered media. This disorder can follow to fractal or non-fractal scenario. In all cases we proved that for description of such kind of disorder the parameter as the coordination of sphere j is the most suitable parameter for description of the heterogeneous object. The distorted lattices (having small amount of “disorder”) also can be described in the framework of the same scheme as the “fractal” object (having large amount of disorder). In this paper we chose objects having a spherical symmetry. Other steps are related to the consideration of disorder obtained on a “substrate” having cylindrical symmetry. The model of QF admits of further generalization. Instead of simple expression (1) one can think about more general expression of a type

$$N(j) = N_1 j^{\alpha_1} + N_2 j^{\alpha_2} + \dots, R(j) = R_1 j^{\beta_1} + R_2 j^{\beta_2} + \dots \quad (3)$$

or more general formulae containing not only a linear combination of power-law functions. The next step will be the verification of these dependencies (1) or (3) on real data or on the computer models containing not only one scenario of the fractal growth.

But nevertheless one important conclusion has remained to mark it. The “fractal” (power law) behavior one can expect in different disordered systems, which do not obey initially the “classical” fractal behavior. This new concept of “disorder reconsideration” helps to find new objects with random/regular voids which can be successfully described in the framework of QF conception. In turn, it considerably facilitates of penetration of the mathematics of the fractional calculus for description of relaxation/transport phenomena in medium with different types of disorder.

Appendix A

Table 1. The calculated fractal dimension for clusters obtained for different systems obtained by the methods described in section 2. For the lattice model with random exclusion/permission we use the abbreviation (LMRE/P)

Type of model, number of neighbors (z)	R_0 (with values of the stdev)	α (with values of the stdev)	N_0 (with values of the stdev)	β (with values of the stdev)	D (with values of the stdev)	Size of the cluster (diameter, number of particles involved)
Conventional DLA-procedure	0.1743 ($\Delta = 0.0044$)	0.9996 ($\Delta = 0.0045$)	2.0235 ($\Delta = 0.0702$)	0.5427 ($\Delta = 0.0034$)	1.8419 ($\Delta = 0.0059$)	$N = 10^4$ $R = 21.8$
A random "rain" model.	0.1782 ($\Delta = 0.0049$)	1.0028 ($\Delta = 0.004$)	1,9807 ($\Delta = 0.0651$)	0,5399 ($\Delta = 0.0034$)	1,858 ($\Delta = 0.0087$)	$N = 10^4$ $R = 21.4$
The lattice DLA model (z = 3)	0.8972 ($\Delta = 0.0276$)	0,991 ($\Delta = 0.0075$)	2.1802 ($\Delta = 0.1379$)	0.6046 ($\Delta = 0.0046$)	1.6396 ($\Delta = 0.0103$)	$N = 5895$ $R = 138$
LMRE (z = 3)	0.8389 ($\Delta = 0.0334$)	1.0012 ($\Delta = 0.0052$)	2.0153 ($\Delta = 0.0841$)	0.6087 ($\Delta = 0.0059$)	1.6458 ($\Delta = 0.0109$)	$N = 6580$ $R = 139$
LMRP (z = 3)	0.7045 ($\Delta = 0.0275$)	1.0065 ($\Delta = 0.0055$)	1.8826 ($\Delta = 0.0922$)	0.6172 ($\Delta = 0.0047$)	1.6312 ($\Delta = 0.0105$)	$N = 7641$ $R = 140$
The lattice DLA model (z = 4)	0.8473 ($\Delta = 0.034$)	0.9902 ($\Delta = 0.0055$)	2.1936 ($\Delta = 0.1043$)	0.612 ($\Delta = 0.0047$)	1.6183 ($\Delta = 0.0097$)	$N = 8042$ $R = 167$
LMRE (z = 4)	0.8343 ($\Delta = 0.0341$)	0.9953 ($\Delta = 0.0062$)	2.0733 ($\Delta = 0.105$)	0.6096 ($\Delta = 0.0059$)	1.6335 ($\Delta = 0.0109$)	$N = 8250$ $R = 165$
LMRP (z = 4)	0.7026 ($\Delta = 0.0383$)	0.9903 ($\Delta = 0.0088$)	2.2237 ($\Delta = 0.1929$)	0.609 ($\Delta = 0.006$)	1.6266 ($\Delta = 0.008$)	$N = 11801$ $R = 167$
The lattice DLA model (z = 6)	1.8495 ($\Delta = 0.0639$)	0.9889 ($\Delta = 0.0058$)	2.2119 ($\Delta = 0.094$)	0.6096 ($\Delta = 0.0055$)	1.6229 ($\Delta = 0.0111$)	$N = 5901$ $R = 296$
LMRE (z = 6)	1.8459 ($\Delta = 0.1624$)	0.9861 ($\Delta = 0.0103$)	2.2872 ($\Delta = 0.216$)	0.6054 ($\Delta = 0.0101$)	1.631 ($\Delta = 0.0173$)	$N = 6118$ $R = 292$
LMRP (z = 6)	1.3517 ($\Delta = 0.0812$)	0.9919 ($\Delta = 0.0077$)	2.1654 ($\Delta = 0.1483$)	0.6081 ($\Delta = 0.0084$)	1.6331 ($\Delta = 0.0188$)	$N = 9778$ $R = 295$

Table 2. Parameters of the QF obtained by the procedure described in section 3

Type of model	R_0 (with values of the stdev)	α (with values of the stdev)	N_0 (with values of the stdev)	β (with values of the stdev)	D (with values of the stdev)	Size of the cluster (diameter, number of particles involved)
Random lattice ($z = 3$)	0.6801 ($\Delta = 0.0066$)	0.9892 ($\Delta = 0.0021$)	2.2152 ($\Delta = 0.0422$)	0.5388 ($\Delta = 0.001$)	1.8358 ($\Delta = 0.0035$)	200×200
Random lattice ($z = 4$)	0.932 ($\Delta = 0.0079$)	0.9462 ($\Delta = 0.0016$)	3.3279 ($\Delta = 0.0571$)	0.4853 ($\Delta = 0.0084$)	1.9497 ($\Delta = 0.0004$)	200×200
Random lattice ($z = 6$)	0.7053 ($\Delta = 0.0069$)	0.9867 ($\Delta = 0.002$)	4.5213 ($\Delta = 0.0909$)	0.5075 ($\Delta = 0.001$)	1.9442 ($\Delta = 0.0003$)	200×200

References

1. Mandelbrot B (1983) *The Fractal Geometry of Nature*. Freeman, San-Francisco.
2. Nigmatullin RR, Alekhin AP (2005) *Realization of the Riemann-Liouville Integral on New Self-Similar Objects*. In: Books of abstracts, Fifth EUROMECH Nonlinear Dynamics Conference August 7–12, pp. 175–176 Prof. Dick H. van Campen (ed.), Eindhoven University of Technology, The Netherlands.
3. Mehaute A, Nigmatullin RR, Nivanen L (1998) *Fleches du Temps et Geometrie Fractale*, Hermès, Paris (in French).
4. Nigmatullin RR, Le Mehaute A (2005) *J. Non-Cryst. Solids*, 351:2888.
5. Nigmatullin RR (2005) Fractional kinetic equations and universal decoupling of a memory function in mesoscale region, *Physica A* (has been accepted for publication).
6. *Fractals in Physics* (1985) The Proceedings of the 6th International Symposium, Triest, Italy, 9–12 July; Pietronero L, Tozatti E (eds.), Elsevier Science, Amsterdam, The Netherlands.

FRACTIONAL DAMPING: STOCHASTIC ORIGIN AND FINITE APPROXIMATIONS

Satwinder Jit Singh and Anindya Chatterjee

Mechanical Engineering Department, Indian Institute of Science, Bangalore
560012, India

Abstract

Fractional-order derivatives appear in various engineering applications including models for viscoelastic damping. Damping behavior of materials, if modeled using linear, constant coefficient differential equations, cannot include the long memory that fractional-order derivatives require. However, sufficiently great microstructural disorder can lead, statistically, to macroscopic behavior well approximated by fractional order derivatives. The idea has appeared in the physics literature, but may interest an engineering audience. This idea in turn leads to an infinite-dimensional system without memory; a routine Galerkin projection on that infinite-dimensional system leads to a finite dimensional system of ordinary differential equations (ODEs) (integer order) that matches the fractional-order behavior over user-specifiable, but finite, frequency ranges. For extreme frequencies (small or large), the approximation is poor. This is unavoidable, and users interested in such extremes or in the fundamental aspects of true fractional derivatives must take note of it. However, mismatch in extreme frequencies outside the range of interest for a particular model of a real material may have little engineering impact.

Keywords

Damping, fractional derivatives, stochastic processes, Galerkin projection, finite dimensional approximation, viscoelasticity

1 Introduction

The idea of fractional-order derivatives has been used in various engineering applications including models for viscoelastic damping. Damping behavior of materials, if modeled using linear, constant coefficient differential equations, cannot include the long memory that fractional-order derivatives require. However, sufficiently great microstructural disorder can lead, statistically, to macroscopic behavior well approximated by fractional order derivatives. The idea has appeared in the physics literature, but may interest an engineering audience. This idea in turn leads to an infinite-dimensional system without memory; a routine Galerkin projection on that infinite-dimensional system leads to a finite dimensional system of ordinary differential equations (ODEs) (integer order) that matches the fractional-order behavior over user-specifiable, but finite, frequency ranges. For extreme frequencies (small or large), the approximation is poor. This is unavoidable, and users interested in such extremes or in the fundamental aspects of true fractional derivatives must take note of it. However, mismatch in extreme frequencies outside the range of interest for a particular model of a real material may have little engineering impact.

integer
fractional

The first term on the right-hand side of (2.1) is the contribution to the total energy from the initial conditions. The second term is the contribution from the source term. The third term is the contribution from the boundary conditions. The fourth term is the contribution from the initial energy. The fifth term is the contribution from the initial momentum. The sixth term is the contribution from the initial angular momentum. The seventh term is the contribution from the initial mass. The eighth term is the contribution from the initial charge. The ninth term is the contribution from the initial magnetic field. The tenth term is the contribution from the initial electric field. The eleventh term is the contribution from the initial gravitational field. The twelfth term is the contribution from the initial electromagnetic field. The thirteenth term is the contribution from the initial matter field. The fourteenth term is the contribution from the initial radiation field. The fifteenth term is the contribution from the initial scalar field. The sixteenth term is the contribution from the initial vector field. The seventeenth term is the contribution from the initial tensor field. The eighteenth term is the contribution from the initial spinor field. The nineteenth term is the contribution from the initial fermion field. The twentieth term is the contribution from the initial boson field.

2 Stochastic Origins

Let $x(t)$ be a stochastic process satisfying the stochastic differential equation

$$D^\alpha x(t) = \frac{d}{dt} \int_0^t \frac{x(\tau)}{t-\tau} d\tau,$$

where $0 < \alpha < 1$, Γ is a constant, and $x(0) = 0$.

$$\frac{d}{dt} \int_0^t \frac{\tau_+^{\alpha-1}}{t-\tau} d\tau = \frac{\pi \delta(t)}{\pi - \alpha [\Gamma - \alpha]},$$

where $\delta(t)$ is the Dirac delta function, $\tau_+ = \max(\tau, 0)$, and $\tau > 0$.

$$D^\alpha x(t) = h(t)$$

where $h(t)$ is a stochastic process satisfying the stochastic differential equation

$$\sigma(t) = E_1 D^\alpha \epsilon(t).$$

The power law decay

The exponential decay

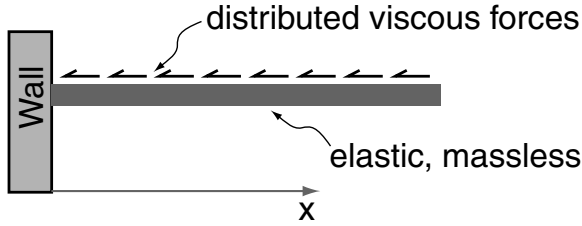


Fig. 1. One dimensional viscoelastic model.

$$\begin{aligned} & \text{The displacement } u(x, t) \text{ of the bar at position } x \text{ and time } t \text{ satisfies the wave equation} \\ & \frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial t^2} = 0, \quad 0 \leq x \leq L, \quad t \geq 0. \end{aligned}$$

$$b \frac{\partial u}{\partial x} \bigg|_{x=0} - c \frac{\partial u}{\partial t} \bigg|_{t=0} = 0, \quad u(0, t) = 0, \quad u_x(L, t) = 0.$$

$$\text{The initial conditions are } u(x, 0) = 0, \quad \frac{\partial u}{\partial t} \bigg|_{t=0} = 0.$$

$$u(x, t) = \sum_{i=1}^n a_i(t) \phi_i(x)$$

$$\begin{aligned} & \text{where } \phi_i(x) \text{ are the eigenfunctions of the problem} \\ & \phi_i'' + \lambda_i \phi_i = 0, \quad 0 \leq x \leq L, \quad \phi_i(0) = 0, \quad \phi_i'(L) = 0. \end{aligned}$$

$$B_{ij} = \int_0^L b \phi_{i,x} \phi_{j,x} dx, \quad C_{ij} = \int_0^L c \phi_i \phi_j dx, \quad a_i(t) = \int_0^L \phi_i u(x, t) dx.$$

$$Ca' = -Ba.$$

$$\text{where } \phi_i \text{ are the eigenfunctions of the problem}$$

$$a' = -Ba.$$

$$\begin{aligned} & \text{The matrix } B \text{ is symmetric and positive definite, and the matrix } C \text{ is symmetric and positive semi-definite.} \\ & \text{The matrix } A \text{ is the } n \times n \text{ matrix defined by } A_{ij} = \int_0^L \phi_i \phi_j dx. \end{aligned}$$

$$x' = -Bx.$$

$$\begin{aligned} & \left[\begin{array}{c} \vdots \\ x_0 \\ \vdots \end{array} \right] \left[\begin{array}{cc} B & x_0 \end{array} \right] \left[\begin{array}{c} \vdots \\ x_0 \\ \vdots \end{array} \right] \quad n \times n \\ & \left[\begin{array}{c} \vdots \\ x_0 \\ \vdots \end{array} \right] \left[\begin{array}{cc} B & x_0 \end{array} \right] \left[\begin{array}{c} \vdots \\ x_0 \\ \vdots \end{array} \right] \quad n \times n \end{aligned}$$

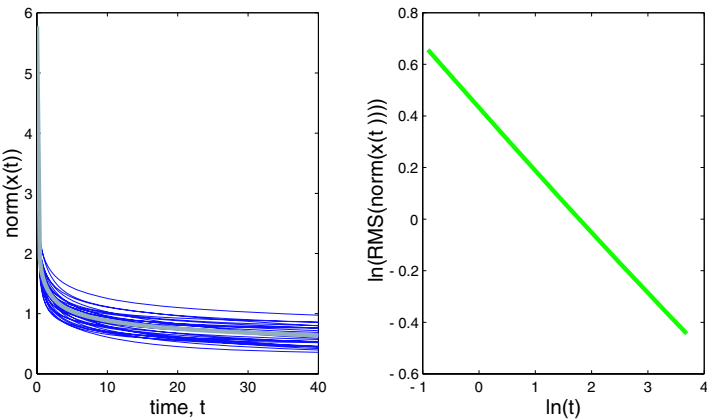


Fig. 2. Left: $\text{norm}(x) = \sqrt{x^T x}$ against time. 30 individual solutions (*thin lines*) as well as their RMS values (*thick gray*). Right: RMS value of $\text{norm}(x)$ against time is a *straight line* on a log-log scale. A *fitted line* has slope $-0.24 \approx -1/4$.

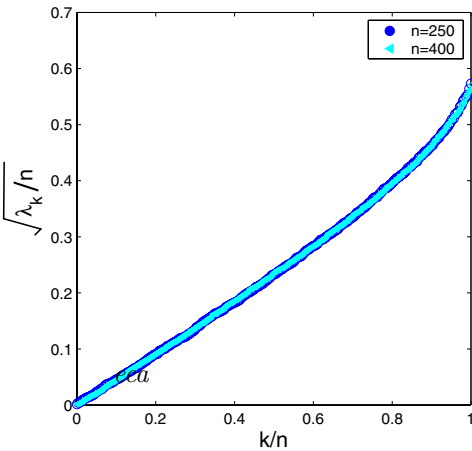


Fig. 3. Eigenvalues of B for $n = 250$ and 400 .

$$t^{-1/4} \left[\begin{array}{c} \vdots \\ x_0 \\ \vdots \end{array} \right] \left[\begin{array}{cc} B & x_0 \end{array} \right] \left[\begin{array}{c} \vdots \\ x_0 \\ \vdots \end{array} \right] \quad n \times n$$

$$\phi_{i+1}(\eta) = \begin{cases} \frac{\eta - p_{i-1}}{p_i - p_{i-1}}, & p_{i-1} \leq \eta \leq p_i, \\ \frac{p_{i+1} - \eta}{p_{i+1} - p_i}, & p_i \leq \eta \leq p_{i+1}, \end{cases} \quad i = 1, \dots, n-1,$$

$$p_0 = 0, \quad p_n = 1.$$

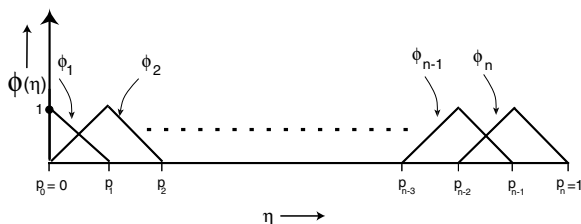


Fig. 4. Hat-shape functions.

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \text{with} \quad y_i = \begin{cases} \eta - p_{i-1} & \text{if } \eta \in [p_{i-1}, p_i] \\ p_{i+1} - \eta & \text{if } \eta \in [p_i, p_{i+1}] \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \text{with} \quad y_i = \begin{cases} \eta - p_{i-1} & \text{if } \eta \in [p_{i-1}, p_i] \\ p_{i+1} - \eta & \text{if } \eta \in [p_i, p_{i+1}] \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \text{with} \quad y_i = \begin{cases} \eta - p_{i-1} & \text{if } \eta \in [p_{i-1}, p_i] \\ p_{i+1} - \eta & \text{if } \eta \in [p_i, p_{i+1}] \\ 0 & \text{otherwise} \end{cases}$$

$$p_i = \frac{y_i^{1/\alpha}}{y_i^{1/\alpha}}, \quad i = 1, \dots, n-1.$$

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \text{with} \quad y_i = \begin{cases} \eta - p_{i-1} & \text{if } \eta \in [p_{i-1}, p_i] \\ p_{i+1} - \eta & \text{if } \eta \in [p_i, p_{i+1}] \\ 0 & \text{otherwise} \end{cases}$$

$$p_i = \frac{y_i^2}{y_i^2}, \quad i = 1, \dots, n-1.$$

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \text{with} \quad y_i = \begin{cases} \eta - p_{i-1} & \text{if } \eta \in [p_{i-1}, p_i] \\ p_{i+1} - \eta & \text{if } \eta \in [p_i, p_{i+1}] \\ 0 & \text{otherwise} \end{cases}$$

$$\int_0^1 \left(\sum_{i=1}^n \left\{ a_i(t) - \frac{\eta}{-\eta} a_i(t) \right\} \phi_i(\eta - x(t)) \right) \times \phi_m(\eta) \frac{\alpha \eta^{\alpha-1}}{-\eta^{1+\alpha}} d\eta \quad ,$$

5 Modeling Issues and Asymptotics

always be wrong.

experimentally indistinguishable

always

¹ A Maple-8 worksheet to compute the matrices A , B , and c is available on [9].

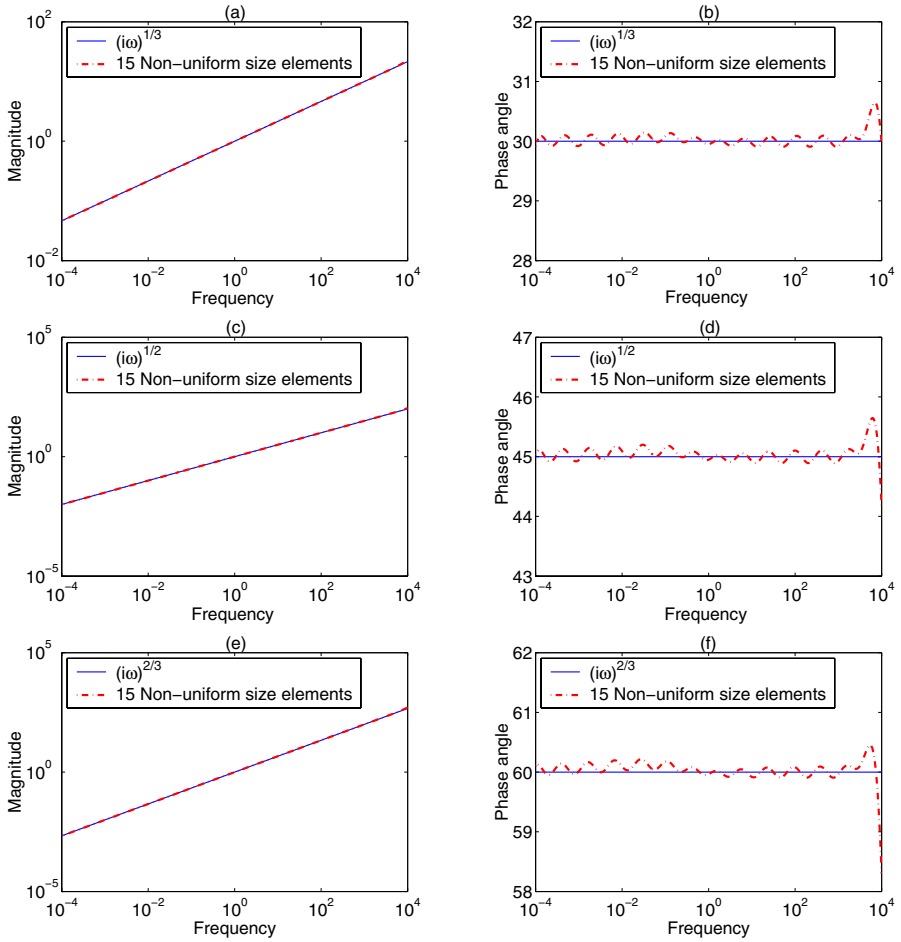


Fig. 5. Magnitude and phase angle comparison in FRFs. Plots (a) and (b): 15 nonuniform hat elements and $\alpha = 1/3$. Plots (c) and (d): 15 nonuniform hat elements and $\alpha = 1/2$. Plots (e) and (f): 15 nonuniform hat elements and $\alpha = 2/3$.

description
same

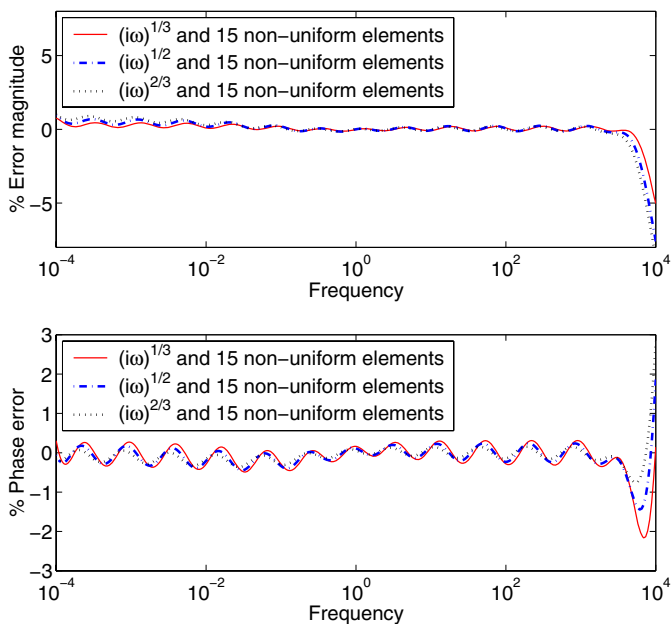


Fig. 6. Percentage errors in the magnitude and the phase angle for $\alpha = 1/3$, $\alpha = 1/2$ and $\alpha = 2/3$.

simulations
accurate enough

References

1. Vlad MO, Schöfnisch B, Mackey MC (1996) Phys. Rev. E 53(5):4703–4710.
2. Chatterjee A (2005) J. Sound Vib. 284:1239–1245.
3. Oustaloup A, Levron F, Mathieu B, Nanot F (2000) IEEE Trans. Circ. Syst. I: *Fundamental Theory and Applications* 47(1):25–39.
4. Chen Y, Vinagre BM, Podlubny I (2004) Nonlinear Dynamics 38:155–170.
5. Yuan L, Agrawal OP (2002) J. Vib. Acoust. 124:321–324.
6. Schmidt A, Gaul L (2006) Mech. Res. Commun. 33(1):99–107.
7. Singh SJ, Chatterjee A (2005) Nonlinear Dynamics (in press).
8. Finlayson BA (1972) *The Method of Weighted Residuals and Variational Principles*. Academic Press, New York.
9. http://www.geocities.com/dynamics_iisc/SystemMatrices.zip
10. Ford NJ, Simpson AC (2001) Numer. Algorithms 26:333–346.

ANALYTICAL MODELLING AND EXPERIMENTAL IDENTIFICATION OF VISCOELASTIC MECHANICAL SYSTEMS

Giuseppe Catania¹ and Silvio Sorrentino²

¹ DIEM, Department of Mechanics, University of Bologna, Viale del Risorgimento 2, 40136 Bologna, Italy; Tel: +39 051 2093447, Fax: +39 051 2093446, E-mail: giuseppe.catania@mail.ing.unibo.it.

² DIEM, Department of Mechanics, University of Bologna, Viale del Risorgimento 2, 40136 Bologna, Italy; Tel: +39 051 2093451, Fax: +39 051 2093446, E-mail: silvio.sorrentino@mail.ing.unibo.it.

Abstract

In the present study non-integer order or fractional derivative rheological models are applied to the dynamical analysis of mechanical systems. Their effectiveness in fitting experimental data on wide intervals of frequency by means of a minimum number of parameters is first discussed in comparison with classical integer order derivative models. A technique for evaluating an equivalent damping ratio valid for fractional derivative models is introduced, making it possible to test their ability in reproducing experimentally obtained damping estimates. A numerical procedure for the experimental identification of the parameters of the Fractional Zener rheological model is then presented and applied to a high-density polyethylene (HDPE) beam in axial and flexural vibrations.

Keywords

Fractional derivative, viscoelasticity, frequency response function, damping.

1 Introduction

The selection of an appropriate rheological model is a relevant problem when studying the dynamic behaviour of mechanical structures made of viscoelastic materials, like polymers for example. The selected model should be accurate in fitting the experimental data on a wide interval of frequencies, from creep and relaxation behaviour to high-frequency vibrations, by means of a minimum number of parameters. In particular, regarding vibrations, it should be able to reproduce the experimentally found behaviour of the damping ratio ζ_n as a function of the natural angular frequency ω_n [1].

In the present study some differential linear rheological models are considered, discussing their effectiveness in solving the above-mentioned problem, in relation to a high-density polyethylene (HDPE) beam in axial and flexural vibrations. Structural and hysteretic damping laws are not included in the analysis, since they lead to non-causal behaviour [2].

Classical integer order differential models are compared to fractional differential ones, which are considered to be very effective in describing the linear viscoelastic dynamic behaviour of mechanical structures made of polymers [3]. Extensive literature exists on this topic [4, 5, 6], the application of fractional calculus to viscoelasticity yielding physically consistent stress-strain constitutive relations with a few parameters, good curve fitting properties and causal behaviour [7].

Since with fractional derivative models the evaluation of closed form expressions of an equivalent damping ratio ζ_n does not seem an easy task, a different approach is proposed [8], based on the standard circle-fit technique [9].

When using fractional derivative models the solution of direct problems, i.e., the evaluation of time or frequency response from a known excitation can still be obtained from the equations of motion using standard tools such as modal analysis [10, 11, 12], but regarding the inverse problem, i.e., the identification from measured input-output vibrations, no general technique has so far been established, since the current methods do not seem to easily work with differential operators of non-integer order [1].

In the present study a frequency-domain method is thus proposed for the experimental identification of the fractional Zener model, also known as fractional standard linear solid [5], to compute the frequency-dependent complex stress-strain relationship parameters related to the material. The procedure is first applied to numerically generated frequency-response functions for testing its accuracy, and then to experimental inertance data.

2 Selection of a Rheological Model

In the present study the uniform, rectangular cross-section, straight axis HDPE beam shown in Fig. 1 is considered, Table 2 showing its geometrical parameters and Table 1 some HDPE material typical values [13].

Table 1. HDPE typical parameters

Average density	954 Kg×m ⁻³
Young's modulus	0.2 to 1.6 GPa

Table 2. Parameters of the beam

Material	HDPE
Density	$\rho = 1006.3 \text{ Kg}\times\text{m}^{-3}$
Length (x direction)	$L = 1000 \text{ mm}$
Thickness (z direction)	$h_z = 96.58 \text{ mm}$
Thickness (y direction)	$h_y = 24.14 \text{ mm}$
Cross-section area	$A = 2.332\times10^{-3} \text{ m}^2$
Section moment of inertia	$I_{zz} = 1.1328\times10^{-7} \text{ m}^4$
Section moment of inertia	$I_{yy} = 1.8125\times10^{-6} \text{ m}^4$
Total mass	$M = 2.346 \text{ Kg}$



Fig. 1. Experimental testing setup.

According to data available in the literature, an appropriate model for the HDPE beam should yield a creep compliance $J(t)$ (response to the unit stress step) reaching 95% of its final value after $100 \div 500$ s and a relaxation modulus $G(t)$ (response to the unit strain step) reaching 5% of its initial value after $10 \div 50$ s [13]. On the other hand, the same model should accurately fit the responses of the system under analysis (in the case considered herein, frequency-response functions), thus reproducing the experimentally found behaviour of the damping ratio ζ_n as a function of the natural angular frequencies ω_n , as shown for example in Fig. 2.

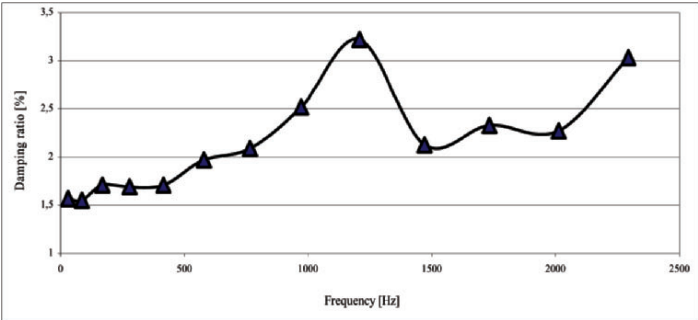


Fig. 2. Experimental damping ratio ζ_n versus natural frequency f_n .

Subsequently, several different integer order and non-integer order derivative rheological models, depicted in Fig. 3, are considered and compared, discussing their ability to satisfy the above mentioned requirements.

2.1 Integer order derivative models

The simplest real, causal, and linear viscoelastic model is the Kelvin–Voigt (Fig. 3a), whose constitutive equation is:

$$\sigma(t) = \left[E + C \frac{d}{dt} \right] \varepsilon(t) \quad (1)$$

yielding the creep compliance and the relaxation modulus:

$$\begin{cases} J(t) = \frac{1}{E} \{1 - \exp[-(t/\tau_\varepsilon)]\} \\ G(t) = E + C\delta(t) \end{cases}, \quad \tau_\varepsilon = \frac{C}{E}, \quad (2)$$

and the following expression for the damping ratio:

$$\zeta_n = \frac{C\omega_n}{2E}. \quad (3)$$

holding for free vibrations of uniform beams.

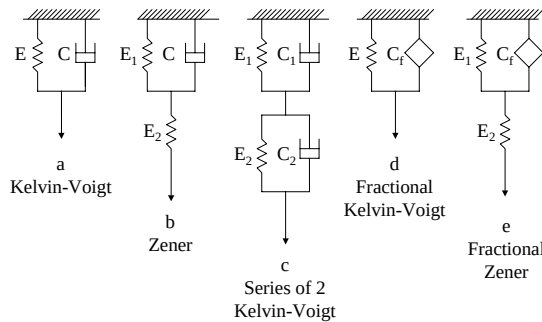


Fig. 3. Analogical models. (The Scott–Blair elements are represented by means of square symbols.)

Eq. (3) is incompatible with experimental results like those shown in Fig. 2, due to flexural vibrations of free-free HDPE beams. Assuming $E = 1.5 \times 10^9$ N/m for the HDPE static Young's modulus and $\zeta_n = 0.05$ at a frequency of 200 Hz,

Eq. (3) yields $C = 1.1937 \times 10^5$ Ns/m. The retardation time τ_e should thus be 8×10^{-3} s, meaning that the creep compliance would reach its steady state value (assumed to be at 95% of its asymptotic value) after less than 3×10^{-2} s, which is too short a time. The relaxation time according to the model should be null, in contradiction to experimental data [13].

The 3-parameter Zener (Fig. 3b) yields the following constitutive equation:

$$\left[1 + \frac{C}{E_1 + E_2} \frac{d}{dt}\right] \sigma(t) = \frac{E_2}{E_1 + E_2} \left[E_1 + C \frac{d}{dt}\right] \varepsilon(t) \quad (4)$$

In this case the creep and relaxation functions take the form:

$$\begin{cases} J(t) = \frac{E_1 + E_2}{E_1 E_2} \{1 - \exp[-(t / \tau_e)]\} \\ G(t) = \frac{E_1 E_2}{E_1 + E_2} \{1 + \exp[-(t / \tau_\sigma)]\} \end{cases}, \quad \tau_e = \frac{C}{E_1}, \tau_\sigma = \frac{C}{E_1 + E_2}, \quad (5)$$

Regarding the free vibrations of uniform beams, the following approximate expression for the damping ratio can be obtained:

$$\zeta_n = \frac{E_2}{2C\omega_n} \quad (6)$$

The experimental ζ_n values reported in Fig. 2 are also clearly incompatible with Eq. (6). Moreover, from Eq. (5), introducing $\tau_e = 100$ s and $\tau_\sigma = 10$ s yields $E_1 = C \cdot 0.01 \text{ s}^{-1}$ and $E_2 = C \cdot 0.09 \text{ s}^{-1}$, so that Eq. (6) at a frequency of 200 Hz yields $\zeta_n = 3.581 \times 10^{-5}$, clearly inconsistent with the experimental evidence.

To take into account both the “slow” and “fast” dynamical behaviour, the 4-parameter model obtained by a series of two Kelvin–Voigt elements and reported in Fig. 3c may be adopted, with $C_2 \ll C_1$. The constitutive equation is:

$$\left[1 + \frac{C_1 + C_2}{E_1 + E_2} \frac{d}{dt}\right] \sigma(t) = \frac{E_1 E_2}{E_1 + E_2} \left[1 + \left(\frac{C_1}{E_1} + \frac{C_2}{E_2}\right) \frac{d}{dt} + \frac{C_1 C_2}{E_1 E_2} \frac{d^2}{dt^2}\right] \varepsilon(t) \quad (7)$$

In this case the creep and relaxation functions take the form:

$$\begin{cases} J(t) = \frac{E_1 + E_2}{E_1 E_2} \left\{1 - \frac{\tau_{e1} \exp[-(t / \tau_{e1})] - \tau_{e2} \exp[-(t / \tau_{e2})]}{\tau_{e1} - \tau_{e2}}\right\} \\ G(t) = \frac{E_1 E_2}{E_1 + E_2} \{1 + \exp[-(t / \tau_\sigma)]\} \end{cases}, \quad \begin{cases} \tau_{e1} = \frac{C_1}{E_1}, \tau_{e2} = \frac{C_2}{E_2} \\ \tau_\sigma = \frac{C_1 + C_2}{E_1 + E_2} \end{cases} \quad (8)$$

Clearly, since $C_2 \ll C_1$, the relevant term in the creep compliance is $\tau_{\varepsilon 1}$. Regarding the free vibrations of uniform beams, the following approximate expression for the damping ratio can be obtained as well:

$$\zeta_n = \frac{1}{2} \left[\frac{E_2}{C_1 \omega_n} + \frac{C_2 \omega_n}{E_2} \right] \quad (9)$$

In comparison to the Zener model, in this case the parameter C_2 can take into account the “fast” dynamics, while it is not influential in the creep compliance and in the relaxation modulus. In the case of HDPE, a possible choice for the parameters is $E_1 = 1.6 \times 10^8$ N/m, $E_2 = 1.5 \times 10^9$ N/m, $C_1 = 1.6 \times 10^{10}$ Ns/m and $C_2 = 1.0 \times 10^6$ Ns/m, yielding a creep compliance which reaches 95% of its final value after 300 s and a relaxation modulus which reaches 5% of its initial value after 30 s. Regarding the modal damping ratio, the model is still not realistic, since by increasing the frequency the modal-damping ratio soon reaches too high values, in contrast with respect to the experimental evidence.

2.2 Non-integer order derivative models

A further enhancement can be obtained by taking into account models with constitutive equations defined through non integer order derivatives (or fractional derivatives, if the orders are assumed to be rational).

Replacing the first derivative (Newton element) with a fractional derivative (Scott–Blair element [5]) in the Kelvin–Voigt model yields the fractional Kelvin–Voigt:

$$\sigma(t) = \left[E + C_f \frac{d^\alpha}{dt^\alpha} \right] \varepsilon(t) \quad (10)$$

The equivalent analogical model is shown in Fig. 3d. The creep compliance and relaxation modulus become:

$$\begin{cases} J(t) = \frac{1}{E} \{ 1 - E_\alpha [-(t/\tau_\varepsilon)^\alpha] \}, \quad \tau_\varepsilon^\alpha = \frac{C_f}{E} \\ G(t) = E + C_f \frac{t^{-\alpha}}{\Gamma(1-\alpha)} \end{cases} \quad (11)$$

where E_α is the Mittag-Leffler function [5, 14], which plays the role of the exponential holding in the case of integer order derivatives.

The fractional Kelvin–Voigt can be found to perform very well in modelling both the “fast” dynamics and the creep behaviour, since for small values of the fractional derivative order the Mittag-Leffler function decreases very slowly.

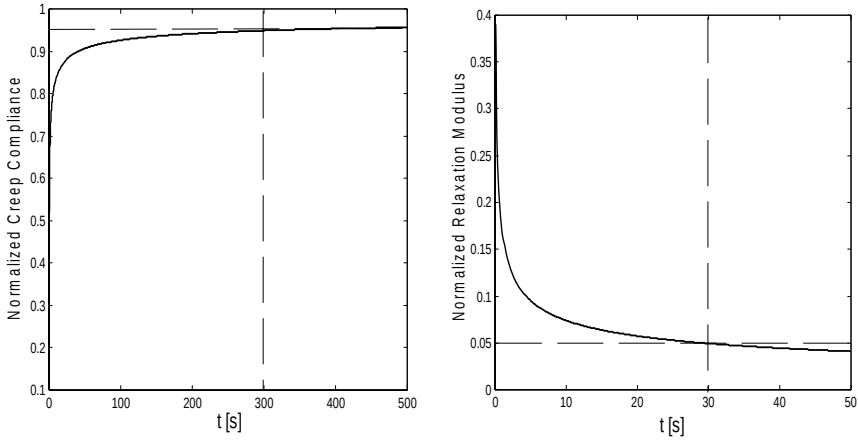


Fig. 4. Example of normalized creep compliance (left) and of normalized relaxation modulus (right).

Regarding the relaxation modulus, however, the results are worse, so another parameter is necessary to control the relaxation, yielding the fractional Zener model, Fig. 3e [5]:

$$\left[1 + \frac{C_f}{E_1 + E_2} \frac{d^\alpha}{dt^\alpha} \right] \sigma(t) = \frac{E_2}{E_1 + E_2} \left[E_1 + C_f \frac{d^\alpha}{dt^\alpha} \right] \varepsilon(t) \quad (12)$$

with creep compliance and relaxation modulus:

$$\begin{cases} J(t) = \frac{1}{E_2} + \frac{1}{E_1} \{ 1 - E_\alpha [-(t/\tau_\varepsilon)^\alpha] \} \\ G(t) = \frac{E_2}{E_1 + E_2} \{ E_0 + E_1 E_\alpha [-(t/\tau_\sigma)^\alpha] \} \end{cases}, \quad \begin{cases} \tau_\varepsilon^\alpha = \frac{C_f}{E_1} \\ \tau_\sigma^\alpha = \frac{C_f}{E_1 + E_2} \end{cases} \quad (13)$$

The creep compliance (normalized asymptotic value = 1) and relaxation modulus (normalized initial value = 1) computed using the parameters in Table 3 identified for HDPE are shown in Fig. 4. Since in the case of the fractional Zener model the evaluation of an approximate closed form expression of ζ_n does not seem an easy task, a different approach is proposed in the following section.

2.3 Evaluation of an equivalent damping ratio

The circle-fit identification technique for the damping ratio is based on the assumption that the Nyquist plot of the mobility $M_n(\omega)$ for any mode n is a circumference [9]. If this assumption is still acceptable when considering non-integer derivative models, then the circle fit can also be applied in such cases, taking into account that the physical meaning of the identified parameter ζ_n changes depending on the selected model.

The angle δ shown in Fig. 5 can be adopted to define a circle shape estimator. It can be expressed as:

$$\delta = \arctan\left(\frac{\text{Im}[M_n(\omega)]}{\text{Re}[M_n(\omega)]}\right) + \arctan\left(\frac{\text{Im}[M_n(\omega)]}{\text{Re}[M_n(\omega_0)] - \text{Re}[M_n(\omega)]}\right) \quad (14)$$

where ω_0 is the value of angular frequency for which $\text{Im}[M_n(\omega_0)] = 0$. If the Nyquist plot of $M_n(\omega)$ is a circumference, $\delta = \pi/2$ for every value of $\omega \neq \omega_0$.

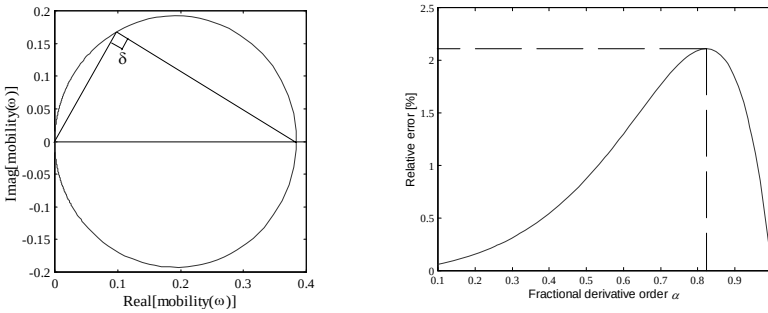


Fig. 5. Nyquist plot of the mobility, mode 3 (left) and maximum relative error in function of α , mode 3 (right).

The difference between the actual value $\delta(\omega)$ and $\pi/2$ thus provides a measure of the error made in approximating the Nyquist plot with a circumference. The absolute value of the relative error can then be expressed as follows:

$$\varepsilon(\omega) = \left| \frac{2\delta(\omega)}{\pi} - 1 \right| \quad (15)$$

Figure 5 show the maximum of $\varepsilon(\omega)$ in function of α . Similar results can be obtained for the other parameters of the fractional Zener model. The maximum

absolute error is in any case very small, which means that the approximation of the Nyquist plots of the mobility with circumferences is perfectly acceptable. The circle-fit technique can thus be adopted as a tool for estimating the ability of a non-integer derivative model to fit experimental measurements in a given frequency range.

3 Experimental Identification

When considering homogeneous free-free beams in flexural or axial vibration, the receptance can be written in general form as:

$$H(x, x_f; \omega) = \sum_{n=1}^{\infty} \frac{\phi_n(x)\phi_n(x_f)}{M[E(\omega)\chi_n - \omega^2]} = \sum_{n=1}^{\infty} \frac{\text{Res}_n}{E(\omega)\chi_n - \omega^2} \quad (16)$$

where M is the total mass of the beam, E is the material Young's modulus, n is the mode order, χ_n is a modal parameter, ϕ_n is the normalized eigenfunction, ω is the angular frequency, x_f and x are the force and response points respectively [15]. The internal dissipation can be modelled replacing the real valued Young's modulus in the modal stiffness by its complex representation [7]. In the case of the fractional Zener model, it can be expressed as:

$$E(\omega) = \frac{E_0 + a(i\omega)^\alpha}{1 + b(i\omega)^\alpha} \quad (17)$$

where i is the imaginary unit.

Under the assumption of well-separated modes (which often holds true for beams in axial or flexural vibrations), $E(\omega)$ can be identified from Eq. (16) writing the following equation for the four unknown parameters E_0 , a , b and α :

$$\begin{aligned} \xi E_0 + \xi(i\omega)^\alpha a - (i\omega)^\alpha b &= 1 \\ \xi &= \frac{H_n(\omega)\chi_n}{\omega^2 H_n(\omega) + \text{Res}_n} \end{aligned} \quad (18)$$

valid for the n th mode in a neighbourhood of its natural frequency.

Assuming a trial value for the fractional derivative order α , and evaluating Eq. (18) in correspondence with different modes, yields a linear system with complex coefficient matrix $\mathbf{A}$. In order to ensure the reality and causality of the model, the constitutive parameters must be real and the system can thus be written in the form:

$$\begin{bmatrix} \text{Re}[\mathbf{A}] \\ \text{Im}[\mathbf{A}] \end{bmatrix} \mathbf{y} = \begin{bmatrix} \mathbf{1} \\ \mathbf{0} \end{bmatrix} \Leftrightarrow \mathbf{A}_r \mathbf{y} = \mathbf{d}_r, \quad \mathbf{y} = [E_0 \ a \ b]^T \quad (19)$$

The solution can be computed using the singular value decomposition algebraic technique, paying attention to the ill-conditioning of the system Eq. (19). This latter problem can be solved by normalizing each variable with respect to the quadratic norm of the corresponding column of the matrix $\mathbf{A}_r$.

An error estimate may help in identifying the optimal solution with respect to the fractional derivative order α . Different expressions for the error estimate can be given in the form:

$$\text{err}_i = \frac{\sqrt{\boldsymbol{\Psi}_i^T \boldsymbol{\Psi}_i}}{N}, \quad i = 1 \dots 4, \quad \begin{cases} \boldsymbol{\Psi}_1 = \mathbf{A}\mathbf{y} - \mathbf{1}; \\ \boldsymbol{\Psi}_2 = \mathbf{A}_r\mathbf{y} - \mathbf{d}_r \\ \boldsymbol{\Psi}_3 = \text{Re}[\mathbf{A}]\mathbf{y} - \mathbf{1} \\ \boldsymbol{\Psi}_4 = \text{Im}[\mathbf{A}]\mathbf{y} - \mathbf{1} \end{cases} \quad (20)$$

where N is the number of equations.

The above-described method generally proves both numerically stable and accurate, even if noise is added to numerically generated frequency response functions. Fig. 6 shows the error functions due to the identification of a complex Young's modulus, $E_0 = 1.2 \times 10^9 \text{ Nm}^{-2}$, $a = 1 \times 10^7 \text{ Nm}^{-2} \text{ s}^\alpha$, $b = 10^{-3} \text{ s}^\alpha$ and $\alpha = 0.3$, using numerically generated data and added white noise with amplitude $E_0 \times 10^{-3}$.

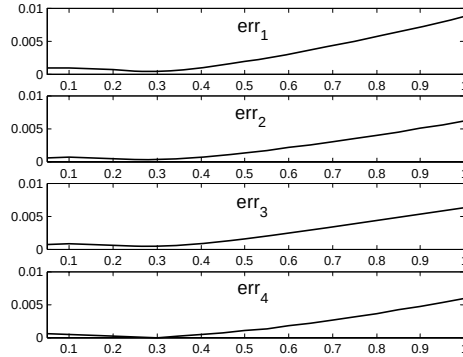


Fig. 6. Errors err_1 , err_2 , err_3 , and err_4 versus the fractional derivative order α .

4 Application

The uniform, rectangular cross section, straight-axis HDPE beam shown in Fig. 1 was tested with respect to flexural and axial free vibration. Restraining the beam by means of flexible, rubber-made couplings to the frame approximates the free-free boundary conditions. The adopted global Cartesian reference frame has its

main direction x along the axis of the beam and directions y, z along the cross-section principal axes of inertia. The selected experimental degrees of freedom (d.o.f.s) are chosen in correspondence with the three displacement components along x, y, z with respect to 11 equally spaced ($\Delta x = 0.1$ m) points on the axis of the beam. The system was excited by means of an instrumented ICP hammer in correspondence with the d.o.f.s x_f and acceleration responses were evaluated by means of miniaturized ICP piezoelectric accelerometers in correspondence with the d.o.f.s x_r .

The frequency response functions (inertances) were estimated by means of the H1 technique [9], 25 averages, with respect to all the combinations of excitation-response in the same vibrational condition (axial, flexural x - y and flexural x - z), using a rectangular force window, without response windowing prior acquisition. The adopted sampling frequency is $f_s = 51200$ Hz, with $N = 2^{16}$ samples, with acquisition time $T = 1.28$ s and frequency resolution $\Delta f = 0.78125$ Hz. The data were acquired by means of a DSP VXI Agilent 16 channel acquisition card, using the MTS-Ideas Test Software to interface the hardware. The coherence was very good up to about 2000 Hz, and the linearity of the system was also checked by comparing the frequency response functions obtained by swapping the force and response d.o.f.s.

Table 3. Identified parameters

Fractional derivative order	$\alpha = 0.358$
Complex impedance parameters	$E_0 = 1.358 \times 10^9 \text{ Nm}^{-2}$
	$a = 7 \times 10^6 \text{ Nm}^{-2} \text{ s}^\alpha$
	$b = 1.5 \times 10^{-5} \text{ s}^\alpha$

Table 3 shows the identified constitutive parameters using the previously described method. Since the functions err_i defined in the previous section do not seem to suggest a clear indication for the fractional derivative order when dealing with experimental data, the selected value for α is the one which yields creep retardation ($\tau_\epsilon = 300$ s) and relaxation ($\tau_\sigma = 30$ s) times according to [5], as shown in Fig. 4, representing the theoretical creep compliance and relaxation modulus of the material. Fig. 7 compares some estimated flexural x - y inertance functions (*thick lines*, according to the parameters of Table 3) with experimental data (*thin lines*). Good agreement can be found in the frequency interval 0-1000 Hz.

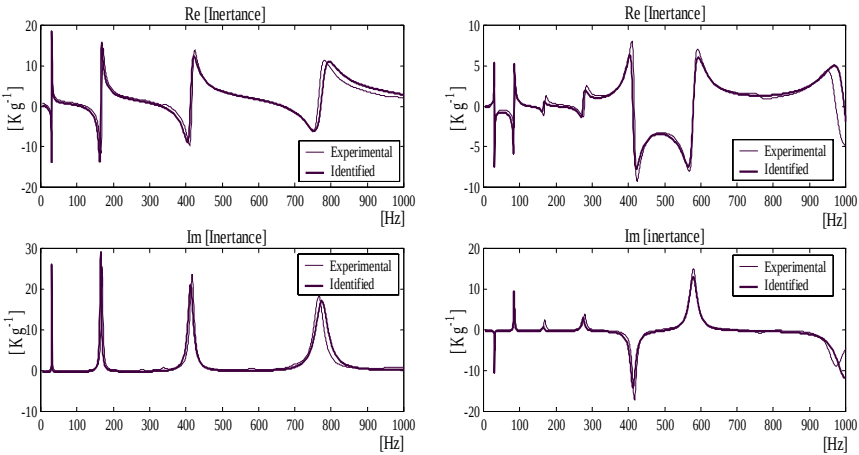


Fig. 7. Inertance: $x_f = 0.5$ m, $x_r = 0.5$ m (left); $x_f = 0.9$ m, $x_r = 0.3$ m (right).

Figure 8 (left) shows the theoretical plot of the equivalent modal damping ratio with respect to natural frequency computed through the circle fit for flexural vibration using the identified parameters of Table 3. It appears that this parameter slowly increases with respect to frequency, with good agreement with identified experimental results. Figure 8 (right) compares the previous result with that related to a series of 2 Kelvin–Voigt elements, with the same parameters as the ones discussed in the introduction, except for C_2 , which is slightly reduced in order to obtain the same value of ζ_n for the 5th mode for both models. This latter model exhibits an unrealistic behaviour beyond the frequency of 250 Hz.

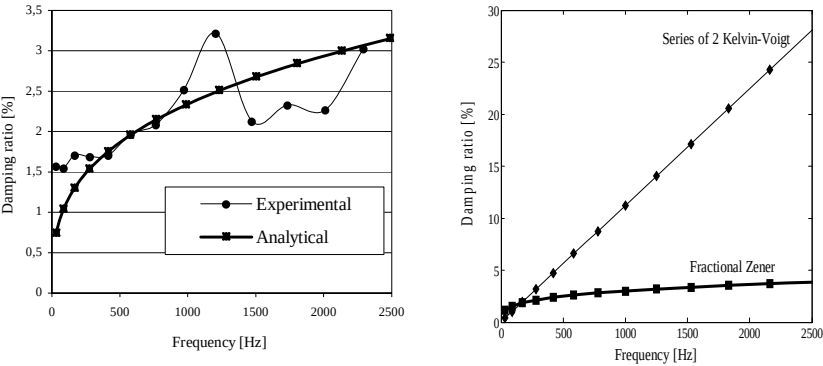


Fig. 8. Damping ratio ζ_n [%] versus natural frequency f_n [Hz].

5 Discussion

The capability of the fractional Zener model to accurately fit experimental data from both creep-relaxation and vibration tests was outlined herein. The equivalent damping ratio versus natural frequency functional relationship can also be modelled with good accuracy, globally matching most experimentally obtained damping estimates. On the other hand, standard integer-order derivative models such as the series of 2 Kelvin–Voigt elements seem to lack this feature, since they exhibit a linear relationship between ζ_n and f_n in the high-frequency range.

Different assumptions for the fractional derivative order do not seem to affect the error estimator, meaning that different solutions $\mathbf{y}$ are equivalent with respect to the identification problem from vibrations. Introducing the creep retardation and relaxation times (equal to experimental known values) as two further constraints, a single optimal fractional derivative order can be obtained, also yielding the optimal $\mathbf{y}$ choice.

Regarding the comparison of experimental and estimated frequency-response functions, the results are good and consistent over the frequency interval 0–1000 Hz. Beyond the frequency of 1000 Hz the estimated frequency-response functions do not seem to show good agreement with respect to experimental data. This could mean that the simple fractional Zener model is still not able to fit the experimental data whenever high frequencies are concerned, and a different model, adopting more parameters, should be investigated as well.

6 Conclusions and Future Work

The complex Young's modulus of a homogeneous beam made from HDPE was identified according to the fractional Zener model using frequency-domain experimental data.

The numerical stability and the accuracy of the adopted technique were successfully tested considering numerically generated frequency-response functions with addition of noise.

Creep retardation and relaxation times obtained from the complex Young's modulus identified parameters are in agreement with those available in the literature while analytically evaluated FRFs also match experimental estimates over the frequency range 0–1000 Hz. A procedure for estimating an equivalent damping ratio was successfully adopted for testing the identified model in reproducing experimental damping estimates.

Future work will be devoted to developing a global, more general multiple degree of freedom (MDOF) identification technique, and new material constitutive models as well, in order to extending the identification with respect to the high-frequency range.

Acknowledgments

ASME permission to publish this paper, a modified and upgraded version of paper DETC 2005-85725, is kindly acknowledged.

References

1. Catania G, Sorrentino S (2005) Experimental identification of a fractional derivative linear model for viscoelastic materials, Long Beach, California, Proceedings of IDETC/CIE 2005 (DETC 2005-85725).
2. Frammartino D (2000) Modelli analitici evoluti per lo studio di sistemi smorzati, Degree thesis, Politecnico di Torino (in Italian).
3. Jones DG (2001) *Handbook of Viscoelastic Vibration Damping*. Wiley, New York.
4. Caputo M, Mainardi F (1971) Linear models of dissipation in anelastic solids, *Rivista del Nuovo Cimento* 1:161–198.
5. Mainardi F (1997) Fractional calculus: some basic problems in continuum and statistical mechanics, in: *Fractals and Fractional Calculus in Continuum Mechanics*. Springer, New York.
6. Beyer H, Kempfle S (1995) Definition of physically consistent damping laws with fractional derivatives, *Zeitschrift für Angewandte Mathematik und Mechanik* 75:623–635.
7. Gaul L (1999) The influence of damping on waves and vibrations, *Mech. Syst. Signal Process.* 13:1–30.
8. Catania G, Sorrentino S (2006) Fractional derivative linear models for describing the viscoelastic dynamic behaviour of polymeric beams, Saint Louis, Missouri, MO Proceedings of IMAC 2006.
9. Ewins DJ (2000) *Modal Testing: Theory, Practice and Application*, 2nd edition, Research Studies Press Baldock, UK.
10. Bagley RL, Torvik PJ (1983) Fractional calculus: a different approach to the analysis of viscoelastically damped structures, *AIAA J.* 21:741–748.
11. Sorrentino S (2003) Metodi analitici per lo studio di sistemi vibranti con operatori differenziali di ordine non intero, PhD thesis, Politecnico di Torino.
12. Sorrentino S, Garibaldi L (2004) Modal analysis of continuous systems with damping distributions defined according to fractional derivative models, Leuven (Belgium), Proceedings of Noise and Vibration Engineering Conference (ISMA 2004).
13. McCrum NG, Buckley CP, Bucknall CB (1988) *Principles of Polymer Engineering*, Oxford University Press, Oxford.
14. Miller KS, Ross B (1993) *An Introduction to the Fractional Calculus and Fractional Differential Equations*, Wiley, New York.
15. Timoshenko S, Young DH (1955) *Vibrations Problems in Engineering*, 3rd edition, Van Nostrand New York.

Part 7

Control

LMI CHARACTERIZATION OF FRACTIONAL SYSTEMS STABILITY

Mathieu Moze, Jocelyn Sabatier, and Alain Oustaloup

LAPS-UMR 5131 CNRS, Université Bordeaux 1, – ENSEIRB, 351 cours de la Libération, F33405 TALENCE Cedex, France; Tel: +33 (0)540 006 607, E-mail: firstname.name@laps.u-bordeaux1.fr

Abstract

The notions of linear matrix inequalities (LMI) and convexity are strongly related. However, with state-space representation of fractional systems, the stability domain for a fractional order ν , $0 < \nu < 1$, is not convex. The classical LMI stability conditions thus cannot be extended to fractional systems. In this paper, three LMI-based methods are used to characterize stability. The first uses the second Lyapunov method and provides a sufficient but nonnecessary condition. The second and new method provides a sufficient and necessary condition, and is based on a geometric analysis of the stability domain. The third method is more conventional but involves nonstrict LMI with a rank constraint.

Keywords

LMI, fractional systems, stability.

1 Introduction

Fractional differentiation is now a well-known tool for controller synthesis. Several presentations and applications of the fractional PID controller [1–4], and the CRONE controller [5] demonstrate their efficiency. Fractional differentiation also permits a simple representation of some high-order complex integer systems [6]. Consequently, basic properties of fractional systems have been investigated these last 10 years and criteria and theorems are now available in the literature concerning stability [7, 8], observability [7], and controllability [7] of fractional systems.

Whereas Lyapunov methods have been developed for stability analysis and control law synthesis of integer linear systems [9] and have been extended to deal with more complex systems such as nonlinear, linear time-varying (LTV), and linear parameter-varying (LPV), only few studies deal with Lyapunov stability of fractional systems, and synthesis of control laws for such systems is

almost exclusively done in the frequency domain [5]. However, using Lyapunov stability conditions or quadratic robust control problems [10, 11] defined by linear matrix inequalities (LMI) would permit to use efficient numerical methods based on convex optimization [12].

This situation could be well explained by the fact that, at the contrary of fractional systems where fractional order $1 < \nu < 2$, for which a new proof of the extended Matignon stability theorem is proposed in section 3 along with a LMI stability condition, the stability domain fractional systems where $0 < \nu < 1$ is not a convex set of the complex plane and a LMI stability condition cannot be directly derived in this case.

LMI stability conditions for fractional systems are however proposed in this paper. After some definitions, a stability condition proposed by [13] is studied. It is shown that this condition is nonnecessary and thus of limited interest. Then, two other LMI conditions for $0 < \nu < 1$ are proposed, one based on a geometric transformation of the stability domain, and the other based on the characterization of the unstability domain.

2 Fractional Calculus

Riemann–Liouville fractional differentiation definition is used in this paper. The fractional integral of a function $f(t)$ is thus defined by [14]

$$I_0^\nu f(t) = \frac{1}{\Gamma(\nu)} \int_0^t (t - \xi)^{\nu-1} f(\xi) d\xi, \quad (1)$$

where $\nu \in \mathbb{R}^+$ denotes the fractional integration order, and where

$$\Gamma(\nu) = \int_0^\infty e^{-x} x^{\nu-1} dx. \quad (2)$$

The order ν fractional derivative of a function f , $\nu \in \mathbb{R}^+$, can consequently be defined by [15]:

$$D^\nu f(t) = D^m [I^{m-\nu} f(t)], \quad (3)$$

where m is the smallest integer that exceeds ν .

3 Stability of Fractional Systems: Theorems and Definitions

Let consider a fractional system S_1 whose input signal $u(t) \in \mathbb{R}$ and output signal $y(t) \in \mathbb{R}$ are linked by the fractional differential equation:

$$\sum_{i=0}^M b_i (D^\nu)^i y(t) = \sum_{i=0}^N a_i (D^\nu)^i u(t), \quad \nu = \frac{m}{q}, \quad (m, q) \in \mathbb{N}^2, \quad (4)$$

where $N < M$, $(N, M) \in \mathbb{N}^2$ and all the differentiation orders are multiples of a commensurate order ν .

Assuming that system S_1 is relaxed at $t=0$, so the Laplace transforms of $D^\alpha u(t)$ and of $D^\alpha y(t)$ are respectively considered as $s^\alpha U(s)$ and $s^\alpha Y(s)$ for any $\alpha \in \mathbb{R}$, transfer function $F(s)$ corresponding to differential equation (4) is

$$F(s) = \frac{Y(s)}{U(s)} = \frac{\sum_{i=0}^N a_i (s^\nu)^i}{\sum_{i=0}^M b_i (s^\nu)^i}, \quad (5)$$

where $Y(s)$ and $U(s)$ are respectively the Laplace transforms of $y(t)$ and $u(t)$.

Given the commensurate order hypothesis, system S_1 also admits the state-space representation:

$$(S_1): \begin{cases} D^\nu x(t) = Ax(t) + Bu(t) \\ y(t) = Cx(t) \end{cases} \quad (6)$$

where $A \in \mathbb{R}^{M \times M}$, $B \in \mathbb{R}^{M \times 1}$, and $C \in \mathbb{R}^{1 \times M}$.

Stability analysis of system S_1 was investigated by Matignon who stated the following theorem for $0 < \nu < 1$.

Theorem 1: [9] *Autonomous system:*

$$D^\nu x(t) = Ax(t), \quad \text{with } x(t_0) = x_0 \quad \text{and} \quad 0 < \nu < 1, \quad (7)$$

is asymptotically stable if and only if $|\arg(\text{spec}(A))| > \nu \frac{\pi}{2}$, where $\text{spec}(A)$ is the set of all eigenvalues of A . Also, state vector $x(t)$ decays towards 0 and meets the following condition: $\|x(t)\| < Nt^{-\nu}$, $t > 0$, $\nu > 0$.

□

Exponential stability thus cannot be used to characterize asymptotic stability of fractional systems. A new definition must be introduced.

Definition: $t^{-\gamma}$ stability

Trajectory $x(t) = 0$ of system $d^\nu / dt^\nu = f(t, x(t))$ is $t^{-\gamma}$ asymptotically stable if the uniform asymptotic stability condition is met and if there is a positive real γ such that :

$$\forall \|x(t_0)\| \leq c, \exists Q(x(t_0)) \text{ such that } \forall t \geq t_0, \|x(t)\| \leq Q t^{-\gamma}.$$

□

$t^{-\gamma}$ stability will thus be used to refer to the asymptotic stability of fractional systems.

As the components of the state $x(t)$ slowly decay towards 0 following $t^{-\nu}$, fractional systems are sometimes called *long memory systems*.

An extension of theorem 1 to the case $1 < \nu < 2$ is given in [16]. A new proof for this theorem is now proposed.

Theorem 2: System (7) is asymptotically stable if and only if $|\arg(\text{spec}(A))| > \nu \frac{\pi}{2}$, when $1 < \nu < 2$.

□

Proof: Any transfer function given by:

$$F(s) = \frac{T(s^\nu)}{R(s^\nu)} \quad (8)$$

can be rewritten as

$$F(s) = \frac{T'(s^{\nu'})}{R'(s^{\nu'})}, \quad (9)$$

where $\nu' = \frac{\nu}{2}$ and $\frac{1}{2} < \nu' < 1$.

For instance the denominator of $F(s)$ given by (5) is:

$$R(s^\nu) = \prod_{i=1}^M (s^\nu - s_{vi}), \quad (10)$$

which can be rewritten as

$$R'(s^{\nu/2}) = \prod_{i=1}^M (s^{\nu/2} - s_{vi}^{1/2}) (s^{\nu/2} + s_{vi}^{1/2}), \quad (11)$$

or

$$R'(s^{\nu/2}) = \prod_{i=1}^M (s^{\nu/2} - s_{vi}^{1/2}) (s^{\nu/2} - e^{j\pi} s_{vi}^{1/2}). \quad (12)$$

Note that for every value of s_ν satisfying $R(s_\nu) = 0$, two values $s_{\nu'}$ verifying $R'(s_{\nu'}) = 0$ arise:

- One such that $\arg(s_{\nu'}) = \frac{\arg(s_\nu)}{2}$
- And another one such that $\arg(s_{\nu'}) = \pi + \frac{\arg(s_\nu)}{2}$

As $0 < \nu' < 1$, stability theorem 1 given by Matignon is applicable and it can be easily checked that:

$$\left| \frac{\arg(s_\nu)}{2} \right| < \nu' \frac{\pi}{2} \text{ and } \left| \pi + \frac{\arg(s_\nu)}{2} \right| < \nu' \frac{\pi}{2} \text{ if and only if } |\arg(s_\nu)| < \nu \frac{\pi}{2}.$$

System described by (8) is hence asymptotically stable if and only if $|\arg(\text{spec}(A))| > \nu \frac{\pi}{2}$, where $1 < \nu < 2$ and A denotes the state transition matrix of its state-space representation. □

Figure 1 shows the stability domain $\mathcal{D}_S$ of a fractional system depending on its differentiation order ν and on the value of $|\arg(\text{spec}(A))|$.

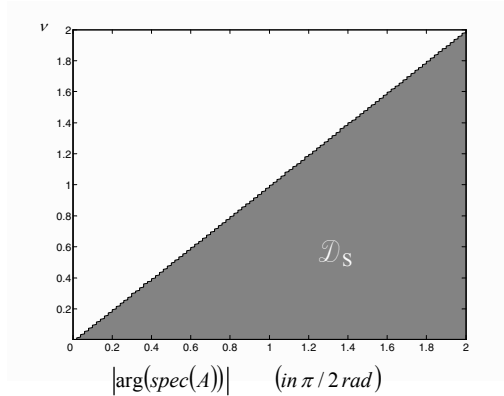


Fig. 1. Stability domain $\mathcal{D}_S$ (■) depending on fractional order ν and $|\arg(\text{spec}(A))|$.

4 Stability of Fractional Systems: LMI Characterization Issues

4.1 LMI and fractional systems

A LMI has the form [17]

$$F(x) = F_0 + \sum_{i=1}^m x_i F_i > 0, \quad (13)$$

where $x \in \mathbb{R}^m$ is the variable and where the symmetric matrices $F_i = F_i^T \in \mathbb{R}^{n \times n}$, $i = 0, \dots, m$ are known.

LMI have played an important role in control theory since the early 1960s due to this particular form. Actually a lot of matters arising in control theory can be expressed in terms of convex optimization problems involving LMI, for which efficient algorithms (interior point methods) have been developed such as [18] or [19].

The main issue when dealing with LMI is the convexity of the optimization set. Briefly, a set is said to be convex if for any two points belonging to the set, the line joining them is also contained in the set [19]. Figure 2 shows the stability domain of a fractional system for two different values of fractional order ν .

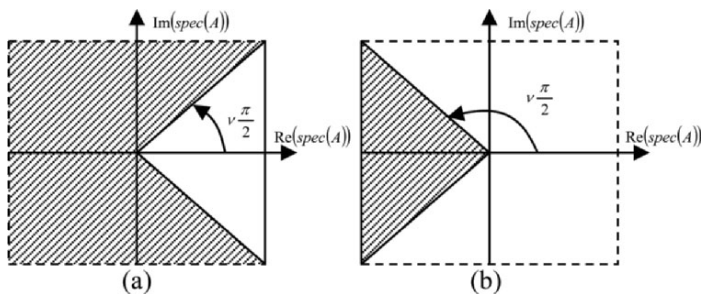


Fig. 2. Stability domain (▨) for: (a) $0 < \nu < 1$, (b) $1 \leq \nu < 2$.

As the stability domain of a fractional system with order $1 \leq \nu < 2$ is a convex set, various LMI methods for defining such a region have already been developed. The placement of the eigenvalues of a given matrix in an angular sector of the left-half complex plane needs indeed to be verified. Hence a LMI-based theorem for the stability of a fractional system with order $1 \leq \nu < 2$ can be formulated as follows.

Theorem 3: *A fractional system described by (6) with order $1 \leq \nu < 2$ is asymptotically stable if and only if there exists a matrix, $P \in \mathbb{R}^{M \times M}$, such that*

$$\begin{pmatrix} (A^T P + P A) \sin\left(\nu \frac{\pi}{2}\right) & (A^T P - P A) \cos\left(\nu \frac{\pi}{2}\right) \\ (P A - A^T P) \cos\left(\nu \frac{\pi}{2}\right) & (A^T P + P A) \sin\left(\nu \frac{\pi}{2}\right) \end{pmatrix} < 0.$$

□

Proof: This result is directly based on the methods described in [17] which provide rigorous proof for characterization of convex domains using LMI.

□

When $0 < \nu < 1$, Fig. 2a indicates that the stability domain is not convex. Its characterization using LMI can thus not be directly derived. Following parts introduce three methods to obtain such LMI.

5 Fractional Stability Theorem Based on an Algebraic Transformation of the System

5.1 Lyapunov stability

The success of LMIs in control theory is mainly due to the efficiency and the particular form of the Lyapunov second method that can be summarized by the following statement.

Autonomous system

$$\dot{x}(t) = Ax(t), \quad x(t_0) = x_0, \quad (14)$$

is asymptotically stable if a function $V(x(t)) > 0$ exists such that $\frac{d}{dt}V(t) < 0, \forall t \geq t_0$.

If function $V(t) = x^T(t)Px(t)$, $P > 0$, $P \in \mathbb{R}^{M \times M}$, is used to demonstrate the stability, the following theorem can be stated [17].

Theorem 4: *Integer order system of relation (14) is exponentially stable iff there is a positive definite matrix $P \in \mathbb{S}$, where $\mathbb{S}$ denotes the set of symmetric matrices, such that $A^T P + P A < 0$.*

□

Note that theorem 4 is satisfied if and only if the eigenvalues of A lie in the left-half complex plane.

5.2 Equivalent integer order system

The following stability condition is from [13]. In order to apply Lyapunov method on fractional systems and therefore to extend it to $0 < \nu \leq 1$, an equivalent integer order system is now derived from the autonomous system described by (7).

Laplace transform of (7) is given by [15]:

$$s^\nu X(s) = AX(s) + I^{1-\nu} x_0, \quad (15)$$

where $X(s)$ is the Laplace transform of $x(t)$.

Note that for $\nu = \frac{m}{q}$, $(m, q) \in \mathbb{N}^2$,

$$s^m X(s) = s^{m-\nu} s^\nu X(s), \quad (16)$$

and successively substituting $s^\nu X(s)$ using relation (3) leads to

$$s^m X(s) = A^q X(s) + \sum_{i=1}^q A^{i-1} I^{1-\nu} x_0 s^{m-i\nu}. \quad (17)$$

Inverse Laplace transform of (17) is given by

$$\left(\frac{d}{dt}\right)^m x(t) = A^q x(t) + \sum_{i=1}^q A^{i-1} I^{1-\nu} x_0 \delta^{(m-i\nu)}, \quad (18)$$

or, using a state-space description, by

$$\begin{cases} \dot{z}(t) = A_f z(t) + B_f \Delta(t) \\ x(t) = C_f z(t) \end{cases}, \quad (19)$$

where

$$A_f = \begin{pmatrix} 0 & \cdots & 0 & A^{\frac{1}{\nu}} \\ A^{\frac{1}{\nu}} & & 0 & 0 \\ & \ddots & & \vdots \\ 0 & & A^{\frac{1}{\nu}} & 0 \end{pmatrix}, \quad B_f = \begin{pmatrix} 1 & A & \cdots & A^{q-1} \\ & 0 & & \end{pmatrix},$$

$$C_f = (0 \quad \cdots \quad 0 \quad 1) \text{ and } \Delta^T(t) = I^{1-\nu} x_0 (\delta^{m-\nu} \quad \delta^{m-2\nu} \quad \cdots \quad \delta).$$

Lyapunov's second method can now be applied to integer order system (19) to determine stability of fractional system (7). Using theorem 4 with matrix A_f , the following theorem can be stated (note that $A_f < 0$ if and only if $A^{1/\nu} < 0$).

Theorem 5: [13] (sufficient condition) Fractional system (7) is $t^{-\gamma}$ stable if matrix $P > 0$, $P \in \mathbb{R}^{M \times M}$, exists, such that $\begin{pmatrix} \frac{1}{A^\nu} \end{pmatrix}^T P + P \begin{pmatrix} \frac{1}{A^\nu} \end{pmatrix} < 0$.

□

Proof: See steps above. Also in [13].

□

5.3 Validity of the stability condition

Figure 3 presents stability domain $\mathcal{D}_S'$ of a fractional system characterized using theorem 5 according to fractional order ν and to $|\arg(\text{spec}(A))|$. A comparison between Fig. 3 and Fig. 1 reveals that the entire stability domain is not identified using theorem 5. It therefore leads to a sufficient but nonnecessary condition.

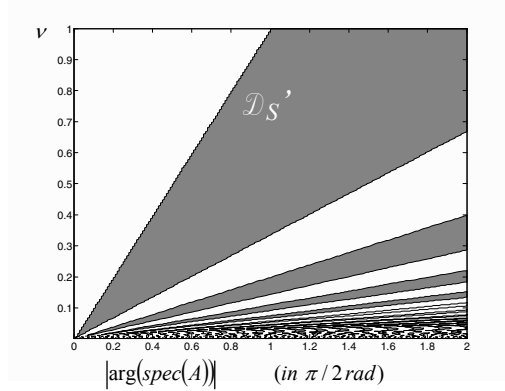


Fig. 3. Stability domain $\mathcal{D}_S'$ (■) determined by criterion 2 according to the values of ν and $|\theta|$.

A simple explanation can be given. Systems (7) and (19) have strictly the same behavior. However, transformations given by relations (16) to (18) produce a matrix A_f whose eigenvalues are outside the left-half complex plane. The unstable modes thus created are compensated by zeros produced by matrix $B_f \Delta(t)$ thus leading to a stable response to nonzero initial conditions. Due to such a situation, a method based on eigenvalue analysis of matrix A_f can only produce pessimistic stability conditions.

In order to analyze such a conservatism, let θ_f be an arguments of an eigenvalue of matrix $A^{1/\nu}$ and θ be the one of system (6) state transition matrix A of system (6). Line (D) in Fig. 4 represents the function F_ν that associates $|\theta|$ to $|\theta|_f$ according to ν :

$$F_\nu : \begin{matrix} [0, \pi] \rightarrow [0, 2\pi] \\ x \rightarrow \frac{1}{\nu} x \end{matrix} \quad (20)$$

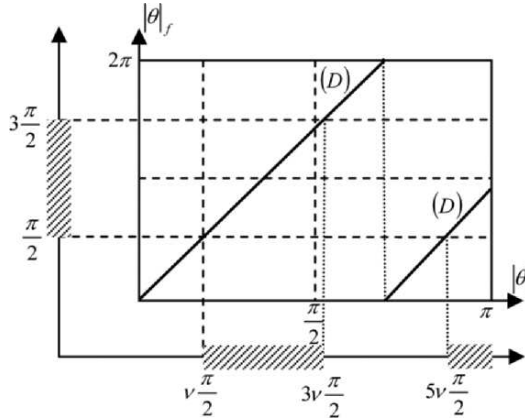


Fig. 4. F_ν as a function of $|\theta|$ and $|\theta_f|$, and deduced stable domain (//).

As ν decays towards, 0, the slope of (D) increases significantly such that high values of $|\theta|$ lead to detection of some instability within the fractional stability domain $\mathcal{D}_S = \left[\nu \frac{\pi}{2}, \pi \right]$ which is thus reduced to:

$$\mathcal{D}_S' = [0, \pi] \cap \bigcup_{i=1,2,\dots} \left[(4i-3)\nu \frac{\pi}{2}, (4i-1)\nu \frac{\pi}{2} \right]. \quad (21)$$

A method leading to a necessary and sufficient condition for stability of fractional systems is therefore necessary.

6 Stability Theorem Based on a Geometric Analysis of the Stability Domain

6.1 Characterization of the entire stability domain

In order to characterize the entire stability domain $\mathcal{D}_S$, it is necessary to define a function that associates every $\lambda \in \mathcal{D}_S$ with λ' belonging to a convex domain of the complex plane. This convex domain may be the left-half complex plane whose characterization is performed through LMI in theorem 4. Such a function can be defined by:

$$F'_\nu : x \rightarrow \frac{1}{2-\nu}x + \frac{1-\nu}{2-\nu}\pi, \quad (22)$$

which is represented by line (D') in Fig. 5.

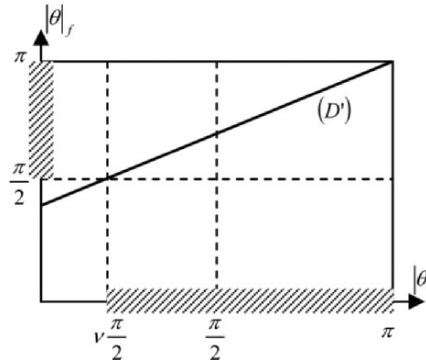


Fig. 5. F'_ν as a function of $|\theta|$ and $|\theta_f|$, and deduced stability domain (//).

6.2 Equivalent integer order system

Using function F'_ν defined by (22), it is now possible to assess stability of a fractional system through stability analysis of an equivalent integer system whose state transition matrix is to be determined.

Let $a = e^{j\alpha}$, where j is the complex variable and $\alpha \in [0, \pi]$. As

$$\alpha = -j \ln(a), \quad (23)$$

one can note that

$$\arg(b) = F'_\nu(\arg(a)) \quad \text{if} \quad b = -(-a)^{\frac{1}{2-\nu}}. \quad (24)$$

Thus, $\arg(a) \in \left] \nu \frac{\pi}{2}, \pi \right]$ iff $\arg\left(-(-a)^{\frac{1}{2-\nu}}\right) \in \left] \frac{\pi}{2}; \pi \right]$.

Stability of system (6) can thus be deduced by applying theorem 4 to a fictive integer system with state transition matrix $-(-A)^{\frac{1}{2-\nu}}$.

Theorem 6: Fractional system (6) is $t^{-\nu}$ stable if and only if a positive definite matrix $P \in \mathbb{S}$ exists such that

$$\left(-(-A)^{\frac{1}{2-\nu}}\right)^T P + P \left(-(-A)^{\frac{1}{2-\nu}}\right) < 0.$$

□

Proof: See steps above.

□

As $-(-A)^{\frac{1}{2-\nu}}$ is a complex matrix, theorem 6 needs to be slightly changed when implemented in a LMI solver. As any complex LMI can be turned into a real one [18], the following LMI is to be implemented:

$$\begin{pmatrix} \text{Re}\left(-(-A)^{\frac{1}{2-\nu}}\right)^T P + P \text{Re}\left(-(-A)^{\frac{1}{2-\nu}}\right) & \text{Im}\left(-(-A)^{\frac{1}{2-\nu}}\right)^T P + P \text{Im}\left(-(-A)^{\frac{1}{2-\nu}}\right) \\ \text{Im}\left(-(-A)^{\frac{1}{2-\nu}}\right)^T P + P \text{Im}\left(-(-A)^{\frac{1}{2-\nu}}\right) & \text{Re}\left(-(-A)^{\frac{1}{2-\nu}}\right)^T P + P \text{Re}\left(-(-A)^{\frac{1}{2-\nu}}\right) \end{pmatrix} < 0. \quad (25)$$

6.3 Validity of the method

Figure 6 presents the stability domain $\mathcal{D}_S''$ determined using theorem 6, according to the values of ν and of $|\theta|$.

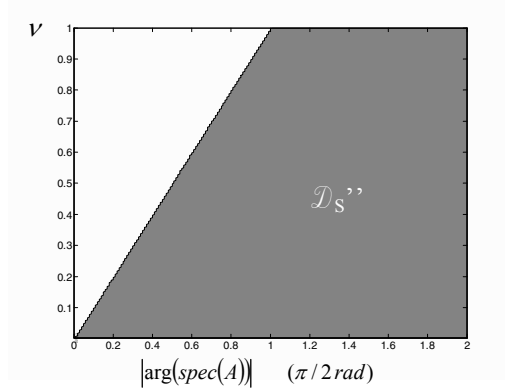


Fig. 6. Stability domain $\mathcal{D}_S''$ (■) determined by theorem 6 according to the values of ν and $|\theta|$.

When compared with Fig. 1, the entire stability domain $\mathcal{D}_s$ is identified here ($\mathcal{D}_s'' = \mathcal{D}_s$). The criterion is therefore not only sufficient but also necessary for stability detection of fractional systems.

However, LMI of theorem 6 is not linear in relation to matrix A , thus limiting its use in more specific control problems.

7 Stability Criterion Based on Unstability Domain Characterization

7.1 Problem definition

This approach is based on the obvious fact that a fractional system is stable if and only if it is not unstable.

Applied to system (6) it emerges that the eigenvalues of the matrix A lie in the stable domain if and only if they do not lie in the unstable one, which is, as previously mentioned, convex.

7.2 Characterization of the entire unstable domain

Let $\mathcal{D}_u$ denote the unstable domain as depicted on Fig. 2a. It is obvious that $\lambda \in \mathbb{C}$ belongs to $\mathcal{D}_u$ if and only if it belongs to both $\mathcal{D}_{u1}$ and $\mathcal{D}_{u2}$ defined by

$$\mathcal{D}_{u1} = \left\{ \lambda \in \mathbb{C} \operatorname{Re} \left(\lambda \exp(j(1-\nu)\frac{\pi}{2}) \right) \geq 0 \right\}, \quad (26)$$

and

$$\mathcal{D}_{u2} = \left\{ \lambda \in \mathbb{C} \operatorname{Re} \left(\lambda \exp(j(\nu-1)\frac{\pi}{2}) \right) \geq 0 \right\}. \quad (27)$$

Thus λ belongs to $\mathcal{D}_u$ if and only if

$$\begin{cases} \operatorname{Re} \left(\lambda \exp(j(1-\nu)\frac{\pi}{2}) \right) \geq 0 \\ \operatorname{Re} \left(\lambda \exp(j(\nu-1)\frac{\pi}{2}) \right) \geq 0 \end{cases}, \quad (28)$$

or if and only if

$$\begin{cases} \lambda \exp(j(1-\nu)\frac{\pi}{2}) + \left(\lambda \exp(j(1-\nu)\frac{\pi}{2}) \right)^* \geq 0 \\ \lambda \exp(j(\nu-1)\frac{\pi}{2}) + \left(\lambda \exp(j(\nu-1)\frac{\pi}{2}) \right)^* \geq 0 \end{cases}, \quad (29)$$

which can be rewritten as:

$$\begin{cases} \lambda r + \lambda^* r^* \geq 0 \\ \lambda r^* + \lambda^* r \geq 0 \end{cases}, \quad (30)$$

where $r = \sin\left(\nu \frac{\pi}{2}\right) + j \cos\left(\nu \frac{\pi}{2}\right)$.

Fractional system (6) is thus $t^{-\gamma}$ stable if and only if

$$\exists \lambda \in \mathcal{D}_u, \quad q \in \mathbb{C}^n: (\lambda I - A)q = 0, \quad q \neq 0, \quad (31)$$

or if and only if

$$\exists \lambda \in \mathbb{C}, \quad q \in \mathbb{C}^n, \quad \begin{cases} \lambda r + \lambda^* r^* \geq 0 \\ \lambda r^* + \lambda^* r \geq 0 \end{cases}: (\lambda I - A)q = 0, \quad q \neq 0. \quad (32)$$

As for some $\lambda \in \text{spec}(A)$, $\lambda^* \in \text{spec}(A)$, and as $\mathcal{D}_{u1}$ and $\mathcal{D}_{u2}$ are symmetric in relation to the real axis of the complex plane, condition (32) becomes

$$\exists \lambda_1 \in \text{spec}(A), \lambda_1 \in \mathcal{D}_{u1} \Leftrightarrow \exists \lambda_2 \in \text{spec}(A), \lambda_2 \in \mathcal{D}_{u2}, \quad (33)$$

and fractional system (6) is $t^{-\gamma}$ stable if and only if

$$\exists \lambda \in \mathbb{C}, q \in \mathbb{C}^n, \quad \lambda r + \lambda^* r^* \geq 0: (\lambda I - A)q = 0, \quad q \neq 0. \quad (34)$$

It is now possible to use the following lemma given in [20].

Lemma 1 [20]: *There exists a vector $p = \lambda q \neq 0$ for some $\lambda + \lambda^* \geq 0$ if and only if $pq^* + qp^* \geq 0$.*

□

Applied to relation (32), fractional system (6) is thus $t^{-\gamma}$ stable if and only if

$$\exists p = \lambda r q \neq 0, \quad q \in \mathbb{C}^n, \quad pq^* + qp^* \geq 0: (\lambda I - A)q = 0, \quad q \neq 0, \quad (35)$$

or if and only if

$$\exists q \neq 0, \quad q \in \mathbb{C}^n, \quad \lambda r q q^* + q q^* \lambda^* r^* \geq 0: (\lambda I - A)q = 0. \quad (36)$$

As $\lambda q = Aq$, fractional system (6) is $t^{-\gamma}$ stable if and only if

$$\exists q \neq 0, \quad q \in \mathbb{C}^n, \quad r A q q^* + q q^* A^T r^* \geq 0. \quad (37)$$

Theorem 7: Fractional system (6) is $t^{-\gamma}$ stable if and only if there does not exist any nonnegative rank one matrix $Q \in \mathbb{C}^{n \times n}$ such that

$$\begin{pmatrix} (AQ + QA^T) \sin\left(\nu \frac{\pi}{2}\right) & (AQ - QA^T) \cos\left(\nu \frac{\pi}{2}\right) \\ (AQ - QA^T) \cos\left(\nu \frac{\pi}{2}\right) & (AQ + QA^T) \sin\left(\nu \frac{\pi}{2}\right) \end{pmatrix} \geq 0.$$

□

Proof: See steps above.

□

8 Conclusion

An analysis of an existing method and two new methods are presented in order to characterize stability of fractional systems through LMI tools. Matignon's theorem developed for stability analysis of fractional systems is first presented. A new proof of its extension to systems whose fractional order ν verifies $1 < \nu < 2$ is proposed. For such derivative orders, stability is granted if all the eigenvalues of its state transition matrix belong to a convex subset of the complex plane, called stability domain. A trivial LMI stability condition is thus presented. For fractional orders ν verifying $0 < \nu < 1$, stability domain is not a convex subset of the complex plane. Three stability conditions involving LMI are however proposed.

The first condition appears in [13] and appears after algebraic transformations of the fractional system state-space representation. The obtained condition is only sufficient. Even if the derived system and the original fractional system have strictly the same behavior, an explanation of the conservatism of the condition is presented.

The second condition is new and relies on a geometric analysis of the stability domain. The resulting LMI stability condition is sufficient and necessary but is not linear in relation to the state transition matrix of the fractional system state-space representation, which can limit its applicability.

In order to overcome this problem, a third condition is proposed. It relies on the fact that instability domain is a convex subset of the complex plane when $0 < \nu < 1$.

This work is a first step in fractional system stability analysis using LMI tools towards new conditions and applications.

Acknowledgment

Thanks go to the American Society of Mechanical Engineers (ASME) for the permission to publish this revised contribution of an ASME article.

References

1. Podlubny I (1999) Fractional-order systems and $PI^{\lambda}D^{\mu}$ -Controllers, *IEEE Trans. Automat. Control*, 44(1):208–214.
2. Monje CA, Vinagre BM, Chen YO, Feliu V, Lanusse P, Sabatier J (2004) Proposals for fractional PID tuning, First IFAC Workshop on Fractional Derivative and its Application, FDA 04, Bordeaux, France, 2004.
3. Caponetto R, Fortuna L, Porto D (2004) A new tuning strategy for a non integer order PID controller, First IFAC Workshop on Fractional Derivative and its Application, FDA 04, Bordeaux, France.
4. Chen YQ, Moore KL, Vinagre BM, Podlubny I (2004) Robust PID controller auto tuning with a phase shaper, First IFAC Workshop on Fractional Derivative and its Application, FDA 04, Bordeaux, France.
5. Oustaloup A, Mathieu B (1999) La commande CRONE du scalaire au multivariable. Hermès, Paris.
6. Battaglia J-L, Cois O, Puissegur L, Oustaloup A (Juillet 2001) Solving an inverse heat conduction problem using a non-integer identified model, *Int. J. Heat Mass Transf.*, 44(14). 2671–2680.
7. Hotzel R, Fliess M (1998) On linear systems with a fractional derivation: Introductory theory and examples, *Math. Comp. Simulation*, special issue: Delay Systems, 45:385–395.
8. Maignon D (July 1996) Stability results on fractional differential equations with applications to control processing, *Comp. Eng. Syst. Appl. multiconference*, 2:963–968, IMACS, IEEE-SMC.
9. Biannic JM (1996) Commande robuste des systèmes à paramètres variables, application en aéronautique, PhD Thesis, ENS de l'Aéronautique et de l'Espace.
10. Balakrishnan V, Kashyap RL (March 1999) Robust stability and performance analysis of uncertain systems using linear matrix inequalities, *J. Optim. Theory Appl.* 100(3):457–478.
11. Balakrishnan V (August 2002) Linear Matrix Inequalities in Robust Control: A Brief Survey, in *Proceedings of the Mathematical Theory of Networks and Systems*. Notre Dame, Indiana.
12. Boyd S, Vandenberghe L (2004) *Convex Optimization*, Cambridge University Press.
13. Momani S, El-Khazali R (November 19–22, 2001) Stability An alysis of Composite Fractional Systems, in *Intelligent Systems and Control*, Tampa, Florida.
14. Samko AG, Kilbas AA, Marichev OI (1987) *Fractional Integrals and Derivatives*. Gordon and Breach Science, Minsk.
15. Miller KS, Ross B (1993) *An Introduction to the Fractional Calculus and Fractional Differential Equation*. Wiley, New York.
16. Malti R, Cois O, Aoun M, Levron F, Oustaloup A (July 21–26 2002) Computing impulse response energy of fractional transfer function, in the 15th IFAC World Congress 2002, Barcelona, Spain.
17. Boyd S, El Ghaoui L, Feron E, Balakrishnan V (June 1994) Linear matrix inequalities in system and control theory. Volume 15 of *Studies in Applied Mathematics*, Philadelphia.
18. Gahinet P, Nemirovski A, Laub AJ, Chilali M (1995) *LMI control toolbox user's guide*, The Math Works.
19. Takab D, Kuo BC (1971) *Optimal Control by Mathematical Programming*. Prentice-Hall, New Jersey.
20. Ben-Tal A, El Ghaoui L, Nemirovski A (2000) Robustness, in *Handbook of Semidefinite Programming: Theory, Algorithms and Applications*. Kluwer Academic, Boston, pp. 68–92.

ACTIVE WAVE CONTROL FOR FLEXIBLE STRUCTURES USING FRACTIONAL CALCULUS

Masaharu Kuroda

National Institute of Advanced Industrial Science and Technology (AIST), 1-2-1 Namiki, Tsukuba, Ibaraki 305-8564, Japan; Tel: +81-29-861-7147, Fax: +81-29-861-7098, E-mail: m-kuroda@aist.go.jp

Abstract

Recently, active wave control theory has attracted great interest as a novel method for vibration control of large space structure (LSS). The method can be applied to suppress vibration in large flexible structures that have high modal density, even for relatively low frequencies. In this report, we formulate a feedback-type active wave control law, described as a transfer function including a Laplace transform with an $s^{1/2}$ or $s^{3/2}$ term. As an example, we present the fractional-order derivatives and integrals of structural responses in the vibration suppression of a thin, light cantilevered beam.

Keywords

Fractional calculus, control, fractional-order transfer function, wave, flexible structure.

1 Introduction

Flexible structures such as large-scale space structures (LSS) have a high vibration-mode density, even in the low-frequency domain. Therefore, to achieve active vibration suppression, vibration control approaches based on modal analysis must determine the limits of the spillover instability phenomenon. Hence it is necessary to establish a new control methodology that can be applied to flexible structures. Among such novel approaches, the active wave (absorption) control method has attracted attention.

It is known that control laws derived from active wave control theory can be expressed using a transfer function including a non-integer order power of the

variable s of the Laplace transform. However, there are difficulties implementing the transfer function due to the non-integer order power. In this report we present a formulation of a feedback-type active wave controller, designed to suppress vibration of a flexible cantilevered beam, described by a transfer function with $\sqrt{s}$ or $s\sqrt{s}$ and introducing a fractional-order derivative and integral.

2 Active Wave Control of a Flexible Structure

Active wave control differs from conventional vibration control in the way it suppresses the vibration modes (standing waves) of a structure. The interaction of progressive and retrogressive waves creates a standing wave, each of which can be treated as a controlled object in the control method developed by von Flotow and Schafer [1].

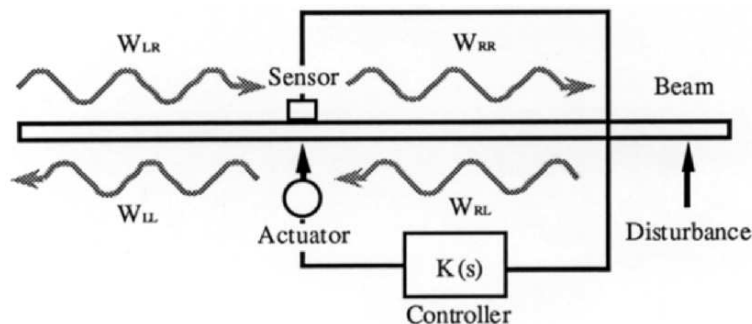


Fig. 1. Schematic diagram of the active wave control method.

As an example, we consider the vibration control of a flexible cantilever (Fig. 1). A sensor and an actuator are placed near the middle of the beam. A disturbance is applied at the free end of the beam. The relationship between the progressive and retrogressive wave vectors generated by the disturbance on the cantilever can be described in matrix form using boundary conditions on the control point. Backward propagating waves are produced by the reflection of the progressive wave, but are also produced by the control input, allowing control of the backward propagating wave. The progressive wave vectors can also be controlled. However, we note that only one control force can control any one of the wave

components. The transfer matrix from the progressive wave vector to the retrogressive wave vector is called the scattering matrix.

In the feedback control of the system, the beam deflection angle is detected and fed to a bending moment actuator, such as a piezoelectric patch, used as control input. Taking a_1 and a_2 as progressive wave components and b_1 and b_2 as retrogressive wave components, we obtain a closed-loop relationship between the progressive wave and the retrogressive wave [2]:

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} 0 \\ -\sqrt{2} \\ p \end{bmatrix} \frac{C}{EI} \mathbf{M}_c, \quad (1)$$

where EI is the bending stiffness of the cantilever, ρ is the mass per unit length, A is the cross-sectional area of the cantilever, $C^2 = EI/\rho A$, and M_c is the bending moment for a control input $p = s/C$. Using Eq. (1), we can devise a controller to minimize the H_∞ -norm of the transfer function of the closed-loop scattering matrix at the actuator point, i.e., the transfer function from the beam deflection angle θ to the bending moment for the control input, M_c . The resultant controller can be expressed as [2, 3]

$$K(s) = \frac{M_c(s)}{\theta(s)} = \frac{\sqrt{2}(EI)^{\frac{3}{4}}(\rho A)^{\frac{1}{4}}}{3} \sqrt{s}. \quad (2)$$

This equation shows that the active wave control law includes the term $\sqrt{s}$: the half-order derivative element. In other words, the control law can be qualified as velocity feedback with a phase shift of 45° , rather than 90° . According to MacMartin and Hall [3], the controller is capable of extracting half of the power input to the structure over the entire frequency range.

Active wave control can be performed using the deflection of the beam as the detected value and the shear force as the control input. In this case, the transfer function of the controller can be expressed as [4]

$$K(s) = \frac{F_c(s)}{w(s)} = 4G_1(EI)^{\frac{1}{2}}(\rho A)^{\frac{3}{4}}s\sqrt{s}. \quad (3)$$

In this equation, the control law includes $s\sqrt{s}$.

We note that the controller must be described by a transfer function in the form of a fractional expression of an integer power series. Customarily, transfer functions including non-integer powers of s have been approximated by introducing a limitation in the frequency range. Following the methods of MacMartin and others, the transfer function is substituted by an approximated

transfer function with a finite number of poles and zeros located in the exponential positions along the negative real axis, as shown in the following equation [2]:

$$\sqrt{s} \approx \frac{(s+10^{-4})(s+10^{-2})(s+10^0)(s+10^2)}{(s+10^{-3})(s+10^{-1})(s+10^1)(s+10^3)}. \quad (4)$$

3 Fractional Calculus

The transfer function can be defined in terms of fractional calculus, whereby the derivatives and integrals of a continuous function can be defined using non-integers [5–8]. The definition of a fractional derivative can be written as

$$D^q[x(t)] = \frac{1}{\Gamma(1-q)} \frac{d}{dt} \left[\int_0^t \frac{x(\tau)}{(t-\tau)^q} d\tau \right], \quad (0 < q < 1). \quad (5)$$

As with a normal integer-order derivative, a fractional derivative satisfies linearity and the composition rule with a zero initial condition [9]:

$$\text{Linearity: } D^\alpha[ax(t) + by(t)] = aD^\alpha[x(t)] + bD^\alpha[y(t)], \quad (6)$$

$$\text{Composition Rule: } D^\alpha[D^\beta[x(t)]] = D^{\alpha+\beta}[x(t)], \quad (x(0) = y(0) = 0), \quad (7)$$

$$\text{Laplace Transform: } L[D^\alpha[x(t)]] = s^\alpha L[x(t)], \quad (x(0) = y(0) = 0). \quad (8)$$

Fractional calculus has been applied to control theory and research into this field has flourished. Examples of recent publications include a report on PID-controller parameter tuning based on fractional calculus [10], a history of the development of control based on fractional-order derivatives and integrals [11] and reports on discrete-time modelling and numerical simulation of fractional systems by transfer function representation and state-space representation [12, 13].

Some implementation techniques in structural dynamics yield fractional-order derivative/integral responses. Examples include a special analogue electrical circuit device [9] and a digital filter designed from a discrete approximated form of the definition equation [14]. However, obtaining fractional-order derivatives/integrals of dynamical response at a single point on a structure presents technological difficulties.

In this study we overcome the difficulties due to fractional derivative responses by constructing the responses at the actuation point from a linear combination of multiple signals at several sensing points, rather than from a signal from a single sensor. In this method, special sensors with additional signal-conversion functions are not required and existing displacement and velocity sensors can be used.

3.1 Calculated response of 1/2-order and 3/2-order derivatives

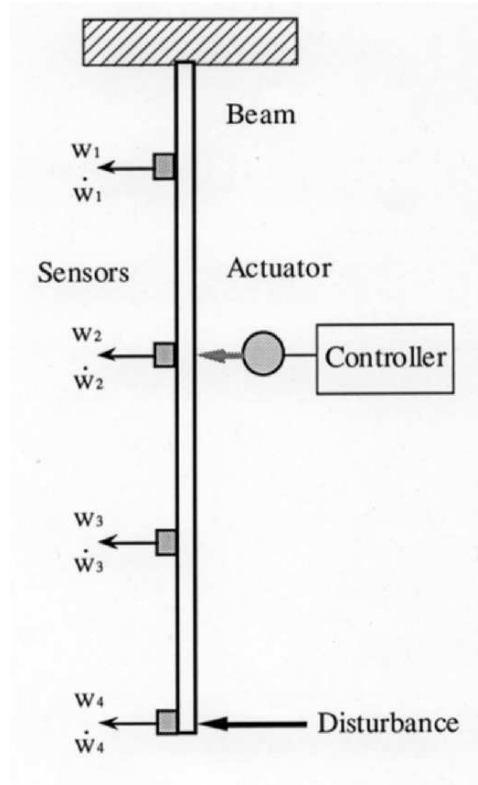


Fig. 2. Active wave control for a flexible cantilevered beam.

Figure 2 shows a flexible cantilevered beam of length L with several sensors and an actuator. Without losing generality, four points on the beam can be designated as sensor points, such as $L/10$, $4L/10$, $7L/10$, and L . In this example, the actuator

is located at the $4L/10$ position of the cantilever. Displacement (the zeroth-order time-derivative of $w_i(t)$, see below) and velocity (the first-order time-derivative of $w_i(t)$) can be detected at each sensor. A linear combination of the displacement and velocity signals at each sensor point is fed back to the actuator. The actuator then supplies shear force or bending moment to the cantilever.

The system response can be characterized by the displacement vector defined as

$$[w(t)] = [w_1(t), w_2(t), w_3(t), w_4(t)]^T. \quad (9)$$

The equation of motion (EOM) for the system under free vibration can be described as

$$[M][\ddot{w}(t)] + [C][\dot{w}(t)] + [K][w(t)] = 0. \quad (10)$$

The expanded system response can then be defined using fractional-order derivatives of $w(t)$:

$$[\hat{w}(t)] = \left[D^{\frac{3}{2}}[w(t)], D^{\frac{2}{2}}[w(t)], D^{\frac{1}{2}}[w(t)], D^0[w(t)] \right]^T. \quad (11)$$

Consequently, the fractional-order EOM can be developed as

$$\begin{aligned} & D^{\frac{1}{2}} [M] \left[D^{\frac{3}{2}} + D^{\frac{2}{2}} + D^{\frac{1}{2}} + D^0 \right] [w(t)] \\ & - [M] \left[D^{\frac{3}{2}} + D^{\frac{2}{2}} + D^{\frac{1}{2}} \right] [w(t)] \\ & + D^{\frac{1}{2}} [C] \left[D^{\frac{1}{2}} + D^0 \right] [w(t)] \\ & - [C] \left[D^{\frac{1}{2}} \right] [w(t)] \\ & + [K] \left[D^0 \right] [w(t)] = 0. \end{aligned} \quad (12)$$

The eigenstructure of the fractional-order EOM of Eq. (12) cannot be solved directly using traditional methods for eigenvalue problems because it includes fractional-order derivatives. To overcome the difficulties arising from the fractional-order derivatives, the expanded EOM can be expressed as

$$D^{\frac{1}{2}} [\hat{M}] [\hat{w}(t)] + [\hat{K}] [\hat{w}(t)] = 0, \quad (13)$$

$$[\hat{M}] = \begin{bmatrix} 0 & 0 & 0 & [M] \\ 0 & 0 & [M] & 0 \\ 0 & [M] & 0 & [C] \\ [M] & 0 & [C] & 0 \end{bmatrix}, \quad (14)$$

$$[\hat{K}] = \begin{bmatrix} 0 & 0 & -[M] & 0 \\ 0 & -[M] & 0 & 0 \\ -[M] & 0 & -[C] & 0 \\ 0 & 0 & 0 & [K] \end{bmatrix}. \quad (15)$$

The matrices $[\hat{M}]$ and $[\hat{K}]$ are the pseudo-mass and pseudo-stiffness matrices, respectively. The formulation of Eq. (13) can be solved using conventional methods for eigenvalue problems.

3.2 Fractional-order derivative feedback

From the above discussion, we can formulate a fractional-order derivative feedback merely by substituting $[G_{3/2}]D^{3/2}[w(t)] + [G_{1/2}]D^{1/2}[w(t)]$ into the right-hand side of Eq. (10), where $[G_{1/2}]$ and $[G_{3/2}]$ are the feedback-gain matrices for the 1/2-order and for the 3/2-order derivatives of the state vector.

3.3 Measured response of 1/2-order and 3/2-order derivatives

The modal expansion of the expanded system response can be written in the following form:

$$[\hat{w}(t)] = [\Phi][\xi]. \quad (16)$$

Here the vector $[\xi]$ is a column vector of the modal coordinates of the system. The modal matrix $[\Phi]$ is composed of conjugate pairs of eigenvectors associated with the conjugate eigenvalues on the principal sheet of the Riemann surface, as shown in the following equation:

$$[\Phi] = [\phi_1, \bar{\phi}_1, \phi_2, \bar{\phi}_2, \phi_3, \bar{\phi}_3, \phi_4, \bar{\phi}_4]. \quad (17)$$

Using Eq. (16), the traditional state vector and the fractional-order state vector for the system can be constructed. The traditional (integer-order) state vector can be described as

$$[x(t)] = [\mathbf{w}_1 \quad \mathbf{w}_2 \quad \mathbf{w}_3 \quad \mathbf{w}_4 \quad \dot{\mathbf{w}}_1 \quad \dot{\mathbf{w}}_2 \quad \dot{\mathbf{w}}_3 \quad \dot{\mathbf{w}}_4]^t. \quad (18)$$

The fractional-order state vector can be given by

$$[y(t)] = \left[D^{\frac{3}{2}}(\mathbf{w}_2, \mathbf{w}_3) \quad D^{\frac{2}{2}}(\mathbf{w}_2, \mathbf{w}_3) \quad D^{\frac{1}{2}}(\mathbf{w}_2, \mathbf{w}_3) \quad D^0(\mathbf{w}_2, \mathbf{w}_3) \right]^t. \quad (19)$$

We can extract the row vectors corresponding to the integer-order state vectors in the matrix $[\Phi]$ to create the smaller matrix $[\phi_A]$, as illustrated in Eq. (20). We can then extract the row vectors corresponding to the fractional-order state vectors in $[\Phi]$ to create a further matrix $[\phi_B]$.

$$[\Phi] = \begin{bmatrix} \phi_{1,1} & \bar{\phi}_{1,1} & \phi_{2,1} & \bar{\phi}_{2,1} & \phi_{3,1} & \bar{\phi}_{3,1} & \phi_{4,1} & \bar{\phi}_{4,1} \\ \phi_{2,2} & \bar{\phi}_{2,2} & \phi_{2,2} & \bar{\phi}_{2,2} & \phi_{3,2} & \bar{\phi}_{3,2} & \phi_{4,2} & \bar{\phi}_{4,2} \\ \phi_{1,3} & \bar{\phi}_{1,3} & \phi_{2,3} & \bar{\phi}_{2,3} & \phi_{3,3} & \bar{\phi}_{3,3} & \phi_{4,3} & \bar{\phi}_{4,3} \\ \phi_{1,4} & \bar{\phi}_{1,4} & \phi_{2,4} & \bar{\phi}_{2,4} & \phi_{3,4} & \bar{\phi}_{3,4} & \phi_{4,4} & \bar{\phi}_{4,4} \\ \phi_{1,5} & \bar{\phi}_{1,5} & \phi_{2,5} & \bar{\phi}_{2,5} & \phi_{3,5} & \bar{\phi}_{3,5} & \phi_{4,5} & \bar{\phi}_{4,5} \\ \phi_{1,6} & \bar{\phi}_{1,6} & \phi_{2,6} & \bar{\phi}_{2,6} & \phi_{3,6} & \bar{\phi}_{3,6} & \phi_{4,6} & \bar{\phi}_{4,6} \\ \phi_{1,7} & \bar{\phi}_{1,7} & \phi_{2,7} & \bar{\phi}_{2,7} & \phi_{3,7} & \bar{\phi}_{3,7} & \phi_{4,7} & \bar{\phi}_{4,7} \\ \phi_{1,8} & \bar{\phi}_{1,8} & \phi_{2,8} & \bar{\phi}_{2,8} & \phi_{3,8} & \bar{\phi}_{3,8} & \phi_{4,8} & \bar{\phi}_{4,8} \\ \phi_{1,9} & \bar{\phi}_{1,9} & \phi_{2,9} & \bar{\phi}_{2,9} & \phi_{3,9} & \bar{\phi}_{3,9} & \phi_{4,9} & \bar{\phi}_{4,9} \\ \phi_{1,10} & \bar{\phi}_{1,10} & \phi_{2,10} & \bar{\phi}_{2,10} & \phi_{3,10} & \bar{\phi}_{3,10} & \phi_{4,10} & \bar{\phi}_{4,10} \\ \phi_{1,11} & \bar{\phi}_{1,11} & \phi_{2,11} & \bar{\phi}_{2,11} & \phi_{3,11} & \bar{\phi}_{3,11} & \phi_{4,11} & \bar{\phi}_{4,11} \\ \phi_{1,12} & \bar{\phi}_{1,12} & \phi_{2,12} & \bar{\phi}_{2,12} & \phi_{3,12} & \bar{\phi}_{3,12} & \phi_{4,12} & \bar{\phi}_{4,12} \\ \phi_{1,13} & \bar{\phi}_{1,13} & \phi_{2,13} & \bar{\phi}_{2,13} & \phi_{3,13} & \bar{\phi}_{3,13} & \phi_{4,13} & \bar{\phi}_{4,13} \\ \phi_{1,14} & \bar{\phi}_{1,14} & \phi_{2,14} & \bar{\phi}_{2,14} & \phi_{3,14} & \bar{\phi}_{3,14} & \phi_{4,14} & \bar{\phi}_{4,14} \\ \phi_{1,15} & \bar{\phi}_{1,15} & \phi_{2,15} & \bar{\phi}_{2,15} & \phi_{3,15} & \bar{\phi}_{3,15} & \phi_{4,15} & \bar{\phi}_{4,15} \\ \phi_{1,16} & \bar{\phi}_{1,16} & \phi_{2,16} & \bar{\phi}_{2,16} & \phi_{3,16} & \bar{\phi}_{3,16} & \phi_{4,16} & \bar{\phi}_{4,16} \end{bmatrix} \quad (20)$$

Consequently, the relationship between the integer-order state vector $[x(t)]$ and the fractional-order state vector $[y(t)]$ can be given by

$$\begin{cases} [x(t)] = [\phi_A][\xi] \\ [y(t)] = [\phi_B][\xi], \end{cases} \quad (21)$$

where $[\phi_A]$ and $[\phi_B]$ are the matrices consisting of row vectors of $[\Phi]$ in Eq. (17) and which are associated with the state vectors defined respectively by Eq. (18) and Eq. (19). Hence, Eq. (21) yields the final equation [15]

$$[y(t)] = [\phi_B][\phi_A]^{-1}[x(t)] = [\phi][x(t)], \quad (22)$$

where $[\phi]$ is the state transformation matrix between the traditional state and the fractional state.

Using these formulae, we can determine fractional-order temporal-derivative terms of the observed value using a linear combination of displacement and velocity signals at each sensor point, as is necessary to implement active wave control.

4 The Active Wave Controller and Its Control Effects

In contrast to traditional vibration suppression methods in which the objective is to control the vibration modes (standing waves) of a structure, the objective of active wave control is to control travelling waves in the structure. Consequently, active wave control is equivalent to control of the power flow in the structure. Ideally, an active wave controller can extract half of the power flow transmitted in the structure at all frequencies [3].

Fractional derivatives enable the formulation of the wave control law directly, rather than using the customary method of approximating the wave control transfer function by a function composed of integer-order power-series of the variable s of the Laplace transform. Furthermore, it provides a deeper understanding of the physical meaning of wave control.

As an example, we consider the wave control of a steel cantilever of length 2.7 m, width 50 mm and thickness 5.8 mm. Sensors and an actuator are placed on the beam as depicted in Fig. 2 and a disturbance is applied at the free end of the beam.

Figure 3 shows the eigenvalues for the expanded system satisfying the equation

$$\hat{\lambda}_j [\hat{M}] [\hat{\phi}_j] + [\hat{K}] [\hat{\phi}_j] = 0. \quad (23)$$

The eigenvalues are in complex conjugate pairs and there exist 4×2 pairs of eigenvalues for the expanded system. The eigenvalues may be mapped onto the Riemann surface for the function $\hat{\lambda}_j = s^{1/2}$, consisting of two Riemann sheets. Four pairs of complex conjugate eigenvalues appear on each sheet.

The eigenvalues on the principal Riemann sheet and the corresponding eigenvectors illustrate the sinusoidal motion of the structure; they form the mode shapes of the structure. The natural frequencies of the original system can be obtained by squaring the eigen pairs. The process gives four conjugate pairs. The imaginary part of each conjugate pair gives the eigenfrequency of the original system while the real part gives the product of the damping ratio and the natural frequencies of the original system.

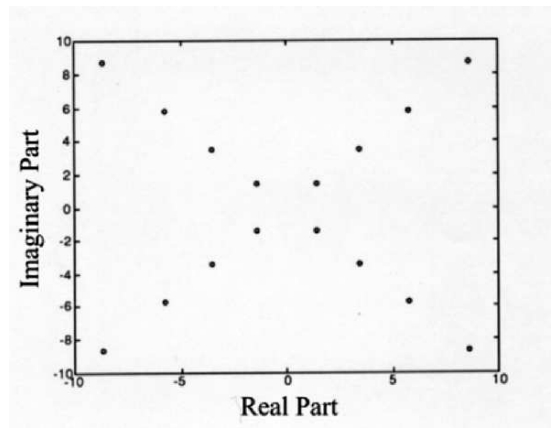


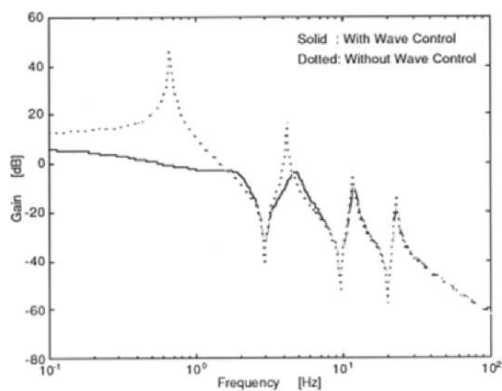
Fig. 3. Eigenvalues of the expanded system.

The eigenvalues on the second Riemann sheet represent poles in the system transfer function, which produce a monotonically decreasing response of the structure. This monotonically decreasing motion describes the creep and relaxation response of the original system.

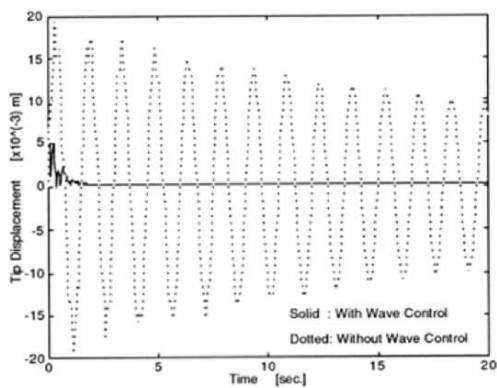
4.1 Combination of beam-slope sensor and bending-moment actuator

An advantage of active wave control is that it yields a controller design that depends directly on the dimensions and material properties of the structure without the necessity of carrying out modal analysis of the structure in advance. Additionally, the controller provides active damping for all structural vibration modes. However, it cannot actively provide strong damping to a specific vibration mode.

We carried out a simulation using a combination of deflection-angle sensors and a bending-moment actuator, as expressed in Eq. (2). Figures 4a and 4b show the driving-point compliance and impulse response evaluate vibration suppression effects of wave control for this simulation.



(a) Driving point compliance

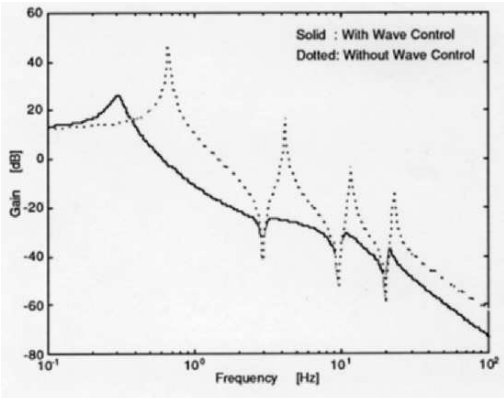


(b) Impulse response

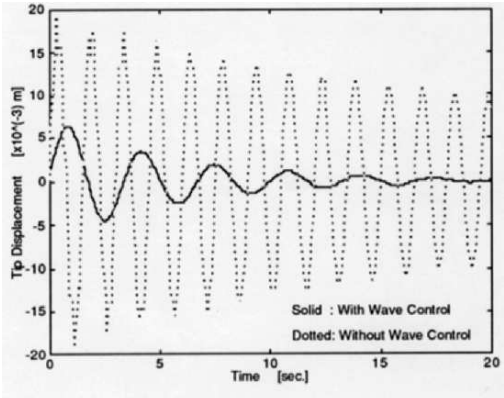
Fig. 4. Active wave control of a flexible cantilevered beam with beam-slope sensors and a bending-moment actuator.

4.2 Combination of beam-deflection sensor and shear-force actuator

We carried out a second simulation using a combination of deflection sensors and a shear-force actuator, as expressed in Eq. (3). Figures 5a and 5b show the wave control for this simulation.



(a) Driving point compliance



(b) Impulse response

Fig. 5. Active wave control of a flexible cantilevered beam with beam-deflection sensors and a shear-force actuator.

The retrogressive wave is eliminated at the control point in both the above simulations, but the results are quite different.

Interestingly, it has been reported that viscoelastic materials, such as silicon gel, also have a frequency response characterized by $\sqrt{s}$ [16]. Accordingly, using a reaction surface, a passive control system that supports a cantilever by viscoelasticity may achieve a similar control effect as active wave control for a sensor/actuator colocation.

5 Conclusions

Active wave control including a $1/2$ -order or a $3/2$ -order derivative element can be formulated using fractional calculus. In this paper we have reported the following:

1. Application of fractional calculus to vibration control
2. The basis for (a) calculating and (b) measuring the responses of $1/2$ -order and $3/2$ -order fractional derivatives of a cantilevered beam
3. Implementation of an active wave controller by means of the response of $1/2$ -order or $3/2$ -order fractional derivatives of a cantilevered beam

We obtained good control results in simulations. In the future, we plan to verify the simulation results through experiments with a flexible cantilevered beam.

Future works will investigate the following:

1. Generalization of vibration control by fractional derivatives and integrals
2. Realization of higher order fractional derivatives and integrals

References

1. Von Flotow AH, Schafer B (1986) Wave-absorbing controllers for a flexible beam, *J. Guid. Control Dynam.*, 9(6):673–680.
2. Agrawal BN (1996) Spacecraft vibration suppression using smart structures, *Proceedings of the 4th International Congress on Sound and Vibration*, pp. 563–570.
3. MacMartin DG, Hall SR (1991) Control of uncertain structures using an H_∞ power flow approach, *J. Guid., Control, Dynam.*, 14(3):521–530.
4. Tanaka N, Kikushima Y, Kuroda M (1992) Active wave control of a flexible beam (on the optimal feedback control) (in Japanese), *Trans. JSME, Series C*, 58(546):360–367.
5. Yang DL (1991) Fractional state feedback control of undamped and viscoelastically-damped structures, *Thesis*, AD-A-220-477, Air Force Institute of Technology, pp. 1–98.
6. Podlubny I (1999) *Fractional Differential Equations*. Academic Press, San Diego.
7. Hilfer R (ed.) (2000) *Applications of Fractional Calculus in Physics*. World Scientific, Singapore.
8. West BJ, Bologna M, Grigolini P (2003) *Physics of Fractal Operators*. Springer, New York.
9. Motoishi K, Koga T (1982) Simulation of a noise source with $1/f$ spectrum by means of an RC circuit (in Japanese), *IEICE Trans.*, J65-A(3):237–244.

10. Barbosa RS, Machado JA, Ferreira IM (2003) A fractional calculus perspective of PID tuning, Proceedings of DETC'03 (ASME), DETC2003/VIB-48375, pp. 651–659.
11. Manabe S (2003) Early development of fractional order control, Proceedings of DETC'03 (ASME), DETC2003/VIB-48370, pp. 609–616.
12. Aoun M, Malti R, Levron F, Oustaloup A (2003) Numerical simulations of fractional systems, Proceedings of DETC'03 (ASME), DETC2003/VIB-48389, pp. 745–752.
13. Poinot T, Trigeassou J-C (2003) Modelling and simulation of fractional systems using a non integer integrator, Proceedings of DETC'03 (ASME), DETC2003/VIB-48390, pp. 753–760.
14. Chen Y, Vinagre BM, Podlubny I (2003) A new discretization method for fractional order differentiators via continued fraction expansion, Proceedings of DETC'03 (ASME), DETC2003/VIB-48391, pp. 761–769.
15. Kuroda M, Kikushima Y, Tanaka N (1996) Active wave control of a flexible structure formulated using fractional calculus (in Japanese), Proceedings of the 74th Annual Meeting of JSME (I), pp. 331–332.
16. Shimizu N, Iijima M (1996) Fractional differential model in engineering problems (in Japanese), Iwaki Meisei University Research Report, No. 9, pp. 48–58.

Part 7

Control

FRACTIONAL-ORDER CONTROL OF A FLEXIBLE MANIPULATOR

Vicente Feliu¹, Blas M. Vinagre², and Concepción A. Monje²

¹ Escuela Técnica Superior de Ingenieros Industriales, Universidad de Castilla-La Mancha, Campus Universitario, 13071 Ciudad Real, Spain; E-mail: Vicente.Feliu@uclm.es

² Escuela de Ingenierías Industriales, Universidad de Extremadura, Avda. de Elvas, s/n, 06071 Badajoz, Spain; E-mail: {bvinagre,cmonje}@unex.es

Abstract

A new method to control single-link lightweight flexible manipulators in the presence of changes in the load is proposed in this paper. The overall control scheme consists of three nested control loops. Once the friction and other nonlinear effects have been compensated, the inner loop is designed to give a fast motor response. The middle loop decouples the dynamics of the system, and reduces its transfer function to a double integrator. A fractional-derivative controller is used to shape the outer loop into the form of a fractional-order integrator. The result is a constant-phase system with, in the time domain, step responses exhibiting constant overshoot, independently of variations in the load (tip mass). In simulation, comparison of the responses to a step command of the manipulator controlled with the controller implemented by different approximations and with the ideal fractional controller showed that the latter could be accurately approximated by standard continuous and discrete controllers of high order preserving the robustness. Simulations also include comparison with standard PD controller, and verification of the assumption of dominant low-frequency vibration mode.

Keywords

Fractional-order controller, robot manipulator, fractional integrator, robust control.

1 Introduction

The control of single-link lightweight manipulators robust to payload changes is a subject of major research interest. Several adaptive and nonadaptive control schemes have been proposed to handle the problem (see, for instance [1–6]).

Since fractional-order controllers have been used successfully in robust control problems, the present work considers a control scheme based on a controller of this type that compensates for undesired changes in the dynamics of the system caused by changes in the payload. The particular design is for the special case of flexible arms that are light in weight compared with the load that they handle. The mechanical structure in this case has a dominant low-frequency vibration mode, and negligible higher frequency modes. It is assumed that any problems caused by the nonlinear Coulomb component of the friction or by changes in the dynamic friction coefficient can be resolved by using the control scheme described in [5]. The general control scheme proposed in this paper consists of three nested loops (Fig. 1):

1. An inner loop that controls the position of the motor. This loop uses a classical PD controller to give a closed-loop transfer function close to unity.
2. A decoupling loop using positive unity-gain feedback. The purpose of this loop is to reduce the dynamics of the system to that of a double integrator.
3. An outer loop that uses a fractional-derivative controller to shape the loop and to give an overshoot independent of payload changes.

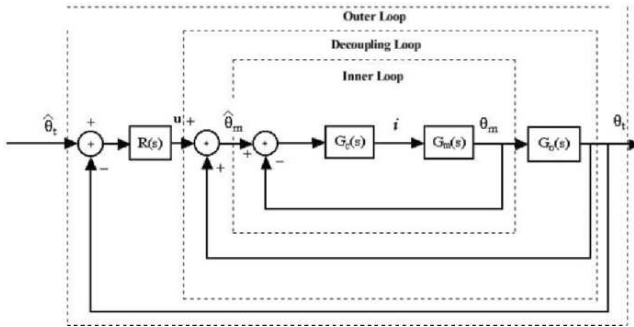


Fig. 1. Proposed general control scheme.

In the figure, $\theta_m(t)$ is the motor angle, $\theta_t(t)$ the tip-position angle, $i(t)$ the motor current, $G_m(s)$ and $G_b(s)$ the transfer functions of motor and beam, respectively, and $G_c(s)$, $R(s)$ the controllers. The design of the first two loops follows [5]. The fractional-order control (FOC) strategy of the outer loop, which is based on the operators of *fractional calculus*, is proposed in this paper.

Fractional calculus generalizes the standard differential and integral operators by defining a single general fundamental operator (see [7–9]). There are two commonly used definitions for the generalized fractional integro-differential: Grünwald–Letnikov (GL) and Riemann–Liouville (RL) [8–9]. The GL definition is

$${}_b D_t^\alpha f(t) = \lim_{h \rightarrow 0} \frac{1}{h^\alpha} \sum_{j=0}^{[(t-b)/h]} (-1)^j \binom{\alpha}{j} f(t - jh), \quad (1)$$

where $[\cdot]$ is a flooring-operator, and $\binom{\alpha}{j}$ is the binomial coefficient. The RL definition is

$${}_b D_t^\alpha f(t) = \frac{1}{\Gamma(n-\alpha)} \frac{d^n}{dt^n} \int_b^t \frac{f(\tau)}{(t-\tau)^{\alpha-n+1}} d\tau, \quad (2)$$

for $(n-1 < \alpha < n)$, where $\Gamma(x)$ is Euler's gamma function. One observes that the notion of the fractional-order operator ${}_b D_t^\alpha$ naturally unifies differentiation and integration. Therefore, terms such as *fractional-order differentiator* or *fractional derivative* should be understood to imply both differentiator and integrator.

If implemented properly, fractional-order controllers will find their place in contributing to many real-world control systems. It has to be borne in mind that a fractional-order controller is an infinite-dimensional linear filter, and that all existing implementation schemes are based on finite-dimensional approximations. These approximate implementations of FOC can be classified into either analogical or digital methods (see [10]). The latter can be further classified into indirect and direct discretization methods. In practice, the FOC implementation should be band-limited with the finite-dimensional approximation being done over an appropriate range of frequencies of practical interest. Furthermore, all the approximations must give stable minimum-phase systems.

The rest of the paper is organized as follows. First, the physical model of the system is presented, followed by a brief description of the control loop for the motor position and the decoupling loop. Details are then given of the fractional-controller-based tip-position control scheme, and results are presented of simulations with particular emphasis on FOC implementations. Finally, some relevant conclusions are drawn.

2 System Model

The model of the electromechanical system to be controlled, described in detail in [5], is depicted in Fig. 2.

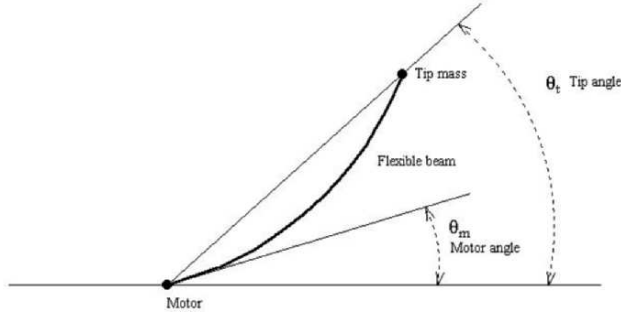


Fig. 2. Electromechanical system to be controlled.

It consists of a DC motor, a slender link attached to the motor hub, and a mass at the end of the link floating on an air table that allows motion of the link in the horizontal plane. The set of differential equations relating the angle of the motor $\theta_m(t)$, the angle of the tip $\theta_t(t)$, and the applied current $i(t)$, is [1]

$$K_m i(t) = J \frac{d^2 \theta_m(t)}{dt^2} + V \frac{d \theta_m(t)}{dt} + C_t(t) + CF(\text{sign}(\frac{d \theta_m(t)}{dt})), \quad (3)$$

$$C_t(t) = C(\theta_m(t) - \theta_t(t)) = mL^2 \frac{d^2 \theta_t(t)}{dt^2}, \quad (4)$$

where K_m is the electromechanical constant of the motor, J the polar moment of inertia of the motor and hub, V the dynamic friction coefficient, $C_t(t)$ the coupling torque between motor and link (the bending moment at the base of the link), m the tip mass, CF the Coulomb friction, $C = (3EI)/L$ a constant that depends on the stiffness EI and the length L of the arm, and t is time. Equation (4) holds approximately because the beam is nearly massless. The magnitude of the Coulomb friction CF can be determined from the spectral analysis of the motor position and the current signals [1]. The coupling torque $C_t(t)$ can be calculated either from strain gauge measurements at the link's base or by the difference between angle measurements of the motor and tip.

3 Motor-Position Control and Decoupling Loops

3.1 Motor-position control loop

The motor-position control loop is the inner loop of Fig. 1. The controller design for this loop has to satisfy two objectives. One is that the modeling errors and nonlinearities introduced by Coulomb friction and changes in the dynamic friction coefficient have to be removed, and the other is that the response of the motor position has to be made much faster than the response of the beam transfer function $G_b(s)$. With the fulfilment of the second objective, the inner loop can be replaced by an equivalent block whose transfer function is approximately equal to unity, that is, the error in the motor position is small and is quickly removed.

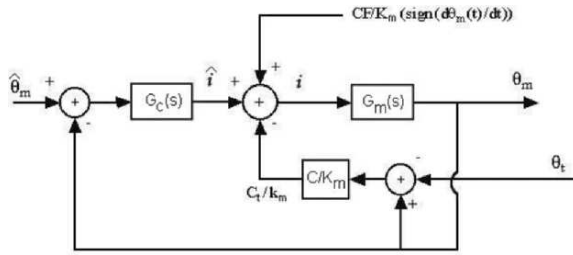


Fig. 3. Motor-position control loop (inner loop).

To simplify the design of the inner loop, the system can be linearized by compensating for the Coulomb friction, and decoupled from the dynamics of the beam by compensating for the coupling torque. This is done by adding the current equivalent to these torques to the control current (Fig. 3). This added current is

$$i_c(t) = \frac{1}{K_m} (C_t(t) + CF(\text{sign}(\frac{d\theta_m}{dt}))). \quad (5)$$

With this compensation, the transfer function between the angle of the motor and the current, $\hat{i}$, is

$$\frac{\theta_m(s)}{\hat{i}(s)} = \hat{G}_m(s) = \frac{K_m/J}{s(s + V/J)}, \quad (6)$$

which corresponds to the free movement of the motor, $C_t(t) = 0$, $CF = 0$, in Eq. (3).

The controller $G_c(s)$ is designed so that the response of the inner loop (position control of the motor) is significantly faster than the response of the outer loop (position control of the tip) without overshoot. When the closed-loop gain of the inner loop is sufficiently high, the motor position will track the reference position with negligible error, and, in the present case, without saturating the actuator (without overpassing the motor current limit). The dynamics of the inner loop ($\theta_m(s)/\hat{\theta}_m(s)$) may then be approximated by “1” when designing the outer-loop controller.

3.2 Decoupling loop

Dynamics of the arm: From Eq. (4) it is obtained that

$$G_b(s) = \frac{\theta_t(s)}{\theta_m(s)} = \frac{\omega_0^2}{s^2 + \omega_0^2}, \quad (7)$$

where the natural resonant frequency of the beam with the motor clamped is ω_0 rad/s, which is related to the tip mass, m , and the stiffness, EI , and the length, L , of the beam by the expression: $\omega_0^2 = 3EI/mL^3$. Assuming a disturbance in the form of an initial conditions polynomial, $P(s) = as + b$ representing the tip's initial angular position and speed, one finds the tip position to be

$$\theta_t(s) = \frac{\omega_0^2}{s^2 + \omega_0^2} \theta_m(s) + \frac{1}{s^2 + \omega_0^2} P(s). \quad (8)$$

If the inner loop has been satisfactorily closed then $\theta_m(t) \sim \hat{\theta}_m(t)$, and the reference input to the motor could be used as input to system of Eq. (8).

Decoupling strategy: The purpose of this loop is to simplify the dynamics of the arm. For the case of a beam with only one vibrational mode, a very simple decoupling loop can be implemented that reduces the dynamics of the system to a double integrator. In particular, one simply closes a positive unity-gain feedback loop around the tip position (Fig. 4). Equation (8) then becomes

$$\theta_t(s) = \frac{\omega_0^2}{s^2} u(s) + \frac{1}{s^2} P(s). \quad (9)$$

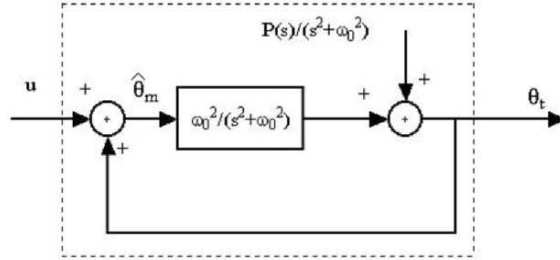


Fig. 4. Decoupling loop (middle loop).

4 Tip-Position Fractional Controller

4.1 Fractional derivative controller

With the inner loop and the decoupling loop closed, the block diagram of Fig. 1 is equivalent to the reduced diagram of Fig. 5, which is based on Eq. (9). From this diagram one obtains for the tip position the expression

$$\theta_t(s) = \frac{1}{1 + \frac{s^2}{R(s)\omega_0^2}} \hat{\theta}_t(s) + \frac{1}{1 + \frac{s^2}{R(s)\omega_0^2}} \frac{P(s)}{R(s)\omega_0^2}. \quad (10)$$

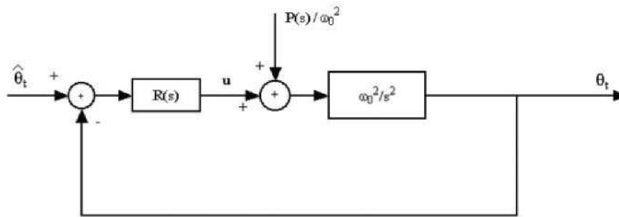


Fig. 5. Reduced diagram for the outer loop.

The controller $R(s)$ has a twofold purpose. One objective is to obtain a constant phase margin in the frequency response, in other words, a constant

overshoot in the step time response, for varying payloads. The other is to remove the effects of the disturbance, represented by the initial-conditions polynomial, on the steady state. To attain these objectives, most authors propose the use of some form of adaptive control scheme (see [5]). Using a fractional-derivative controller, however, both objectives can be achieved without the need for any kind of adaptive algorithm, as will now be shown.

Condition for constant phase margin: From Eq. (10), in the particular case considered in this paper, the condition for a constant phase margin can be expressed as

$$\arg \left[R(j\omega) \frac{\omega_0^2}{(j\omega)^2} \right] = \text{constant}, \quad \forall \omega, \quad (11)$$

and the resulting phase margin ϕ_m is

$$\phi_m = \arg[R(j\omega)]. \quad (12)$$

For a constant phase margin $0 < \phi_m < \pi/2$ the controller must be of the form

$$R(s) = Ks^\beta, \quad \beta = \frac{2}{\pi} \phi_m, \quad (13)$$

so that $0 < \beta < 1$. This $R(s)$ is a fractional derivative controller of order β , in other words, it is a system that performs the fractional derivative of order β defined in Eq. (2).

Condition for removing the effects of disturbances: From the final value theorem, the condition for the effects of the disturbance to be removed is

$$\lim_{s \rightarrow 0} s \frac{1}{1 + \frac{s^2}{R(s)\omega_0^2}} \frac{P(s)}{R(s)\omega_0^2} = 0. \quad (14)$$

Substituting $R(s) = Ks^\beta$ and $P(s) = as + b$, this condition becomes

$$\lim_{s \rightarrow 0} \frac{1}{1 + \frac{s^{2-\beta}}{K\omega_0^2}} \frac{b}{K\omega_0^2} s^{1-\beta} = 0, \quad (15)$$

which implies that $\beta < 1$.

4.2 Ideal response to a step command

Assuming that the dynamics of the inner loop can be approximated by unity and that disturbances are absent, one has for the closed-loop transfer function with the controller of Eq. (13)

$$H(s) = \frac{\theta_t(s)}{\hat{\theta}_t(s)} = \frac{1}{1 + \frac{s^2}{R(s)\omega_0^2}} = \frac{K\omega_0^2}{s^{2-\beta} + K\omega_0^2}, \quad (16)$$

which corresponds in form to Bode's ideal-loop transfer function [7]. The corresponding step response is

$$\theta_t(t) = \mathcal{L}^{-1} \left\{ \frac{K\omega_0^2}{s(s^{2-\beta} + K\omega_0^2)} \right\} = K\omega_0^2 t^{2-\beta} E_{2-\beta, 3-\beta}(-K\omega_0^2 t^{2-\beta}), \quad (17)$$

where $E_{\alpha, \alpha+1}(-At^\alpha)$ is the two-parameter Mittag-Leffler function (see [7]), the overshoot is fixed by $2-\beta$, which is independent of the payload, and the speed is fixed by $K\omega_0^2$, that is, by the payload and the controller gain. To obtain a required step response, it is then necessary to select the values of two parameters. The first is the order β to adjust the overshoot between 0 ($\beta=1$) and 1 ($\beta=0$), or, equivalently, a phase margin between 90° and 0° . The second is the gain K to adjust the crossover frequency, or, equivalently, the speed of the response for a nominal payload. It is interesting to note that increasing β decreases α ($\alpha=2-\beta$) and the overshoot, but increases the time required to correct the disturbance effects.

5 Application Case and Simulation Results

5.1 Mechanical system

The link is a piece of music wire 0.18 m long clamped at the motor hub. The tip mass is a fiberglass disk attached at its center to the end of the link with a freely pivoting pin-joint. The disk has a nominal mass of 54 g, and floats on the horizontal air table with minimal friction. Since the mass of the link is small relative to that of the disk, and the pinned joint prevents generation of torque at

the end of the link, the mechanical system behaves practically as an ideal, single-degree-of-freedom, undamped spring-mass system. The parameters of the motor are $K_m = 0.2468 \text{ N} \cdot \text{m/A}$ and $J = 6.2477 \cdot 10^{-4} \text{ kg} \cdot \text{m}^2$.

The assumption of a single-vibration-mode arm, the *lumped mass model*, was verified numerically by calculating the distributed mass model of the arm, taking the linear density of the wire to be $1.4 \cdot 10^{-3} \text{ kg/m}$. The frequency of the lumped mass model is $\omega_0 = 6.614 \text{ rad/s}$ and the motor friction is $V = 1.374 \cdot 10^{-3} \text{ N} \cdot \text{m/rad/s}$. Therefore, the transfer functions for motor and beam are

$$\hat{G}_m(s) = \frac{395.008}{s(s + 2.1961)}, \quad G_b(s) = \frac{43.75}{s^2 + 43.75}. \quad (18)$$

The motor angle, $\theta_m(t)$, can be measured with an encoder, and the tip angle, $\theta_t(t)$, with a camera that determines, in real time, the x - y position of a LED placed on the tip of the arm.

5.2 Inner-loop control design

With the assumption that Coulomb friction and coupling between motor and beam have been compensated by adding the current of Eq. (5) to the motor current, an inner loop PD controller was designed using the root locus technique. The resulting controller is $G_c(s) = 2.019s + 560.605$, and the transfer function for the inner loop, with the derivative applied to the output (tachometer structure), is

$$G_{in}(s) = \frac{2.215 \cdot 10^5}{s^2 + 800s + 2.215 \cdot 10^5} \quad (19)$$

with poles in $s = -400 \pm j248$. To a step change in its reference the system has time constant τ about 6m. Then $1/\tau \gg \omega_0$, which guarantees that the dynamics of the motor with the inner loop closed is much faster than the dynamics of the beam (7). It may thus be assumed that the equivalent transfer function of the inner loop is unity. Notice that this behavior is independent of the mass placed at the tip, because the coupling torque between the motor and the arm is compensated by using the first equality of expression (4), which only depends on the measured angles θ_m and θ_t , C being a constant.

5.3 Outer-loop control design

With the control scheme of Fig. 5 and a controller given by expression (13), the closed-loop transfer function is

$$H(s) = \frac{\theta_t(s)}{\theta_i(s)} = \frac{43.75K}{s^{2-\beta} + 43.75K}, \quad (20)$$

and its step response is

$$\theta_t(t) = 43.75Kt^{2-\beta} E_{2-\beta, 3-\beta}(-43.75Kt^{2-\beta}). \quad (21)$$

The design of the controller thus involves the selection of two parameters:

- β , the order of the derivative, which determines: (a) the overshoot of the step response, (b) the phase margin, or (c) the damping
- K , the controller gain, which determines for a given β : (a) the speed of the step response, or (b) the crossover frequency

To select these parameters, one may work in the complex plane, the frequency domain or the time domain. In the frequency domain, the selection can be regraded as choosing a fixed phase margin by selecting β , and choosing a crossover frequency ω_c , by selecting K for a given β . That is,

$$\beta = \frac{2}{\pi} \phi_m, \quad K = \frac{\omega_c^{2-\beta}}{43.75}. \quad (22)$$

In our case the controller parameters will be chosen in the frequency domain approach for the following specifications: phase margin $\phi_m = 76.5^\circ$, and crossover frequency $\omega_c = 27$ rad/s. With these specifications, and applying expressions in (22), the controller parameters are $K = 1$ and $\beta = 0.85$.

5.4 Fractional controller implementations

Evidently, if no physical device is available to perform the fractional derivative, approximations are needed to implement the fractional controller. Here we use some of the approximations studied in [10]. For a continuous implementation, we use a frequency domain identification technique. An integer-order transfer function is obtained which fits the frequency response of the fractional-order derivative controller in the range $\omega \in (10^{-1}, 10^2)$. The resulting controller will be denoted $R_{ID}(s)$. For discrete implementations with sample period $T = 0.003$ s, we use three different methods: (i) discretization of the above continuous approximation by using the Tustin rule with pre-warping, the resulting controller being denoted $R_{IDZ}(z)$; (ii) direct discretization of the fractional operator by using continuous fraction expansion (CFE) of the Tustin discrete equivalent of the

Laplace operator s , the resulting controller being denoted $R_{TCFE}(z)$; and (iii) direct discretization of the fractional operator by using a Taylor series expansion of the backward discrete equivalent of the Laplace operator s , that is, using the truncated Grünwald–Letnikov formula (1), the resulting controller being denoted $R_{GL}(z)$. In comparing the results, it must be borne in mind that the controllers $R_{ID}(s)$, $R_{IDZ}(z)$, and $R_{TCFE}(z)$ are 7th-order analogical or digital IIR filters, and controller $R_{GL}(z)$ is a 100th-order FIR filter.

5.5 Step responses

The response of the flexible arm to a step reference with the fractional controller $R(s) = s^{0.85}$ has an overshoot of $M_p \approx 5\%$, rise time $t_r \approx 0.09$ s, and peak time $t_p \approx 0.14$ s. Fig. 6 shows the corresponding step responses of the controlled system with ideal and nonideal inner loop for the different approximations described previously. In Table 1 the corresponding step response characteristics of the controlled system are presented. Finally, Fig. 7 shows the step responses obtained with a traditional continuous PD controller (tuned for nominal mass and specified phase margin and crossover frequency) in the outer loop, and the ideal inner loop. One sees that the requirement of constant overshoot is not satisfied even for this ideal case.

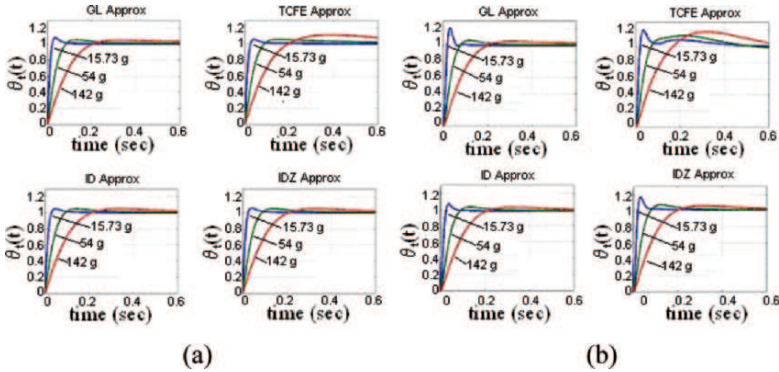
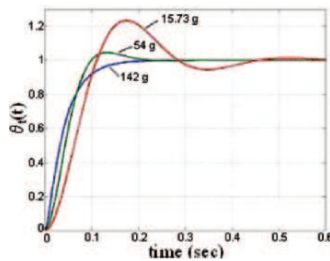


Fig. 6. (a) Step responses of the controlled system with ideal inner loop ($G_{IN}(S) = 1$); (b) Step responses of the controlled system with ideal nonideal inner loop ($G_{IN}(S)$ IN (20)).

Table 1. Summarized step response characteristics of the controlled system with ideal inner loop/nonideal inner loop

Ideal/non-ideal inner loop	Overshoot (%)			Peak time (s)			Rise time (s)		
	m_1	m_2	m_3	m_1	m_2	m_3	m_1	m_2	m_3
R_{GL}	8.1/21.7	5.6/6.0	5.2/5.0	0.04/0.03	0.14/0.12	0.33/0.32	0.02/0.02	0.08/0.08	0.21/0.21
R_{TCFE}	6.2/20.0	6.2/13	12/8	0.04/0.04	0.16/0.22	0.38/0.32	0.02/0.02	0.09/0.07	0.20/0.17
R_{ID}	5.0/6.2	5.0/3.0	5.0/3.0	0.04/0.04	0.14/0.13	0.34/0.33	0.03/0.02	0.09/0.10	0.21/0.23
R_{IDZ}	6.2/16.5	5.2/7.0	5.1/6.0	0.04/0.03	0.14/0.12	0.34/0.32	0.03/0.02	0.09/0.08	0.21/0.19

**Fig. 7.** Step responses with ideal inner loop and PD controller in the outer loop.

6 Conclusions

A new method to control single-link lightweight flexible arms in the presence of changes in the load has been presented in this article. The overall controller consists of three nested control loops. Once the friction and other nonlinear effects have been compensated, the inner loop is designed, following [5], to give a fast motor response. The middle loop decouples the dynamics of the system, and reduces its transfer function to a double integrator. The fractional-derivative controller is used to shape the outer loop into the form of a fractional-order integrator. The result is a constant-phase system with, in the time domain, step responses exhibiting constant overshoot, independently of variations in the load. This control strategy can be viewed as a particular case of the QFT method using fractional-order controllers. An interesting feature of this control scheme is that the overshoot is independent of the tip mass. This allows a constant safety zone to be delimited for any given placement task of the arm, independently of the load being carried, thereby making it easier to plan collision avoidance.

Acknowledgments

This work has been financially supported by the Spanish Research Grants 2PR02A024 (Junta de Extremadura) and DPI 2003-03326 (Ministerio de Ciencia y Tecnología).

References

1. Feliu V, Rattan KS, Brown HB (1993) Control of flexible arms with friction in the joints, *IEEE Trans. Robotics Autom.*, 9(4):467–475.
2. Ge SS, Lee TH, Zhu G (1998) Improving regulation of a single-link flexible manipulator with strain feedback. *IEEE Trans. Robotics Automation*, 14(1):179–185.
3. Feliu JJ, Feliu V, Cerrada C (1999) Load adaptive control of single-link flexible arms based on a new modeling technique. *IEEE Trans. Robotics Automation*. 15(5):793–804.
4. Torfs DE, Vuerinckx R, Schoukens J (1998) Comparison of two feedforward design methods aiming at accurate trajectory tracking of the end point of a flexible robot arm. *IEEE Trans. Control Syst. Technol.*, 6(1):2–14.
5. Feliu V, Rattan KS, Brown HB (1990) Adaptive control of a single-link flexible manipulator. *IEEE Control Syst. Mag.*, 10(2):29–33.
6. Geniele H, Patel RV, Khorasani K (1997) End-point control of a flexible-link manipulator: theory and experiments. *IEEE Trans. Control Syst. Technol.*, 5(6):556–570.
7. Podlubny I (1999) *Fractional Differential Equations*. Academic Press, San Diego.
8. Miller KS, Ross B (1993) *An Introduction to the Fractional Calculus and Fractional Differential Equations*. Wiley, New York.
9. Oustaloup A (1995) *La Dérivation Non Entière. Théorie, Synthèse et Applications*. Hermès, Paris.
10. Vinagre BM, Podlubny I, Hernandez A, Feliu V (2000) Some approximations of fractional order operators used in control theory and applications. *Fract. Cal. Appl. Anal.* 3(3):231–248.

TUNING RULES FOR FRACTIONAL PID_s

Technical University of Lisbon, Instituto Superior Técnico, Department of Mechanical Engineering, GCAR, Av. Rovisco Pais 1, 1049-001 Lisboa, Portugal;

E-mail {dvalerio,sadacosta}@dem.ist.utl.pt. Duarte Valério was partially supported by Fundação para a Ciência e a Tecnologia, grant SFRH/BPD/20636/2004, funded by POCI 2010, POS_C, FSE and MCTES.

Abstract

In this paper, we propose a new tuning rule for fractional PID controllers. The rule is based on the Ziegler-Nichols method, which is one of the most widely used tuning rules. The proposed rule is simple and easy to use, and it provides good results for a wide range of processes.

Keywords

fractional PID controllers, Ziegler-Nichols method, tuning rules

1 Introduction

The fractional PID controller is a generalization of the classical PID controller. It is defined by the transfer function

$$C(s) = K_p \left(1 + \frac{I}{s} + Ds \right)$$

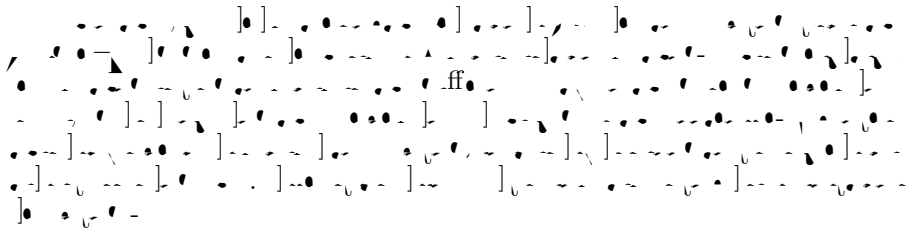
where K_p is the proportional gain, I is the integral gain, and D is the derivative gain. The fractional part is represented by the term $\frac{I}{s}$. The fractional PID controller has several advantages over the classical PID controller. It is more flexible and can be tuned to achieve better performance for a wide range of processes. It is also more robust to parameter variations and disturbances.

$$C(s) = K_p \left(1 + \frac{I}{s} + Ds \right)$$

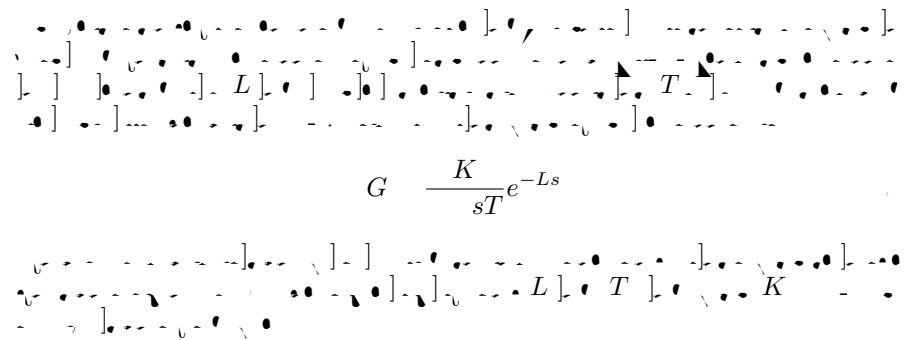
$$\begin{aligned}
 & \left[\frac{C}{s} - P - \frac{I}{s^\lambda} - Ds^\mu \right] \\
 & \left[\frac{\lambda}{s} - \mu \right] \\
 & \text{tuning rules}
 \end{aligned}$$

2 Tuning by Minimisation

$$\begin{aligned}
 & \left[C \omega_{cg} - G \omega_{cg} \right] \\
 & \left[\varphi_m - \pi - \varphi_m \right] \left[C \omega_{cg} - G \omega_{cg} \right] \\
 & \left[\omega_h \right] H, \\
 & \left| \frac{C \omega_h - G \omega_h}{C \omega_h - G \omega_h} \right| < H \\
 & \left[\omega_l \right] N, \\
 & \left| \frac{C \omega_l - G \omega_l}{C \omega_l - G \omega_l} \right| < N \\
 & \left. \frac{d}{d\omega} \right|_{\omega=\omega_{cg}} [C \omega - G \omega]
 \end{aligned}$$



3 A First Set of S-shaped Response-Based Tuning Rules



$$G = \frac{K}{sT} e^{-Ls}$$

$$\omega_{cg} = \frac{1}{L} \text{ / } \text{rad/s}$$

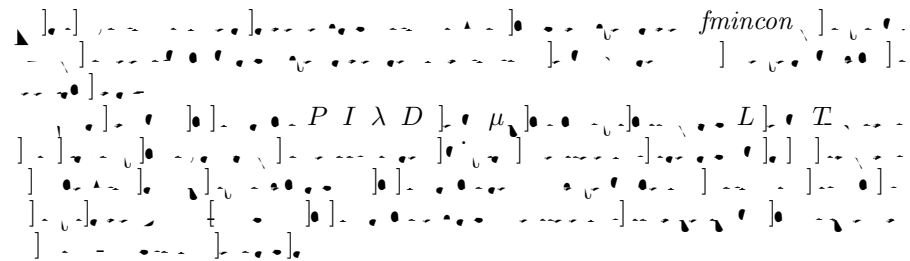
$$\varphi_m = -\frac{\pi}{2} \text{ / } \text{rad}$$

$$\omega_h = \frac{1}{L} \text{ / } \text{rad/s}$$

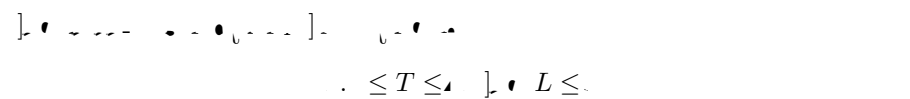
$$\omega_l = \frac{1}{L} \text{ / } \text{rad/s}$$

$$H = \frac{K}{L} \text{ / } \text{rad/s}$$

$$N = \frac{K}{L} \text{ / } \text{rad/s}$$



$$P = \frac{K}{L} \text{ / } \text{rad/s}, L = \frac{K}{L} \text{ / } \text{rad/s}, T = \frac{K}{L} \text{ / } \text{rad/s}$$



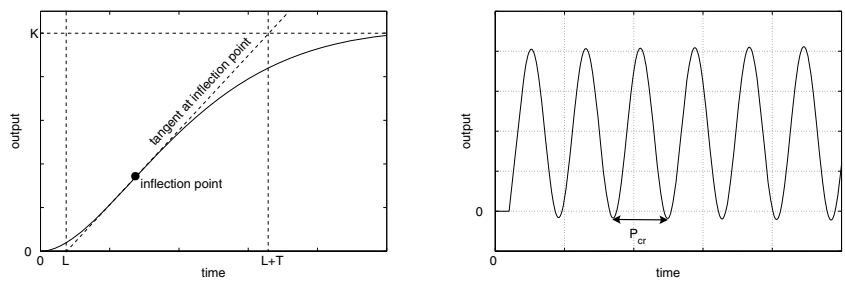


Fig. 1. Left: S-shaped unit-step response; right: plant output with critical gain control.

Table 1. Parameters for the first set of tuning rules for S-shaped response plants

Parameters to use when $0.1 \leq T \leq 5$					
	P	I	λ	D	μ
1	-0.0048	0.3254	1.5766	0.0662	0.8736
L	0.2664	0.2478	-0.2098	-0.2528	0.2746
T	0.4982	0.1429	-0.1313	0.1081	0.1489
L^2	0.0232	-0.1330	0.0713	0.0702	-0.1557
T^2	-0.0720	0.0258	0.0016	0.0328	-0.0250
LT	-0.0348	-0.0171	0.0114	0.2202	-0.0323
Parameters to use when $5 \leq T \leq 50$					
	P	I	λ	D	μ
1	2.1187	-0.5201	1.0645	1.1421	1.2902
L	-3.5207	2.6643	-0.3268	-1.3707	-0.5371
T	-0.1563	0.3453	-0.0229	0.0357	-0.0381
L^2	1.5827	-1.0944	0.2018	0.5552	0.2208
T^2	0.0025	0.0002	0.0003	-0.0002	0.0007
LT	0.1824	-0.1054	0.0028	0.2630	-0.0014

4 A Second Set of S-shaped Response-Based Tuning Rules

$$\begin{aligned} \omega_{cg} &= \frac{1}{P} \sqrt{\frac{1}{1 + \frac{D^2}{P^2}}} \\ \varphi_m &= \frac{1}{P} \sqrt{\frac{1}{1 + \frac{D^2}{P^2}}} \\ \omega_h &= \frac{1}{P} \sqrt{\frac{1}{1 + \frac{D^2}{P^2}}} \end{aligned}$$

ω_l	$\frac{1}{\sqrt{1+\lambda^2}}$	$\frac{1}{\sqrt{1+\lambda^2}}$	$\frac{1}{\sqrt{1+\lambda^2}}$	$\frac{1}{\sqrt{1+\lambda^2}}$	$\frac{1}{\sqrt{1+\lambda^2}}$
H	$\frac{1}{\sqrt{1+\lambda^2}}$	$\frac{1}{\sqrt{1+\lambda^2}}$	$\frac{1}{\sqrt{1+\lambda^2}}$	$\frac{1}{\sqrt{1+\lambda^2}}$	$\frac{1}{\sqrt{1+\lambda^2}}$
N	$\frac{1}{\sqrt{1+\lambda^2}}$	$\frac{1}{\sqrt{1+\lambda^2}}$	$\frac{1}{\sqrt{1+\lambda^2}}$	$\frac{1}{\sqrt{1+\lambda^2}}$	$\frac{1}{\sqrt{1+\lambda^2}}$

Table 2. Parameters for the second set of tuning rules for S-shaped response plants

	P	I	λ	D	μ
1	-1.0574	0.6014	1.1851	0.8793	0.2778
L	24.5420	0.4025	-0.3464	-15.0846	-2.1522
T	0.3544	0.7921	-0.0492	-0.0771	0.0675
L^2	-46.7325	-0.4508	1.7317	28.0388	2.4387
T^2	-0.0021	0.0018	0.0006	-0.0000	-0.0013
LT	-0.3106	-1.2050	0.0380	1.6711	0.0021

5 A First Set of Critical Gain-Based Tuning Rules

Consider the S-shaped response plant $G(s)$ with the critical gain K_{cr} and critical period P_{cr} . The critical gain K_{cr} is the gain at which the system becomes marginally stable. The critical period P_{cr} is the period of the sustained oscillation at the critical gain. The critical gain K_{cr} and critical period P_{cr} are determined by the system parameters P , I , λ , D , and μ . The critical gain K_{cr} and critical period P_{cr} are given by the following equations:

$$P_{cr} = \frac{1}{\omega_{cr}}$$

$$K_{cr} = \frac{1}{P_{cr}}$$

The critical gain K_{cr} and critical period P_{cr} are given by the following equations:

$$P_{cr} \leq \frac{1}{\omega_{cr}}, \quad K_{cr} P_{cr} \leq 1$$

Table 3. Parameters for the first set of tuning rules for plants with critical gain and period

Parameters to use when $0.1 \leq T \leq 5$					
	P	I	λ	D	μ
1	0.4139	0.7067	1.3240	0.2293	0.8804
K_{cr}	0.0145	0.0101	-0.0081	0.0153	-0.0048
P_{cr}	0.1584	-0.0049	-0.0163	0.0936	0.0061
$1/K_{cr}$	-0.4384	-0.2951	0.1393	-0.5293	0.0749
$1/P_{cr}$	-0.0855	-0.1001	0.0791	-0.0440	0.0810

Parameters to use when $5 \leq T \leq 50$					
	P	I	λ	D	μ
1	-1.4405	5.7800	0.4712	1.3190	0.5425
K_{cr}	0.0000	0.0238	-0.0003	-0.0024	-0.0023
P_{cr}	0.4795	0.2783	-0.0029	2.6251	-0.0281
$1/K_{cr}$	32.2516	-56.2373	7.0519	-138.9333	5.0073
$1/P_{cr}$	0.6893	-2.5917	0.1355	0.1941	0.2873

6 A Second Set of Critical Gain-Based Tuning Rules

$$P_{cr} \leq \frac{1}{\omega_{cr}} \sqrt{\frac{K_{cr}}{P_{cr}}} \quad (6.1)$$

Table 4. Parameters for the second set of tuning rules for plants with critical gain and period

	P	I	λ	D	μ
1	1.0101	10.5528	0.6213	15.7620	1.0101
K_{cr}	0.0024	0.2352	-0.0034	-0.1771	0.0024
P_{cr}	-0.8606	-17.0426	0.2257	-23.0396	-0.8606
P_{cr}^2	0.1991	6.3144	0.1069	8.2724	0.1991
$K_{cr}P_{cr}$	-0.0005	-0.0617	0.0008	0.1987	-0.0005
$1/K_{cr}$	-0.9300	-0.9399	1.1809	-0.8892	-0.9300
$1/P_{cr}$	-0.1609	-1.5547	0.0904	-2.9981	-0.1609
K_{cr}/P_{cr}	-0.0009	-0.0687	0.0010	0.0389	-0.0009
P_{cr}/K_{cr}	0.5846	3.4357	-0.8139	2.8619	0.5846

7 A Third Set of Critical Gain-Based Tuning Rules

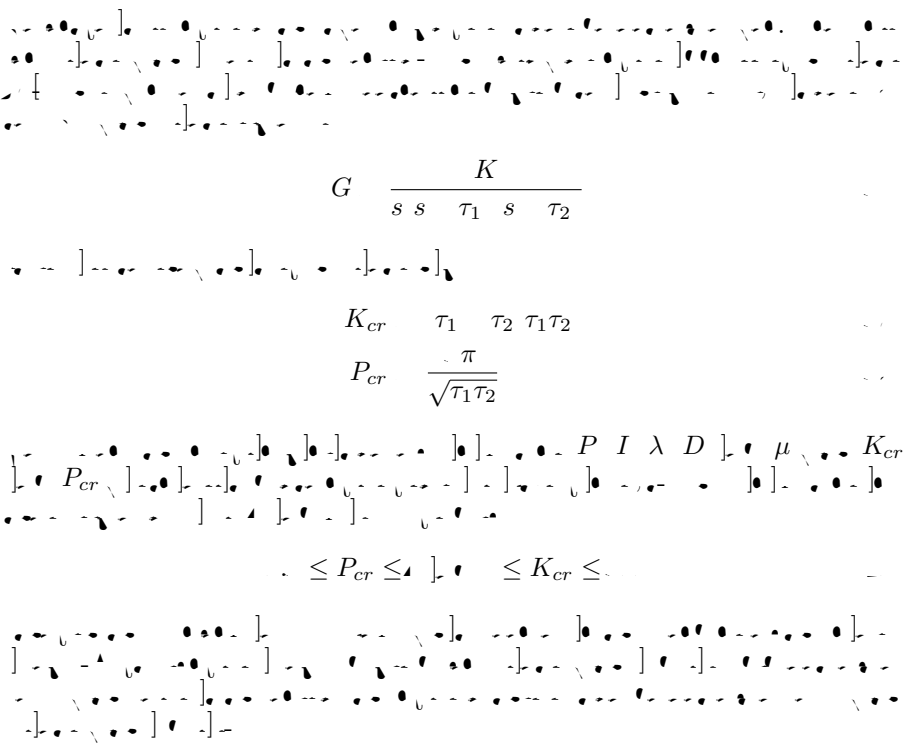
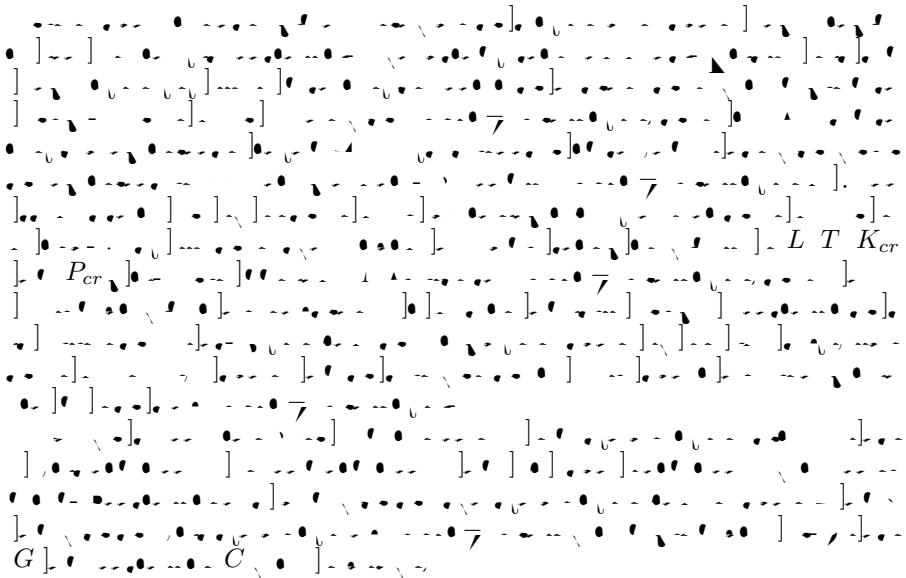


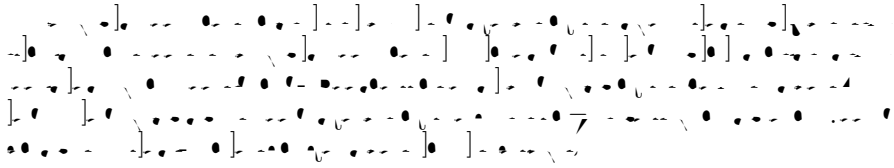
Table 5. Parameters for the third set of tuning rules for plants with critical gain and period

	P	I	λ	D	μ
1	-1.6403	-92.5612	0.7381	-8.6771	0.6688
K_{cr}	0.0046	0.0071	-0.0004	-0.0636	0.0000
P_{cr}	-1.6769	-33.0655	-0.1907	-1.0487	0.4765
$K_{cr}P_{cr}$	0.0002	-0.0020	0.0000	0.0529	-0.0002
$1/K_{cr}$	0.8615	-1.0680	-0.0167	-2.1166	0.3695
$1/P_{cr}$	2.9089	133.7959	0.0360	8.4563	-0.4083
K_{cr}/P_{cr}	-0.0012	-0.0011	0.0000	0.0113	-0.0001
P_{cr}/K_{cr}	-0.7635	-5.6721	0.0792	2.3350	0.0639
$\log_{10}(K_{cr})$	0.4049	-0.9487	0.0164	-0.0002	0.1714
$\log_{10}(P_{cr})$	12.6948	336.1220	0.4636	16.6034	-3.6738

8 Robustness



G_1	s	$\frac{K}{s}e^{-0.1s}$	-
C_{1a}	s	$\frac{1}{s^{1.4277}} \approx s^{1.0202}$	-
C_{1b}	s	$\frac{1}{s^{1.1230}} \approx s^{0.1533}$	--
C_{1c}	s	$\frac{1}{s}$	-
G_2	s	$\frac{K}{s^2} \approx \frac{K}{s}e^{-0.2s}$	-
C_{2a}	s	$\frac{1}{s^{0.6751}} \approx s^{0.6957}$	-
C_{2b}	s	$\frac{1}{s^{0.6000}} \approx s^{0.7805}$	-
C_{2c}	s	$\frac{1}{s}$	-
G_3	s	$\frac{K}{\sqrt{s}}e^{-0.5s} \approx \frac{K}{s}e^{-0.1s}$	-
C_{3a}	s	$\frac{1}{s^{1.3646}} \approx s^{1.0618}$	-
C_{3b}	s	$\frac{1}{s^{1.1011}} \approx s^{0.1855}$	-
C_{3c}	s	$\frac{1}{s}$	-



$$G_{4a}(s) = \frac{K}{s} e^{-0.2s}$$

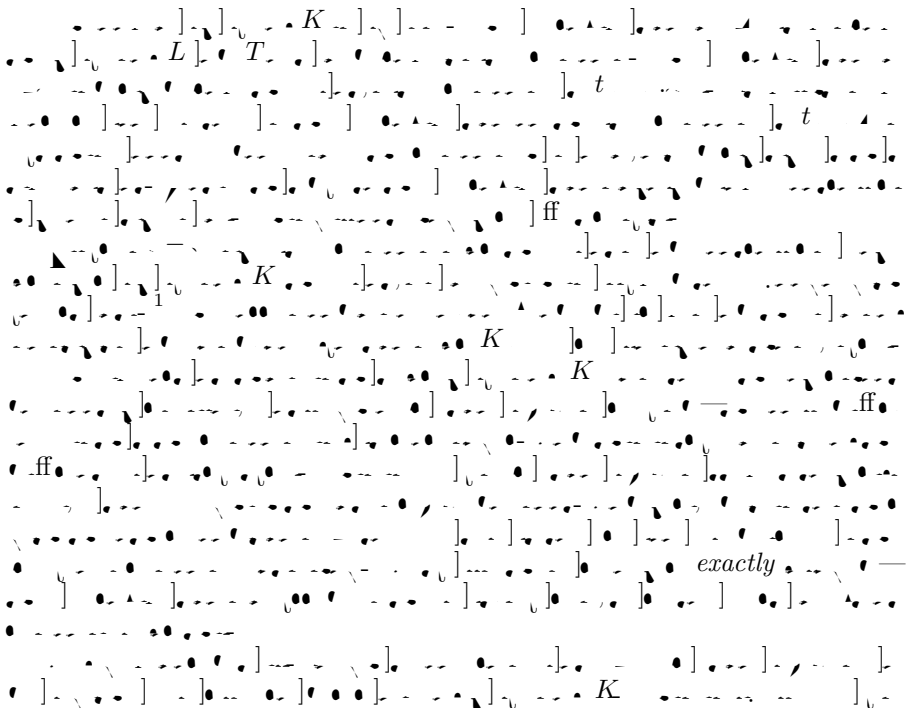
$$G_{4b}(s) = \frac{K}{s^3 + s^2 + s} \approx \frac{K}{s} e^{-0.2s}$$

$$C_{4a}(s) = \frac{1}{s^{0.6363}} \quad s^{0.5494}$$

$$C_{4b}(s) = \frac{1}{s^{0.7480}} \quad s^{0.3835}$$

$$C_{4c}(s) = \frac{1}{s^{0.5588}} \quad s^{1.2328}$$

$$C_{4d}(s) = \frac{1}{s} \quad s$$



¹ Those time-responses involving fractional derivatives and integrals were obtained using Oustaloup's approximations [4] for the fractional terms. Approximations were conceived for the frequency range $[\omega_l, \omega_h] = [10^{-3}, 10^3]$ rad/s and make use of 7 poles and 7 zeros.

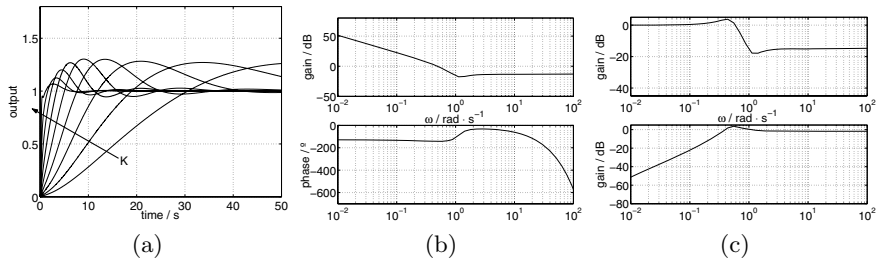


Fig. 2. (a) Step response of (31) controlled with (32) when K is 1/32, 1/16, 1/8, 1/4, 1/2, 1 (thick line), 2, 4, and 8. (b) Open-loop Bode diagram when $K=1$. (c) Closed-loop function gain (top) and sensitivity function gain (bottom) when $K=1$.

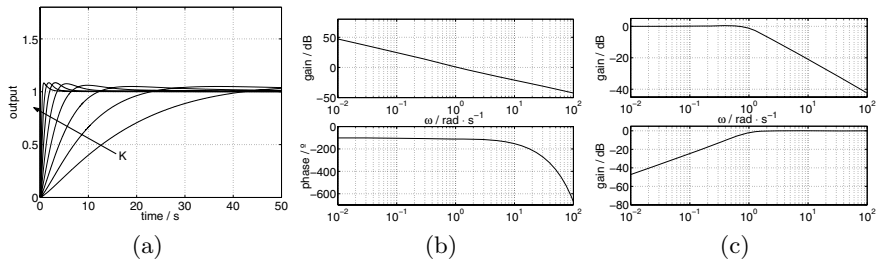


Fig. 3. (a) Step response of (31) controlled with (33) when K is 1/32, 1/16, 1/8, 1/4, 1/2, 1 (thick line), 2, and 4. (b) Open-loop Bode diagram when $K=1$. (c) Closed-loop function gain (top) and sensitivity function gain (bottom) when $K=1$.

ff. K K /

9 Conclusions

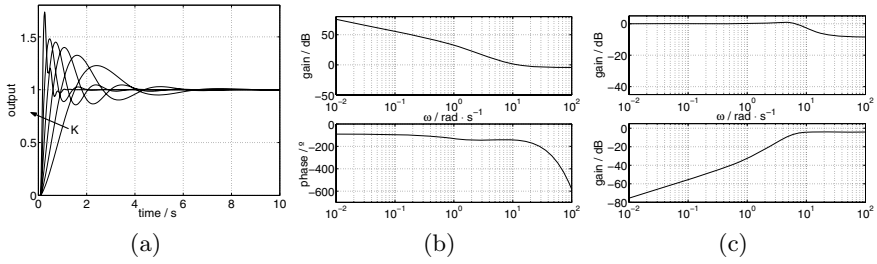


Fig. 4. (a) Step response of (31) controlled with (34) when K is $1/32$, $1/16$, $1/8$, $1/4$, $1/2$, and 1 (thick line). (b) Open-loop Bode diagram when $K=1$ (c) Closed-loop function gain (top) and sensitivity function gain (bottom) when $K=1$.

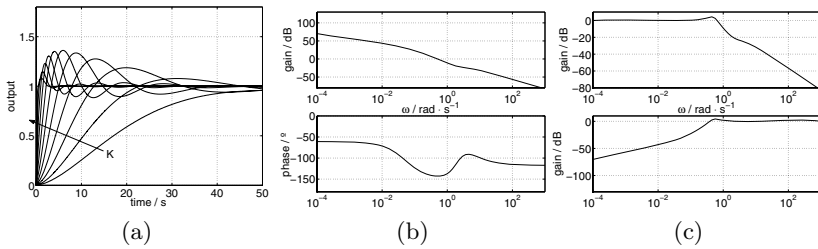


Fig. 5. (a) Step response of (35) controlled with (36) when K is $1/32$, $1/16$, $1/8$, $1/4$, $1/2$, 1 (thick line), 2 , 4 , 8 , 16 , and 32 . (b) Open-loop Bode diagram when $K=1$. (c) Closed-loop function gain (top) and sensitivity function gain (bottom) when $K=1$.

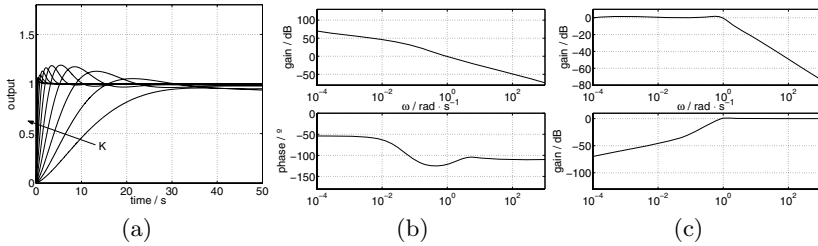
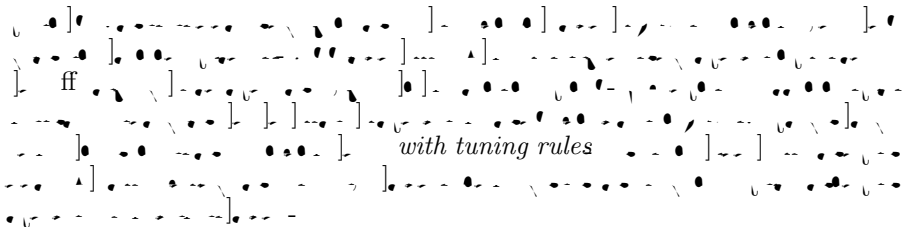


Fig. 6. (a) Step response of (35) controlled with (37) when K is $1/32$, $1/16$, $1/8$, $1/4$, $1/2$, 1 (thick line), 2 , 4 , 8 , 16 , and 32 . (b) Open-loop Bode diagram when $K=1$. (c) Closed-loop function gain (top) and sensitivity function gain (bottom) when $K=1$.



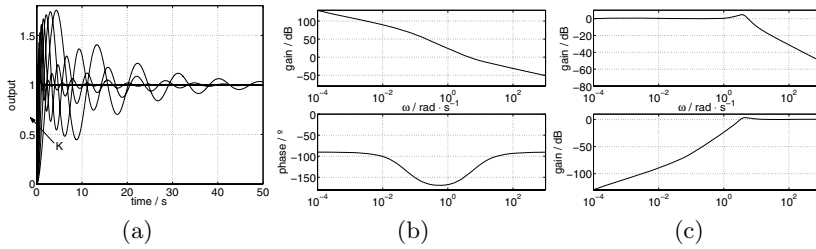


Fig. 7. (a) Step response of (35) controlled with (38) when K is $1/32, 1/16, 1/8, 1/4, 1/2, 1$ (thick line), $2, 4, 8, 16$, and 32 . (b) Open-loop Bode diagram when $K = 1$. (c) Closed-loop function gain (top) and sensitivity function gain (bottom) when $K = 1$.

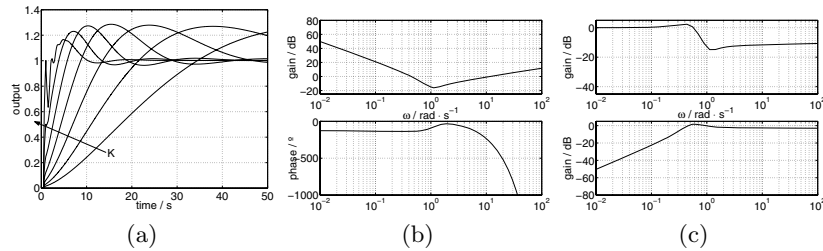


Fig. 8. (a) Step response of (39) controlled with (40) when K is $1/32, 1/16, 1/8, 1/4, 1/2, 1$ (thick line), and 2 . (b) Open-loop Bode diagram when $K = 1$. (c) Closed-loop function gain (top) and sensitivity function gain (bottom) when $K = 1$.

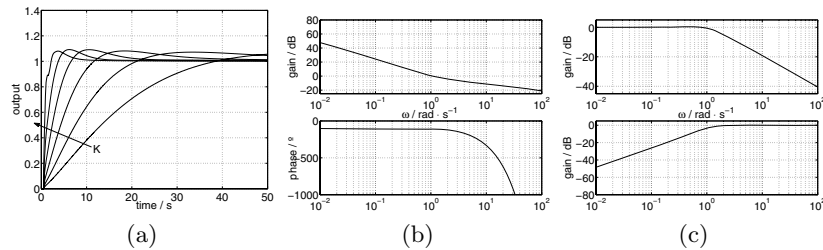


Fig. 9. (a) Step response of (39) controlled with (41) when K is $1/32, 1/16, 1/8, 1/4, 1/2$, and 1 (thick line). (b) Open-loop Bode diagram when $K = 1$. (c) Closed-loop function gain (top) and sensitivity function gain (bottom) when $K = 1$.

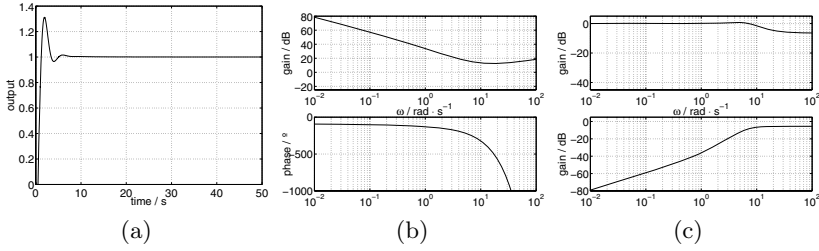


Fig. 10. (a) Step response of (39) controlled with (42) when K is $1/32$. (b) Open-loop Bode diagram when $K = 1$. (c) Closed-loop function gain (top) and sensitivity function gain (bottom) when $K = 1$.

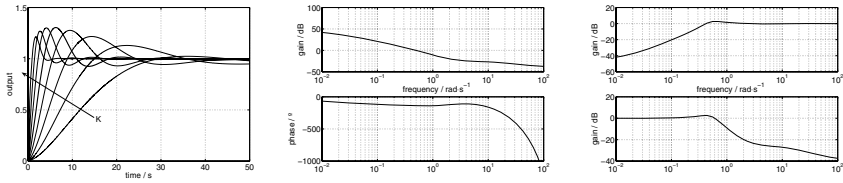


Fig. 11. Left: Step response of (43) controlled with (45) when K is $1/16, 1/8, 1/4, 1/2, 1$ (thick line), $2, 4$, and 8 ; centre: open-loop Bode diagram when $K = 1$; right: sensitivity function gain (top) and closed-loop gain (bottom) when $K = 1$.

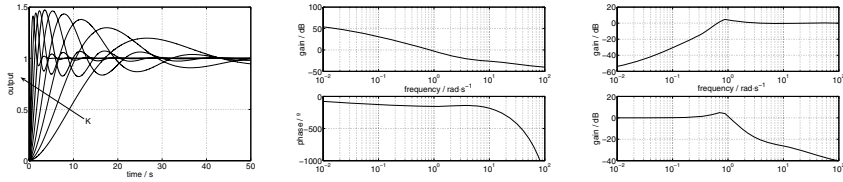


Fig. 12. Left: Step response of (43) controlled with (46) when K is $1/32, 1/16, 1/8, 1/4, 1/2, 1$ (thick line), $2, 4$, and 8 ; centre: open-loop Bode diagram when $K = 1$; right: sensitivity function gain (top) and closed-loop gain (bottom) when $K = 1$.

Acknowledgment

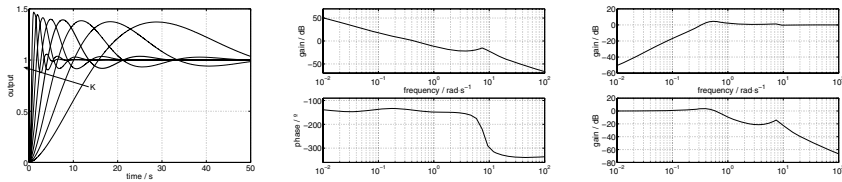


Fig. 13. Left: Step response of (44) controlled with (47) when K is 1/8, 1/4, 1/2, 1 (thick line), 2, 4, 8, and 16; centre: open-loop Bode diagram when $K = 1$; right: sensitivity function gain (top) and closed-loop gain (bottom) when $K = 1$.

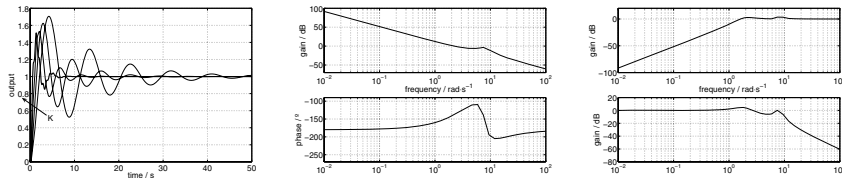


Fig. 14. Left: Step response of (44) controlled with (48) when K is 1/4, 1/2, 1 (thick line), 2, 4, and 8; centre: open-loop Bode diagram when $K = 1$; right: sensitivity function gain (top) and closed-loop gain (bottom) when $K = 1$.

References

1. Caponetto R, Fortuna L, Porto D (2002) Parameter tuning of a non integer order PID controller. In Electronic proceedings of the 15th International Symposium on Mathematical Theory of Networks and Systems, University of Notre Dame, Indiana.
2. Caponetto R, Fortuna L, Porto D (2004) A new tuning strategy for a non integer order PID controller. In First IFAC Workshop on Fractional Differentiation and its Applications, Bordeaux.
3. Monje CA, Vinagre BM, Chen YQ, Feliu V, Lanusse P, Sabatier J (2004) Proposals for fractional PID tuning. In First IFAC Workshop on Fractional Differentiation and its Applications, Bordeaux.
4. Oustaloup A (1991) *La commande CRONE: commande robuste d'ordre non entier*. Hermès, Paris, in French.
5. Podlubny I (1999) *Fractional Differential Equations*. Academic Press, San Diego.
6. Valério D, da Costa JS (2005) Ziegler-nichols type tuning rules for fractional PID controllers. In Proceedings of ASME 2005 Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Long Beach.
7. Valério D, da Costa JS (2006) Tuning of fractional PID controllers with ziegler-nichols type rules. Signal Processing. Accepted for publication.
8. Valério D, da Costa JS (2006) Tuning rules for fractional PID controllers. In Fractional Differentiation and its Applications, Porto.
9. Vinagre B (2001) Modelado y control de sistemas dinámicos caracterizados por ecuaciones íntegro-diferenciales de orden fraccional. PhD thesis, Universidad Nacional de Educación a Distancia, Madrid, In Spanish.

FREQUENCY BAND-LIMITED FRACTIONAL DIFFERENTIATOR PREFILTER IN PATH TRACKING DESIGN

Pierre Melchior, Alexandre Poty, and Alain Oustaloup

LAPS, UMR 5131 CNRS, Université Bordeaux 1, ENSEIRB, 351 cours de la Libération, F33405 TALENCE Cedex, France; Tél: +33 (0)5 40 00 66 07, Fax: +33 (0)5 40 00 66 44, E-mail: surname.name@laps.u-bordeaux1.fr, URL: <http://www.laps.u-bordeaux1.fr>

Abstract

A new approach to path tracking using a fractional differentiation prefilter applied to nonvarying plants is proposed in this paper. In previous works, a first approach, based on a Davidson–Cole prefilter, has been presented; it permits the generation of optimal movement reference input leading to a minimum path completion time, taking into account both the physical actuators constraints (maximum velocity, acceleration, and jerk) and the bandwidth of the closed-loop system. In this paper, an extension of this method is presented: the reference input results from the step response of a frequency band-limited fractional differentiator (FBLFD) prefilter whose main properties are having no overshoot on the plant and to have maximum control value for starting time. Moreover, it can be implemented as a classical digital filter. A simulation on a motor model validates the methodology.

Keywords

Fractional prefilter, Davidson–Cole filter, motion control, path tracking, control, fractional systems, testing bench.

1 Introduction

To increase the speed of machine tools, lighter materials are used increasing their flexibility. Execution times must be optimized without exciting resonance. A prefilter is used in industrial path tracking designs, as it is easy to implement and adapt for reducing overshoots. This reduces the high-frequency energy of the path planning signal using a low-pass filter with trial-and-error determined parameters. Nevertheless, for classic linear prefilter

approaches, when overshoots are reduced, dynamic performances are also reduced. This type of path tracking, based on position step filtering, does not permit separate control over maximal values of velocity and acceleration, which stay proportional to the amplitude of the step applied.

Using a time domain bound in the frequency domain is difficult [4]. It is not therefore easy to take into account actuator saturations in the design of prefilters. So, the prefilter often only narrows the frequency band of the control loop reference input.

For the polynomial approach [3, 5], maximal values of velocity and acceleration cannot be kept. The path completion time is thus over optimal. The Bang–Bang approach [3, 5] takes into account the same physical constraints but does provide a minimal path completion time. However, as for the polynomial approach, the dynamics (bandwidth) of the control loop are not taken into account, so overshoots can appear on the end actuator. Cubic spline functions (order 3 piecewise polynomials) are now widely used in robotics. They are minimal curvature curves [3] and the optimization proposed by Lin [7], or De Luca, [8], based on the nonlinear simplex optimization algorithm [15] offers a complete-path reference solution. However, as in the polynomial or Bang–Bang approach, the dynamics of the control loop are not taken into account: overshoots on the end actuator appear for small displacements. Thus, to limit actuator saturation during transitions, the actuator dynamics must be taken into account, so the above techniques are often combined with a prefilter.

When the mathematical expression of the trajectory is known, and the control loop is defined perfectly, algorithms of Shin and Mac Kay [21, 22] or Bobrow [1] allow synthesis of the optimal control which takes into account constraints on the control inputs and the details of the manipulator dynamics. The dynamic model of the process must be designed by applying Lagrange formalism. The use of curvilinear abscissa allows reduction of the number of variables without loss of information. The minimal path time is determined from the phase curve using the Pontryagin maximum principle. However, this is fastidious and must be done for each trajectory. Furthermore, for such tasks as painting or cloth cutting, the trajectories are very complex and numerous. So, the algebraic calculus takes too long without providing tracking accuracy [6]. Moreover, during this time the task will not take place.

In the aerospace industry, flexible mode frequencies are well defined, but weakly damped. Here, the input shaper technique reduces vibration in path tracking design. Input shaping is obtained by convolving desired input with an impulse sequence. This generates vibration-reducing shaped command, which is more effective than conventional filters [23].

When the target is unknown, nonlinearity, such as saturation, causes the integral of the error to accumulate to a much larger value than in the linear

case. This large integrated error, known as integral windup, causes a large percentage overshoot and a long settling time. The aim of antiwindup compensation is thus to modify the dynamics of a control loop when control signals saturate [16]. This technique uses a fast control loop but no prefilter: the control loop reference input signal is equal to a real-time target position. The antiwindup compensation does not take into account the reference input when the trajectory is known in advance. Here, the actuator must not be saturated, so there is no need for antiwindup compensation.

A recent approach to path tracking using this fractional (or noninteger order) derivative [13, 18, 20] have been developed by Melchior [11, 12, 17]. With a Davidson–Cole [2, 9] prefilter, the reference-input results from its step response. It is thus possible to limit the resonance of the feedback control loop, by a continuous variation on τ , but also on n . It permits the generation of optimal movement reference-input, leading to a minimum path completion time, taking into account both the physical constraints of the actuators (maximum velocity, acceleration, and torque) and the bandwidth of the closed-loop system. The filter can be implemented as a classical digital filter. It is synthesized in the frequency domain, thus the power spectral density of the position permits absolute control of the high-frequency energy. To separate speed and acceleration control, a Davidson–Cole speed filter has been developed allowing intermediate speed control for path tracking [11]. As the spline function, made up of one jerk step per point, is a reference in robotics, a Davidson–Cole jerk filter has also been developed [11]. However, only the control loop reference input is optimized, and not the plant output.

In this paper, a method based on a FBLFD prefilter is proposed to optimize plant output. The FBLFD type transfer function (Fig. 2) properties are having no overshoot on the plant output and by adding numerator to have maximum control value for starting time. These properties are available whatever the values of its constitutive parameters (n, τ_1, τ_2) which are optimized by minimizing the output settling time of the plant, by maximizing the bandwidth energy transfer between the input and the output, and including the time domain bound on the control signal.

The transmission of energy from input to control is maximized. Overshoots are avoided on the control signal by including a frequency bound on the transfer function.

The remainder of this paper is divided as follow. Section 2, defines the generalized differentiator. Section 3 presents the FBLFD and the prefilter synthesis methodology. Section 4 gives simulation performances obtained using this method on a Parvex RX 120 DC motor. Finally, a conclusion is given in section 5.

2 Generalized Differentiator

2.1 Definition

A differentiator of any order n , called generalized differentiator [6,7], is such that its output magnitude $s(t)$ is proportional to the n th derivative of its input magnitude $e(t)$, that is

$$s(t) = \tau^n e^n(t) = \tau^n D^n e(t), \quad (1)$$

where $D = d/dt$, and n can be integer, noninteger, real or complex. τ is the and positive time differentiation constant raised to the n th power to simplify both the canonical transmittance (3) and the expression of the unit gain frequency which is a characteristic of the differentiator within the frequency domain.

Since differentiation using an order n with a negative a part is simply integration, generalized differentiator, or generalized integrator terms can be used indifferently.

2.2 Symbolic characterization

Assuming that the initial conditions are null, translation of time equation (1) into the s domain determines the symbolic equation

$$S(s) = \tau^n s^n E(s), \quad (2)$$

which gives the transmittance

$$D(s) = (\tau s)^n. \quad (3)$$

The frequency response corresponding to transmittance $D(s)$ is of the form:

$$D(j\omega) = (j\omega)^n, \quad (4)$$

or

$$D(j\omega) = \left(j \frac{\omega}{\omega_0} \right)^n, \quad (5)$$

assuming that $\omega_0 = 1/\tau$ which is called unit gain frequency or transition frequency.

Gain and phase are given by:

$$|D(j\omega)| = \left(\frac{\omega}{\omega_0} \right)^n \quad (6)$$

and

$$\arg D(j\omega) = n \frac{\pi}{2}. \quad (7)$$

Figure 1 represents Bode diagrams in the case of a positive real differentiation order. The gain increases by $6n$ dB per octave.

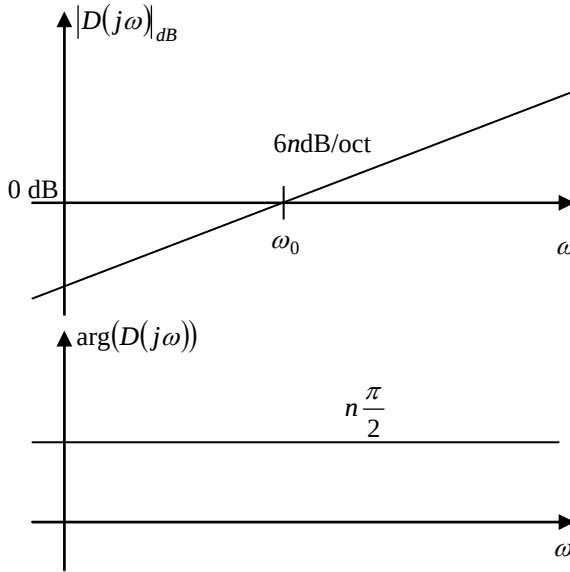


Fig. 1. Bode diagrams of a positive real order differentiator.

3 Frequency Band-Limited Fractional Differentiator

3.1 Introduction

For a single input–single output (SISO) path tracking design (Fig. 2), the filter $F(s)$ decouples the dynamics behaviors in position control and regulation.

However, the accuracy on the output position depends on the controller efficiency to reject noise and disturbances. Also, to allow the controller to reduce effects of these unexpected signals, the power spectral density of the reference input must be within the sensitivity bandwidth.

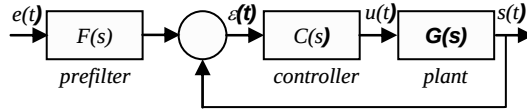


Fig. 2. Filtered unity-feedback control loop.

The transfer function, $A(s)$, of the filtered unity-feedback control is given by:

$$A(s) = \frac{S(s)}{E(s)} = F(s) \frac{C(s)G(s)}{1 + C(s)G(s)}, \quad (8)$$

where $F(s)$, $C(s)$, and $G(s)$ are the transfer functions of the filter, the controller and the plant.

Good tracking performances require that $S_e(s)$, the sensitivity transfer function

$$S_e(s) = \frac{1}{1 + C(s)G(s)} \quad (9)$$

be small in magnitude for small frequencies, so that effect of disturbances is attenuated.

It is also required that $T(s)$, the complementary sensitivity transfer function:

$$T(s) = \frac{C(s)G(s)}{1 + C(s)G(s)}, \quad (10)$$

be small in magnitude for large frequencies, so that effect of the sensor noise is attenuated, and be unity for small frequencies to follow asymptotically the reference input.

The transfer function between control and input is called reference sensibility transfer function:

$$S_1(s) = \frac{U(s)}{E(s)} = F(s) \frac{C(s)}{1 + C(s)G(s)} = \frac{A(s)}{G(s)}. \quad (11)$$

Final value theorem leads to a condition for having a maximum static constant value, $u_{\max}$, on the control signal in response to a constant signal $e_{\max}$ applied on the prefilter input:

$$\lim_{s \rightarrow 0} S_1(s) = \frac{u_{\max}}{e_{\max}}. \quad (12)$$

Otherwise, comparison of expressions (10) and (11) gives:

$$S_1(s) = F(s) \frac{T(s)}{G(s)}. \quad (13)$$

The transfer function of the filter could also be expressed:

$$F(s) = \frac{G(s)S_1(s)}{T(s)}. \quad (14)$$

As the complementary sensivity transfer function verifies:

$$\lim_{s \rightarrow 0} T(s) = 1, \quad (15)$$

the static behavior of the filter transfer function can also be deduced:

$$\lim_{s \rightarrow 0} F(s) = \lim_{s \rightarrow 0} \frac{G(s) S_1(s)}{T(s)}, \quad (16)$$

and using expressions (13) and (15):

$$\lim_{s \rightarrow 0} F(s) = \lim_{s \rightarrow 0} G(s) \frac{u_{\max}}{e_{\max}}. \quad (17)$$

It is now convenient to break down the plant transfer function into:

$$G(s) = G_0(s) G_1(s), \quad (18)$$

where

$$G_0(s) \triangleq \lim_{s \rightarrow 0} G(s), \quad (19)$$

which correspond to the static behavior of the plant.

It is also deduced from (17), that if the low-frequency behavior of the plant is:

$$G_0(s) = \frac{K_0}{s^m}, \quad (20)$$

the low-frequency behavior of the prefilter must be:

$$\lim_{s \rightarrow 0} F(s) = \frac{K_0}{s^m} \frac{u_{\max}}{e_{\max}}. \quad (21)$$

Thus, the low-frequency integration number of the prefilter and the plant must be the same. This result is used in section 3.2.1 to fix the structure of the FBLFD transfer function for the path tracking design.

3.2 Frequency band-limited fractional differentiator prefilter characteristics

In the S domain, a fundamental system is qualified as a FBLFD system when it is defined by the generalized transmittance:

$$F(s) = \left(\frac{1 + \tau_2 s}{1 + \tau_1 s} \right)^n \quad \text{with } n \in \mathbb{C}, \quad (\tau_1, \tau_2) \in \mathbb{R}^2. \quad (22)$$

The expression given by (22), where parameter n is real and no longer restricted to be an integer, is an FBLFD filter [18].

The impulse response of a band-limited fractional differentiator system is:

$$s_{\text{imp}}(t) = \left(\frac{\tau_2}{\tau_1} \right)^n [\delta(t) + \sum_{k=1}^{\infty} \frac{(-1)^k}{k!(k-1)!} \frac{\Gamma(n+k)}{\Gamma(n)} \left(\frac{1}{\tau_1} - \frac{1}{\tau_2} \right)^k t^{k-1} e^{-\frac{t}{\tau_1}} u(t)]. \quad (23)$$

The step response is obtained by integration of relation (23):

$$\dot{q}(t) = \left(\frac{\tau_2}{\tau_1} \right)^n [u(t) + \sum_{k=1}^{\infty} \frac{(-1)^k}{k!(k-1)!} \frac{\Gamma(n+k)\Gamma(k)}{\Gamma(n)} \left(\frac{\tau_2}{\tau_1} - 1 \right)^k \left(1 - \frac{\Gamma(k, t/\tau_2)}{\Gamma(k)} \right)], \quad (24)$$

where

$$\Gamma(n, t/\tau_2) = \int_0^{t/\tau_2} x^{n-1} e^{-x} dx, \quad (25)$$

is the incomplete gamma function.

The FBLFD position filter methodology defines analytic profile expressions of position, speed, acceleration, and their maxima, using only three parameters (n, τ_1, τ_2) .

In the approach presented, only values (τ_1, τ_2) such as $0 < \tau_2 < \tau_1$ are considered and $n \in \mathbb{R}$. The use of real poles prevents frequency resonance and the choice of identical poles allows the greatest possible energy on a given bandwidth (Fig. 3).

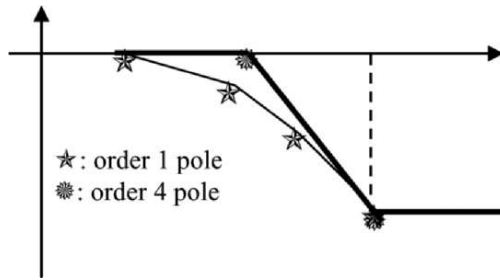


Fig. 3. Pole assignment for a maximum energy in a given pass-band.

The filter given by (22) reduces energy of the signal at high frequencies by defining bandwidth (time constants τ_1 and τ_2) and, through the continuous nature of the selectivity (real order n) as can be seen in Fig. 4. The optimization of parameters (n, τ_1, τ_2) considers the static constraints $(V_{\max}, A_{\max}, J_{\max})$ and the dynamic constraints $(\omega_c < \omega_r)$ to reduce resonance.

Figure 4 represents power spectral density assignment of the FBLFD filter compared to resonance frequency placement of the control loop which is applied.

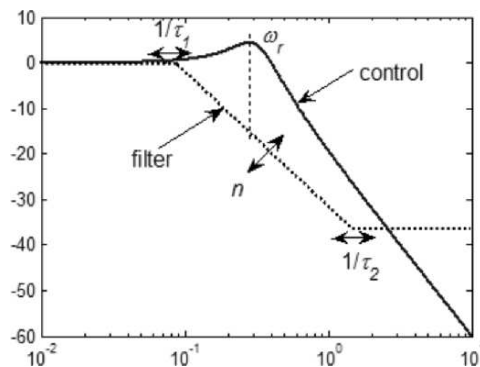


Fig. 4. Power spectral density assignment of the FBLFD filter compared to resonance frequency placement of the control loop which is applied.

A method is now proposed using an FBLFD prefilter to optimize output plant. This method is an extension of Davidson–Cole prefilter synthesis [17].

The main characteristics of the FBLFD prefilter are:

- No overshoot on the plant output
- Maximum bandwidth energy
- And maximum control at starting time

3.2.1 Time domain bound into frequency domain

From Eq. (13), the frequency constraint which keeps the control signals below its maximum value, is:

$$\forall \omega > 0, \left| \frac{1 + j\tau_2\omega}{1 + j\tau_1\omega} \right|^n \frac{C(s)}{1 + C(s)G(s)} < \frac{u_{\max}}{e_{\max}}. \quad (26)$$

This expression is interesting as:

- the I/O transfer function can be designed without knowing which controller is used
- the first corner frequency of the plant limits the I/O transfer function bandwidth

The high frequency $\omega_2 = 1/\tau_2$ can be expressed in function of (n, τ_1) thanks to the initial value theorem:

$$\lim_{t \rightarrow 0} h(t) = \lim_{s \rightarrow \infty} sH(s). \quad (27)$$

So, applying (27) to (26), it leads to

$$\tau_2 = \left[\lim_{s \rightarrow \infty} \left(\frac{1 + C(s)G(s)}{C(s)} \right) \frac{u_{\max}}{e_{\max}} \right]^{1/n} \tau_1. \quad (28)$$

Thanks to (28), only the two parameters (n, τ_1) have to be found.

3.2.2 Integral gap optimization

The fastest FBLFD transfer function is now to be determined. Using the frequency constraint (26), saturation of the control input signal is avoided. Integral gap is often used to determine the dynamic performance of a step response without overshoot. The integral gap analytic expression for the FBLFD step response is [17]:

$$I_e = n(\tau_1 - \tau_2). \quad (29)$$

Remark: if $\tau_2 = 0$ s., the integral gap analytical expression is, for a Davidson–Cole transfer function $I_e = n\tau_1$.

The optimal values of parameters (n, τ_1, τ_2) are found by the Matlab toolbox optimization (using fmincon function for example). The fractional prefilter being known, by using the identification unit of CRONE software [10], a simple expression of $F(s)$ can be found.

4 Simulation on a DC Motor PARVEX RX 120

The DC motor PARVEX RX 120 characteristics are given in Table 1.

Table 1. Parvex RX 120 DC motor characteristics

Motor characteristic	Value
Inertia moment (J)	$5 \cdot 10^{-5} \text{ kg.m}^2$
Viscous friction (f)	$4.2 \cdot 10^{-5} \text{ m.s/rd}$
Electromagnetic torque ratio (K_c)	0.11 V.m/A
Induced inductance (L)	$7.5 \cdot 10^{-3} \text{ H}$
Induced resistance (r)	2.5Ω
Amp/volt ratio (K_i)	1.93 A/V
Maximal control (u_{sat})	3 V

The plant modelization and the identification of the various parameters lead to the following transfer function:

$$G(s) = \frac{K_0}{1 + 2z \frac{s}{\omega_n} + \frac{s^2}{\omega_n^2}}, \quad (30)$$

where $K_0 = 750 \text{ rad/s/V}$, $\omega_n = 0.476 \text{ rad/s}$, and $z = 0.09$.

4.1 Static parameters

The PARVEX DC motor maximum control value is

$$u_{sat} = 3V. \quad (31)$$

The controller is designed so that 20% of the control signal may be used for the regulation function. The maximum value of the control signal available for the positioning function is thus:

$$u_{\max} = 0.8u_{sat}, \quad (32)$$

and the maximum desired is set to

$$e_{\max} = 1800 \text{ rad.s}^{-1}. \quad (33)$$

4.2 Dynamic optimization

The optimization according criteria (26) and constraint (29) and $(u_{\max} \cdot e_{\max})$ leads to:

$$n = 3.3 \quad \tau_1 = 1.21 \text{ s and } \tau_2 = 8.81 \cdot 10^{-2} \text{ s} \quad (34)$$

From expression , the prefilter is deduced using the Crone Identification module [19]:

$$F_{FBLFD} = \frac{F_0(s+b_1)(s+b_2)(s^2+b_3s+b_4)}{(s+a_1)(s+a_2)(s^2+a_3s+a_4)}, \quad (35)$$

where numerator and denominator coefficients are in Table 2.

Table 2. Numerator and denominator coefficients of the FBLFD prefilter

Numerator	Denominator
$F_0 = 1501.4026$	$a_1 = 10^8$
$b_1 = 11.35$	$a_2 = 1.12$
$b_2 = 25.1$	$a_3 = 1.12$
$b_3 = 16 \quad b_4 = 100$	$a_4 = 1.112 \quad a_5 = 0.3819$

To validate the synthesis methodology, a $e_{\max} = 1800 \text{ rad/s}$ is applied. A PID controller is designed with crossover frequency $\omega_u = 6.57 \text{ rad/s}$, corner frequency for the integral action $\omega_i = 1.77 \text{ rad/s}$, for high-frequency filter $\omega_f = 100 \text{ rad/s}$, and phase margin $\Phi_m = 45^\circ$. The following controller is obtained:

$$C(s) = 87.6204 \frac{(s+1.825)(s+1.774)}{s(s+100)(s+23.65)}. \quad (36)$$

A filtered noise is added on the feedback.

The simulation is also done for Davidson–Cole prefilter. The optimal parameters (n, τ) are:

$$n = 3.1 \quad \tau = 1.22. \quad (37)$$

The Davidson–Cole prefilter is deduced using the Crone Identification module [19]:

$$F_{DC} = \frac{F_0}{(s+a_1)(s^2+a_2s+a_3)}, \quad (38)$$

where numerator and denominator coefficients are in Table 3.

Table 3. Numerator and denominator coefficients of the FBLFD prefilter

Numerator	Denominator
$F_0 = 0.55614$	$a_1 = 1.55$
	$a_2 = 1.078 \quad a_4 = 0.3588$

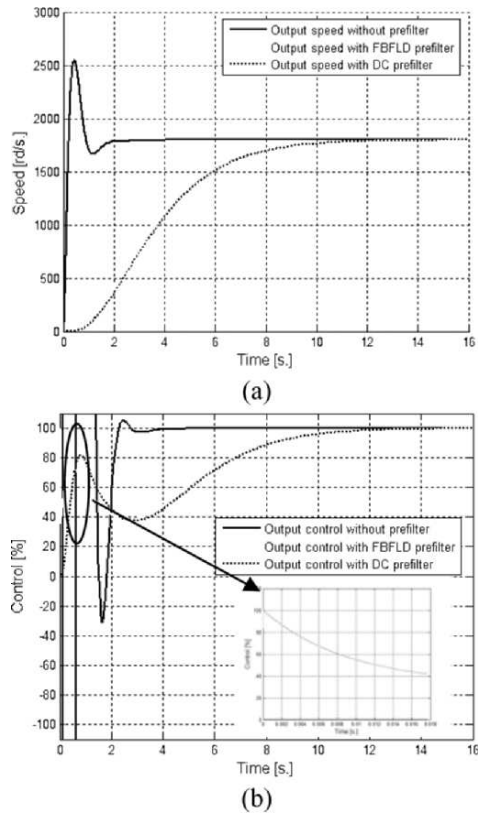


Fig. 5. Output speed (a) and control (b) for maximum speed ($V = 1800$ rad/s).

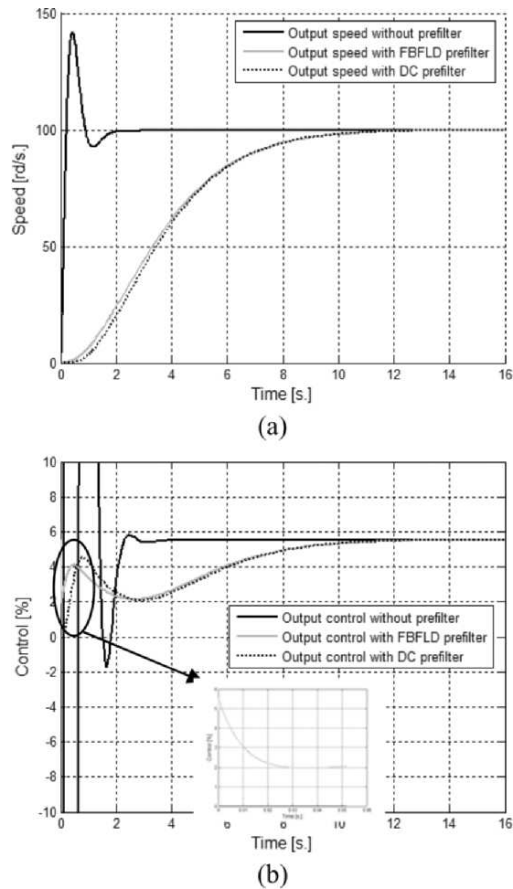


Fig. 6. Output speed (a) and control (b) for short speed ($V = 100$ rad/s).

Simulation results for a maximum speed ($V = 1,800$ rad/s) and small speed ($V = 100$ rad/s) are respectively given in Fig. 5 and 6.

The prefilter increase the settling time ($t_{90\%} \approx 6.86$ s) but the control signal stays below its maximum value. However, the maximum value of the control signal is reached for the short time and kept. Without a prefilter, $u_{\max}$ is not respected and is greater than the maximum admissible value.

5 Conclusion

In this paper, an extension of the method based on a Davidson–Cole prefilter is presented: the reference input results from the step response of a FBLED

prefilter whose main properties are having no overshoot on the plant and to have maximum control value for starting time. It permits the generation of optimal movement reference input leading to a minimum path completion time, taking into account the bandwidth of the closed-loop system.

It is synthesized in the frequency domain, thus the power spectral density of the position allows absolute control of the high-frequency energy. Thanks to the frequential constraint, the maximum control value is set at the initial instant without saturation. Moreover, the prefilter can be implemented as a classical digital filter.

A simulation on a Parvex DC RX 120 motor model validates the methodology.

This approach is complementary to CRONE control which allows a robust control law, and which is based on real or complex noninteger order differentiation.

References

1. Bobrow JE, Dubowsky S, Gibson JS (1985) Time-optimal control of robotic manipulators along specified paths, *Int. J. Robotics Res.*, 4(3):3–17.
2. Davidson D, Cole R (1951) *J. Chem. Phys.*, 19:1484–1490.
3. Dombre E, Khalil W (1988) *Modélisation et commande des robots*. Editions Hermès, Paris.
4. Horowitz I (1992) *Quantitative Feedback Design Theory (QFT)*. QFT, Colorado.
5. Khalil W, Dombre E (1999) *Modélisation, identification et commande des robots*, Editions Hermès, Paris.
6. Kieffer J, Cahill AJ, James MR (1997) Robust and accurate time-optimal path-tracking control for robot manipulators, *IEEE Trans. Robotics Automation*, 13(6):880–890.
7. Lin CS, Chang PR, Luh JYS (1983) Formulation and optimisation of cubic polynomial joint trajectories for industrial robots, *IEEE Trans. Automatic Control*, 28(12):1066–1073.
8. De Luca A, Lanari L, Oriolo G (1991) A sensitivity approach to optimal spline robot trajectories, *IEEE Trans. Automatic Control*, 27(3):535–539.
9. Le Mehaute A, Heliodore F, Oustaloup A (1991) Cole-Cole relaxation and CRONE relaxation, IMACS-IFAC symposium MCTS 91, Lille, France, May.
10. Melchior P, Lanusse P, Dancla F, Cois O (1999) Valorisation de l'approche non entière par le logiciel CRONE, CETSIS-EEA'99, Montpellier, France.
11. Melchior P, Orsoni B, Badie Th, Robin G (2000) Génération de consigne optimale par filtre à dérivée généralisée implicite: Application au véhicule électrique, IEEE CIFA'2000, Lille, France.

12. Melchior P, Orsoni B, Badie Th, Robin G, Oustaloup A (2000) Non-integer motion control: application to an XY cutting table, 1st IFAC Conference on Mechatronic Systems, Darmstadt, Germany, September.
13. Melchior P, Poty A, Oustaloup A (2005) Path tracking design by frequency band-limited fractional differentiator prefilter, Fifth EUROMECH Nonlinear Dynamics Conference (ENOC'05), Eindhoven, The Netherlands, August 7–12.
14. Miller KS, Ross B (1993) *An Introduction to the Fractional Calculus and Fractional Differential Equations*, Wiley, New York.
15. Nedler JA, Mead R (1965) A simplex method for function minimization, *Comput. J.*, 7:308–313.
16. Öhr J, Sternad M, Rönnbäck S (1998) H₂-optimal anti-windup performance in SISO control systems, 4th SIAM Conference on Control and its Applications, Jacksonville, USA, May.
17. Orsoni B, Melchior P, Oustaloup A (2001) Davidson-Cole transfer function in path tracking, 6th IEEE European Control Conference ECC'2001, Porto, Portugal, September 4–7, pp. 1174–1179.
18. Oustaloup A (1995) *La dérivation non entière: théorie, synthèse et applications*, Editions Hermès, Paris.
19. Oustaloup A, Melchior P, Lanusse P, Cois O, Dancla F (2000) The CRONE toolbox for Matlab, IEEE International Symposium on Computer-Aided Control-System Design, Anchorage, USA.
20. Samko SG, Kilbas AA, Marichev OI (1993) *Fractional Integrals and Derivatives*. Gordon and Breach, New York.
21. Shin KG, McKay ND (1985) Minimum-time control of robotic manipulators with geometric path constraints, *IEEE Trans. Automatic Control*, 30(6):531–541.
22. Shin KG, McKay ND (1987) Robust trajectory planning for robotic manipulators under payload uncertainties, *IEEE Trans. Automatic Control*, 32(12):1044–1054.
23. Singhose W, Singer N, Seering W (1995) Comparison of command shaping methods for reducing residual vibration, 3rd European Control Conference, Rome, Italy, September.

FLATNESS CONTROL OF A FRACTIONAL THERMAL SYSTEM

Pierre Melchior, Mikael Cugnet, Jocelyn Sabatier, Alexandre Poty, and Alain Oustaloup

LAPS - UMR 5131 CNRS, Université Bordeaux 1, ENSEIRB 351 cours de la Libération - F33405 TALENCE Cedex, France; Tél: +33 (0)5 40 00 66 07, Fax: +33 (0)5 40 00 66 44, E-mail: surname.name@laps.u-bordeaux1.fr, URL: <http://www.laps.u-bordeaux1.fr>

Abstract

This paper concerns the application of flatness principle to fractional systems. In path planning, the flatness concept is used when the trajectory is fixed (in space and in time), to determine the controls inputs to apply without having to integrate any differential equations. A lot of developments have been made but, in the case of non-integer differential systems (or fractional systems), few developments are still to be made. So, the aim of this paper is to apply flatness principle to a fractional system. As soon as the path has been obtained by flatness, a new robust path tracking based on CRONE control is presented. Firstly, flatness principle definitions used in control's theory are reminded. The fractional systems dynamic inversion is studied. A robust path tracking based on CRONE control is presented. Finally, simulations on a thermal testing bench model, with two different controllers (PID and CRONE), illustrate the path tracking robustness.

Keywords

Flatness, motion control, robust path tracking, robust control, Crone control, fractional systems, thermal systems.

1 Introduction

The systems control theory has been enriched recently with the discovery of a new property characterizing a certain class of non-linear systems which allows the achievement of a simple and robust control [3–5]. This property called flatness has been introduced in 1992 by M. Fliess, J. Lévine, Ph. Martin, and

P. Rouchon [1, 2, 15, 16], then applied to planes and cranes piloting control processing. Thus, each flat system has got a variable said flat output or linearizing output which summarize, on its own, all the system dynamics. That leads us to tell that all system variables deduct themselves from it without integrating differential equations.

Controlled dynamic system flatness notion has been studied abundantly in the case of finished state's dimension non-linear systems. Indeed, it permits the actual realization of path planning and linearization by feedback, with an easy implementation. Flatness offers, just as well, the possibility to solve, straightforwardly, the path tracking problem. Also, it is used in a lot of control problems. As well as its obvious practical interest, dynamic systems flatness in finished state's dimension is a very rich theoretical study domain. Despite some undoubted improvements, the flat systems characterization problem is still an open-ended problem left [17].

This study involves controllable fractional linear systems control. Nonlinear effects in the thermal system are not considered because its transfer is obtained by identification in a linear form. Therefore, it is not a matter of determining if flatness applies to these systems, since it is proved that any controllable linear system is flat, but demonstrating the corresponding flat output can be generalized under a well-known form in state space. The originality of our work is to perform the dynamic inversion of fractional systems by means of flatness concept, without having to integrate any differential equations, but in using the flat output in the case of a fractional transfer along a chosen path [11, 20]. The following sections, show how the Laplace formalism enables us from a desired output to design the control necessary for its achievement.

After this introduction, part 2 summarizes the flatness principle. In part 3, the dynamic inversion of fractional system is detailed. The thermal testing bench is presented in part 4. Part 5 presents the controller design, and part 6 the simulation results. Finally, a conclusion is given in part 7.

2 The Flatness Principle

2.1 Definition in theory control

Let the system be shown by the following differential equation:

$$\dot{x} = f(x, u), \quad (1)$$

$$\begin{aligned} x = (x_1, \dots, x_n) \in \mathbf{R}^n, u = (u_1, \dots, u_m) \in \mathbf{R}^m \\ \text{with } m \leq n \text{ and } f = (f_1, \dots, f_n) \end{aligned}, \quad (2)$$

a regular function of x and u whose the rank of $\frac{\partial f}{\partial u}$ is equal to m . This system is said differentially flat if there are m scalar functions depends on x , u and a β finite number of its derivatives:

$$z_j = h_j(x, u, \dot{u}, \dots, u^{(\beta)}), \quad (3)$$

in which $j = 1, \dots, m$, such as inverse of $\dot{x} = f(x, u)$ system, does not admit any dynamics, u being flat system input, and z being flat system output. Variables z are called flat output or linearizing output. This definition necessitates the existence of two functions A and B as:

$$x = A(z, \dot{z}, \dots, z^{(\alpha)}), \quad (4)$$

$$u = B(z, \dot{z}, \dots, z^{(\alpha+1)}), \quad (5)$$

where α is an integer. System output is also a function of the flat output:

$$y = C(z, \dot{z}, \dots, z^{(\sigma)}), \quad (6)$$

in which σ is an integer. So we can calculate system paths from the flat output z path definition without having to integrate any differential equations. We are able to conceive a linearizing feedback and a diffeomorphism (a C^1 class continue and bijective function) which transform the feedback system to an integral elements chain formed by z too. The linearizing feedback designed in this way will be called endogenous. A flat system is also linearizable by endogenous feedback and inversely. Therefore a flat system is a particular case of linearizable systems and a controllable and linear system is always flat: taking the Brunovsky's outputs stemming from controllability canonical forms as flat outputs is sufficient.

2.2 Continuous linear systems flatness

Let the single input–single output (SISO) time-invariant continuous linear system be defined by the following transfer function in [3]:

$$A(s) Y(s) = B(s) U(s), \quad (7)$$

in which $A(s)$ and $B(s)$ polynomials, both prime, are given by:

$$A(s) = s^n + \sum_{i=0}^{n-1} a_i s^i = s^n + A^*(s) \quad (8)$$

$$\text{and} \quad B(s) = \sum_{i=0}^{n-1} b_i s^i. \quad (9)$$

If the system is controllable, then, it is flat and the flat output is defined by:

$$Z(s) = N(s)Y(s) + D(s)U(s), \quad (10)$$

in which $N(s)$ and $D(s)$ satisfy the following Bézout's identity:

$$N(s)B(s) + D(s)A(s) = 1. \quad (11)$$

Theorem: If $z(t)$ is a flat output, then, we can write:

$$U(s) = A(s)Z(s), \quad (12)$$

$$Y(s) = B(s)Z(s). \quad (13)$$

Comment: $y(t)$ is a flat output if and only if $B(s)$ is a constant.

3 Dynamic Inversion of a Fractional System

3.1 Fractional order transfer model

Let the SISO time-invariant continuous fractional system be defined by the following transfer function:

$$G(s) = \frac{b_0 s^{\beta_0} + b_1 s^{\beta_1} + \dots + b_{m-1} s^{\beta_{m-1}} + b_m s^{\beta_m}}{a_0 s^{\alpha_0} + a_1 s^{\alpha_1} + \dots + a_{n-1} s^{\alpha_{n-1}} + a_n s^{\alpha_n}} \quad (14)$$

with $(\alpha_0, \dots, \alpha_n) \in \mathbb{C}^{n+1}$ and $(\beta_0, \dots, \beta_m) \in \mathbb{C}^{m+1}$

$$\alpha_0 > \alpha_1 > \dots > \alpha_{n-1} > \alpha_n, \quad \beta_0 > \beta_1 > \dots > \beta_{n-1} > \beta_n$$

and $\alpha_0 > \beta_0$.

The case of fractional systems [19, 20] poses a tricky problem in the calculus of controls necessary to obtain the arbitrary paths we have chosen. In fact, there is a result for this kind of uncommon calculus using the well-known Grünwald–Letnikov formula for the fractional derivative [13]:

$$D_h^\alpha f(t) = h^{-\alpha} \sum_{k=0}^{\left\lfloor \frac{t}{h} \right\rfloor} (-1)^k \binom{\alpha}{k} f(t - kh) \quad (15)$$

with $\binom{\alpha}{k} = \frac{\alpha!}{k!(\alpha-k)!} = \frac{\Gamma(\alpha+1)}{\Gamma(k+1)\Gamma(\alpha-k+1)}$ and $\left\lfloor \frac{t}{h} \right\rfloor$ the integer part of the ratio t/h .

Through the flat output $z(t-kh)$, which introduce the terms $z(t)$, $z(t-h)$, $z(t-2h)$... as samples of the past, the Grünwald–Letnikov formula shows the non-integer derivative of a function, at given time t , takes account of all previous function values. Integer derivative gives a local characterization of a function (graph tangent slope at time t), while non-integer derivative gives a global characterization as explained in [7, 10].

From the relation (15), the α order generalized derivative of $f(t)$ is:

$$D^\alpha f(t) = \lim_{h \rightarrow 0} D_h^\alpha f(t) . \quad (16)$$

The error made by the use of the Grünwald–Letnikov formula in our calculus algorithms is of the order of h :

$$D^\alpha f(t) - D_h^\alpha f(t) = O(h) . \quad (17)$$

3.2 Determination of the system flat output

From Eq. (13), as a rule, whatever way you choose the path, the flat output expression is:

$$Z(s) = \frac{Y(s)}{b_0 s^{\beta_0} + b_1 s^{\beta_1} + \dots + b_{m-1} s^{\beta_{m-1}} + b_m s^{\beta_m}} , \quad (18)$$

with $Y(s)$: the Laplace transform of the wanted output $y(t)$. Its temporal expression is deduced from the relation (18):

$$\begin{aligned} & b_0 D^{\beta_0} z(t) + b_1 D^{\beta_1} z(t) + \dots \\ & \dots + b_{m-1} D^{\beta_{m-1}} z(t) + b_m D^{\beta_m} z(t) = y(t) \end{aligned} . \quad (19)$$

$D^{\beta} z(t)$ is approximated by $D_h^{\beta} z(t)$ (error in h) but there are different methods to solve this kind of n order differential equation [14]. With the Grünwald–Letnikov approach, the discretization of the previous differential equation (19) leads to:

$$\begin{aligned}
& b_0 D_h^{\beta_0} z(kh) + b_1 D_h^{\beta_1} z(kh) + \dots \\
& \dots + b_{m-1} D_h^{\beta_{m-1}} z(kh) + b_m D_h^{\beta_m} z(kh) = y(kh)
\end{aligned} \tag{20}$$

which, under developed form, give the following expression:

$$\begin{aligned}
y(t) = & b_0 h^{-\beta_0} \sum_{k=0}^{\left\lfloor \frac{t}{h} \right\rfloor} (-1)^k \binom{\beta_0}{k} z(t - kh) \\
& + b_1 h^{-\beta_1} \sum_{k=0}^{\left\lfloor \frac{t}{h} \right\rfloor} (-1)^k \binom{\beta_1}{k} z(t - kh) \\
& + \dots \\
& + b_{m-1} h^{-\beta_{m-1}} \sum_{k=0}^{\left\lfloor \frac{t}{h} \right\rfloor} (-1)^k \binom{\beta_{m-1}}{k} z(t - kh) \\
& + b_m h^{-\beta_m} \sum_{k=0}^{\left\lfloor \frac{t}{h} \right\rfloor} (-1)^k \binom{\beta_m}{k} z(t - kh)
\end{aligned} \tag{21}$$

For $\frac{t}{h} = 0$, we have:

$$\begin{aligned}
& (b_0 h^{-\beta_0} + b_1 h^{-\beta_1} + \dots \\
& \dots + b_{m-1} h^{-\beta_{m-1}} + b_m h^{-\beta_m}) z(0) = y(0)
\end{aligned} \tag{22}$$

This allows to find the initial value of the flat output. A recursive process permits us to get the following values which depend, at each case, only on the output value set at time t and on the flat output values whole previously calculated.

3.3 Determination of the control

With the (13) and (22) relations, the SISO time-invariant continuous linear system control is written by [11]:

$$\begin{aligned}
u(t) = & a_0 h^{-\alpha_0} \sum_{k=0}^{\left\lfloor \frac{t}{h} \right\rfloor} (-1)^k \binom{\alpha_0}{k} z(t - kh) \\
& + a_1 h^{-\alpha_1} \sum_{k=0}^{\left\lfloor \frac{t}{h} \right\rfloor} (-1)^k \binom{\alpha_1}{k} z(t - kh) \\
& + \dots \\
& + a_{n-1} h^{-\alpha_{n-1}} \sum_{k=0}^{\left\lfloor \frac{t}{h} \right\rfloor} (-1)^k \binom{\alpha_{n-1}}{k} z(t - kh) \\
& + a_n h^{-\alpha_n} \sum_{k=0}^{\left\lfloor \frac{t}{h} \right\rfloor} (-1)^k \binom{\alpha_n}{k} z(t - kh)
\end{aligned} \tag{23}$$

with $(\alpha_0, \dots, \alpha_n) \in \mathbb{C}^{n+1}$.

Through this result, we can observe that each control sample is obtained according to the corresponding flat output sample as well as all those that precede it. In the same way, the control has been introduced, in the second part, as a function of the flat output and its successive derivatives. Also, we notice this assertion is always true in the discrete case and finally the flatness principle applies to fractional systems.

4 Thermal Testing Bench

4.1 Description of the thermal testing bench

The testing bench copying the behaviour of a non-integer derivatives system involves a semi-infinite-dimensional thermal system, namely, an aluminium rod of large dimension (40 cm) (Fig. 1):

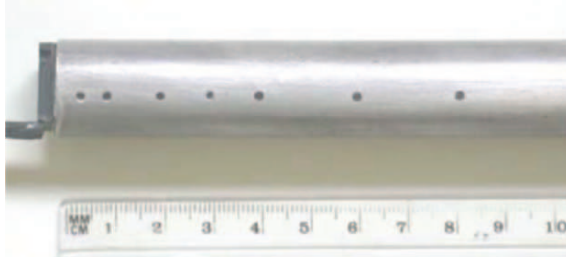


Fig. 1. Aluminium bar, heating resistor 0–12 W and measurement slot.

As illustrated in Fig. 2, the input of this system is a thermal flux and its output is the rod temperature gauged at a d distance from the heated surface. In order to maintain an unidirectional heat transfer, the entire rod surface is isolated (Fig. 3). The thermal flux is generated by a heating resistor stuck on a rod end (high thermal conductivity glue). The maximal flux which can be generated by the rod is 12 W (1 A under 12 V).

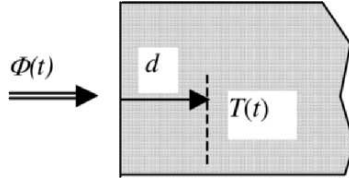


Fig. 2. Thermal system principle.

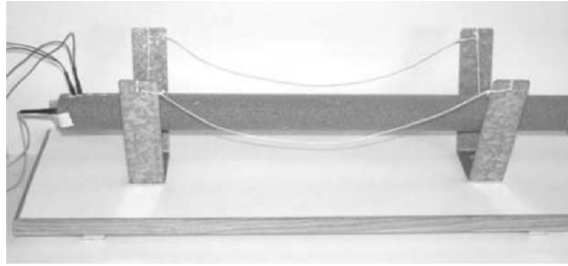


Fig. 3. Photography of the isolated thermal system equipped with the heat resistor and the temperature probes.

4.2 Requirements

This testing bench is composed of an aluminium rod entirely isolated and heated at one end. The length of the aluminium rod allows to look upon it as a semi-infinite media and to demonstrate the existence of a non-integer transfer linking the thermal flux applied at one of its ends to the temperature inside. The non-integer physical behaviour proof of this system is not exposed here, due to the fact it is already developed in [9]. In a first approach, a linear model is obtained by identification. The testing bench is characterized by the following transfer:

$$H(s) = \frac{T(s)}{\varphi(s)}, \quad (24)$$

$$\text{thus} \quad H(s) = \frac{-0.11716s + 0.094626s^{0.5} + 0.0052955}{s^{1.5} + 0.42833s + 0.060125s^{0.5}}. \quad (25)$$

One can notice the positive real part zero, which introduces strong constraints on the performances.

The actuator stresses are studied for the adjustment of an enabled path respecting the maximal temperature like the first, second, and third derivatives maxima. In our case, we are going to consider that the first and second derivatives of temperature are equal to zero at initial and final temperatures. The aluminium rod temperature will have to rise 30°C above ambient temperature in 1,250 s, according to the chosen path, a polynomial interpolation of degree 5 (PI5) [6, 8]:

$$y(t) = q_i + 80(q_f - q_i)\frac{t^3}{t_f^3} - 240(q_f - q_i)\frac{t^4}{t_f^4} + 192(q_f - q_i)\frac{t^5}{t_f^5} \quad (26)$$

with: $q_i = 0^\circ$, $q_f = 30^\circ$, $t_f = 2,500$ s.

The maximum value of the control $u(t)$ is fixed to 10 V.

4.3 Control algorithm test in open-loop

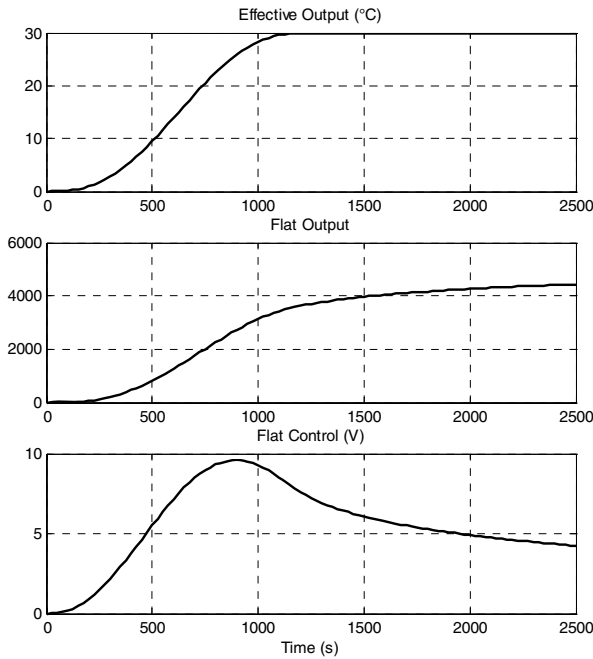


Fig. 4. Simulation of the testing bench in open-loop.

An algorithm able to generate the flat output and the control, necessary to obtain a chosen path, has been created for fractional systems applications. The control calculated in this way, is used in simulation to provide an off-line computed control for the thermal testing bench Simulink model. Of course, all the simulations are carried out with the software Matlab. The open-loop control scheme is presented in Fig. 5. The results obtained are given Fig. 4. Due to the difficulty to give to the flat output a concrete meaning, no unit is employed to define it.

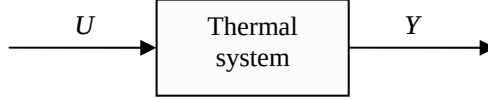


Fig. 5. Open-loop control scheme.

5 Controllers Design

5.1 The PID controller design

The proportional, integral, differential (PID) controller is designed for a desired open-loop gain crossover frequency ω_{cg} equal to 0.01 rad/s. It seems to be not the best choice because it makes the controller a bit slow but it is the only choice which allows to be not too sensible to the positive real part zero contained in the thermal testing bench transfer. A phase margin of 60° is chosen in order to reduce the overshoot. All these specification sheets lead to the PID controller described by the following transfer function explaining the proportional, integral, differential, and Filtering (to reduce noise in high frequencies) action parts:

$$C(s) = C_0 \cdot \left(\frac{1 + \frac{s}{\omega_i}}{\frac{s}{\omega_i}} \right) \cdot \left(\frac{1 + \frac{s}{\omega_a}}{1 + \frac{s}{\omega_b}} \right) \cdot \left(\frac{1}{1 + \frac{s}{\omega_f}} \right) \quad (27)$$

with $C_0 = 3.27$, $\omega_i = 0.001$, $\omega_a = 0.0437$, $\omega_b = 0.00229$, $\omega_f = 0.1$.

5.2 The CRONE controller design

The CRONE controller (a French acronym which means: fractional-order robust control [7, 12]) is defined within the frequency range [0.001, 0.1]

around the desired open-loop gain crossover frequency ω_{cg} in order to ensure a constant phase and more particularly to ensure small variations of the closed-loop system stability-degree. Its transfer function is the following one:

$$C(s) = K \frac{\left(1 + \frac{s}{\omega_b}\right)^{n_b} \left(1 + \frac{s}{\omega_h}\right)^n \left(-s^{0.5} + z\right) \left(s^{1.5} + 0.42833s + 0.060125s^{0.5}\right)}{\left(\frac{s}{\omega_b}\right)^{n_b} \left(1 + \frac{s}{\omega_b}\right)^n \left(1 + \frac{s}{5\omega_h}\right)^{n_h} \left(-0.11716s + 0.094626s^{0.5} + 0.0052955\right)} \quad (28)$$

with $K=460$, $\omega_b=10^{-4}$ rad/s, $\omega_h=1$ rad/s, $z=0.86$, $n_b=1.5$, $n_h=2$ et $n=1.3$.

The desired open-loop gain crossover frequency ω_{cg} is the same with the both controllers, to obtain the same rapidity.

The problem is now to implement this controller on the Simulink model. The solution is contained in the CRONE toolbox (Fig. 6), the tool is named named computer aided frequency identification. With this tool, an achievable achievable rational version $C_R(s)$ of the controller, which can be implemented defined by a transfer function resulting from a recursive distribution of cells of real negative zeros and poles.

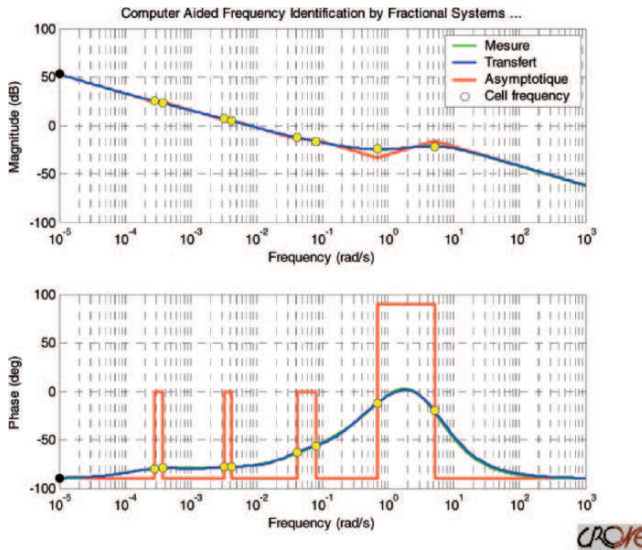


Fig. 6. The CRONE toolbox user interface.

The rational transfer function calculated in this way is:

$$C_R(s) = \frac{2.456 \times 10^5 s^5 + 4.843 \times 10^5 s^4 + 1.38 \times 10^5 s^3 + 5403 s^2 + 0.004416}{2.995 \times 10^5 s^6 + 3.11 \times 10^6 s^5 + 8.202 \times 10^6 s^4 + 6.677 \times 10^5 s^3 + 2865 s^2 + s}. \quad (29)$$

5.3 Comparison of both controllers

A comparison of both controllers is given by the Fig. 7 in which Bode diagrams in open-loop are presented to see the frequency characteristics of each one of them. The constant phase around ω_{cg} is visible in the Crone controller case.

A comparison of both controllers is given by the Fig. 8 in which Bode diagrams in open-loop are presented. Some Bode diagrams for different gain variations (1/50, 1, 50, and 80 times as much gain) are also presented in order to prove the interest to have a quasi-constant phase around ω_{cg} in the Crone controller case.

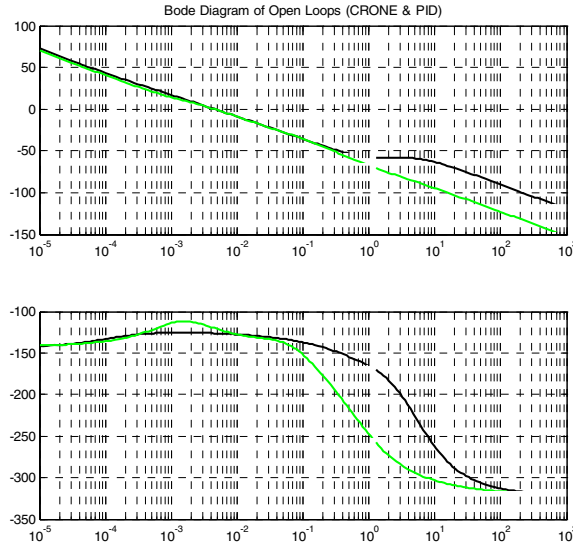


Fig. 7. Comparison of both PID (grey) and CRONE (black) open-loop Bode diagrams.

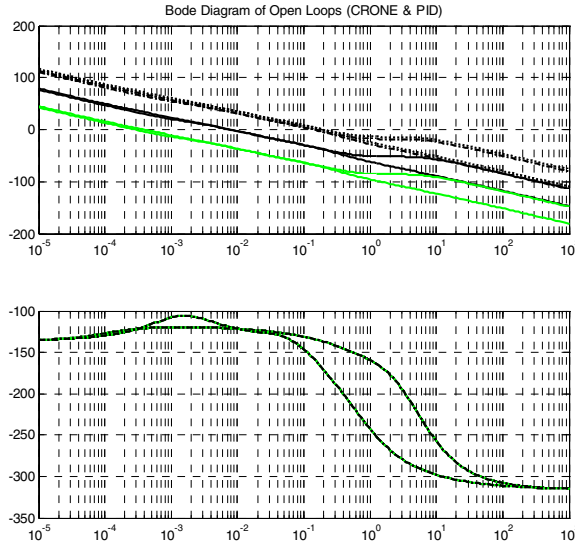


Fig. 8. Comparison of both PID and CRONE open-loop Bode diagrams with different gain variations: $G_0/50$ (grey), G_0 (solid), $G_0 \times 50$ (dash dotted), and $G_0 \times 50$ (dotted).

6 Simulation Results

Now, the system is studied in closed-loop so as to measure its immunity to different disturbances applied to its input (ΔU) and its output (ΔY). The control scheme is presented by Fig. 9, with, U_{ref} , the control obtained by the flatness principle using the chosen reference trajectory Y_{ref} .

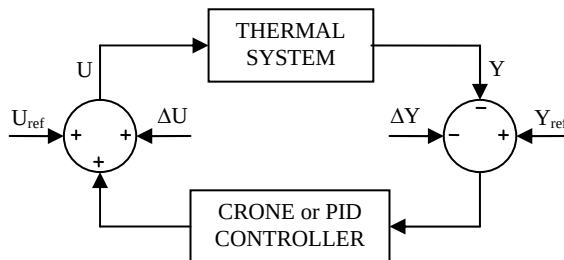


Fig. 9. Closed-loop control scheme.

The PID controller and the CRONE controller are both used in simulation. For this, we study the disturbances and gain variation influences on path tracking. For this, a 1° control input disturbance is applied at 500 s and a 3°

output disturbance is applied at 1,500 s. Time responses are given for different gain variations (1, 50, and 80 times as much gain).

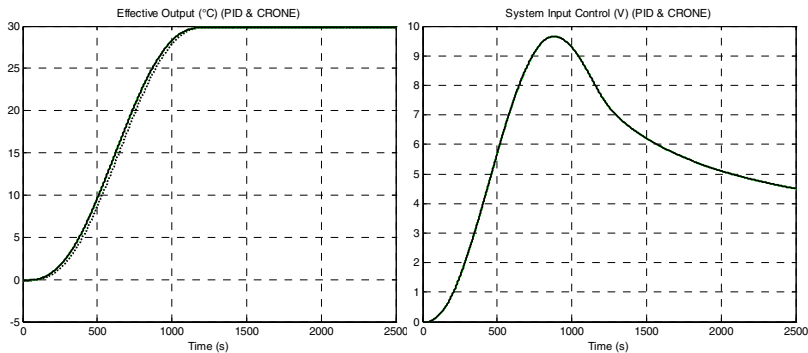


Fig. 10. Simulation with no disturbance and no gain variation; path (*dotted*), CRONE (*black*) and PID (*grey*).

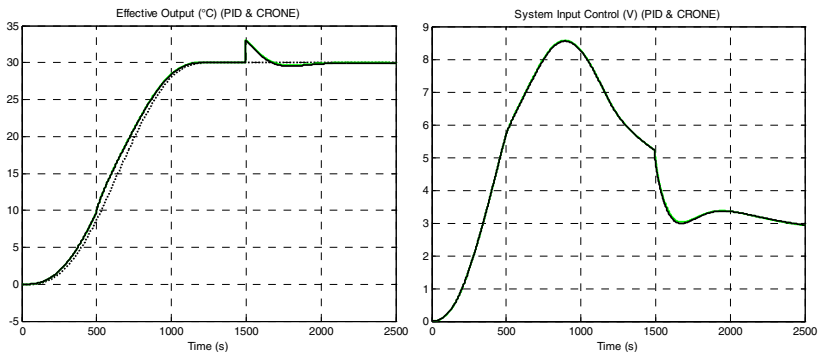


Fig. 11. Simulation with disturbances and no gain variation; path (*dotted*), CRONE (*black*), and PID (*grey*).

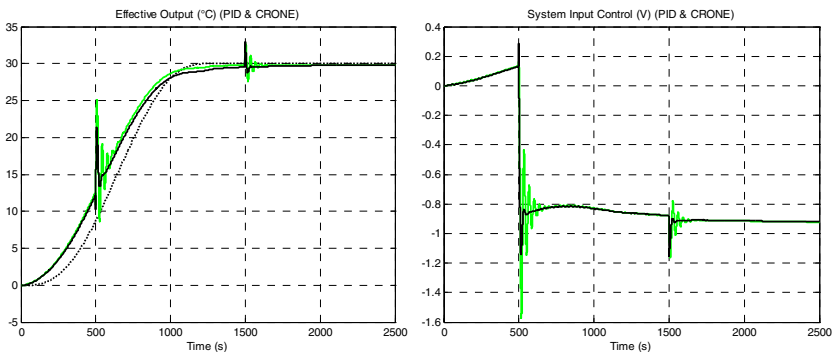


Fig. 12. Simulation with disturbances and $G_0 \times 50$ gain variation; path (*dotted*), CRONE (*black*), and PID (*grey*).

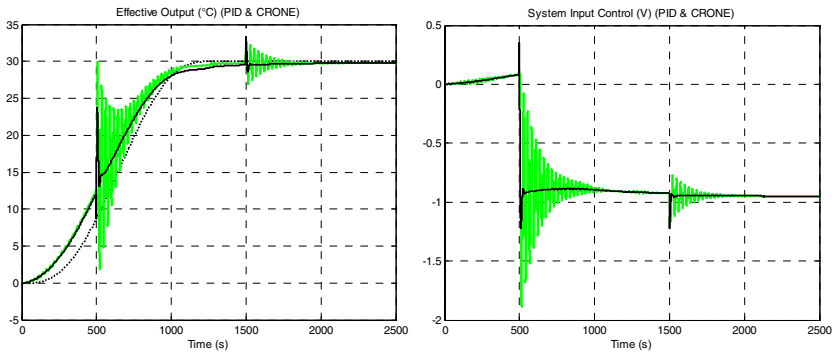


Fig. 13. Simulation with disturbances and $G0 \times 80$ gain variation; path (*dotted*), CRONE (*black*) and PID (*grey*).

Figure 10 shows the same path tracking for PID and CRONE controllers. In fact, the loop has no role in the nominal case.

Figure 11 shows a good path tracking in presence of disturbances due to the loop. PID and CRONE have the same dynamic behaviour (same ωcg).

The robustness study is presented by Figs. 12 and 13. We can see clearly a better path tracking performance with the CRONE controller as well as beside the disturbances than gain variations, due to a quasi-constant phase around ωcg in the Crone controller case.

7 Conclusion

In this paper, a new robust path tracking design based on flatness and CRONE control approaches was presented. Therefore, this method was applied to a fractional system: a thermal testing bench. Firstly, flatness principle definitions used in control's theory were reminded. Then, the fractional systems dynamic inversion was studied. Simulations with two different controllers (PID and CRONE) illustrated the robustness of the proposed path tracking strategy. The study of robust path tracking via a third-generation CRONE control can also be integrated in future designs. Flatness principle application through non-linear fractional systems dynamic inversion can be conceivable.

Acknowledgment

This paper is a modified version of a paper published in proceedings of IDETC/CIE 2005, September 24–28, 2005, Long Beach, California, USA. The authors would like to thank the American Society of Mechanical

Engineers (ASME) for allowing them to publish this revised contribution of an ASME article in this book.

References

1. Fliess M, Lévine J, Martin Ph, Rouchon P (1992) Sur les systèmes non linéaires différentiellement plats, *C. R. Acad. Sci. Paris*, I-315:619–624.
2. Fliess M, Lévine J, Martin Ph, Rouchon P (1995) Flatness and defect of nonlinear systems: introductory theory and examples, *Int. J. Control*, 61(6): 1327–1361.
3. Ayadi M (2002) Contributions à la commande des systèmes linéaires plats de dimension finie, PhD thesis, Institut National Polytechnique de Toulouse.
4. Cazaurang F (1997) Commande robuste des systèmes plats, application à la commande d'une machine synchrone, PhD thesis, Université Bordeaux 1, Paris.
5. Lavigne L (2003) Outils d'analyse et de synthèse des lois de commande robuste des systèmes dynamiques plats, PhD thesis, Université Bordeaux 1, Paris.
6. Khalil W, Dombre E (1999) Modélisation, identification et commande des robots, 2ème édition, Editions Hermès, Paris.
7. Oustaloup A (1995) La dérivation non entière: théorie, synthèse et applications, *Traité des Nouvelles Technologies, série automatique*, Editions Hermès, Paris.
8. Orsoni B (2002) Dérivée généralisée en planification de trajectoire et génération de mouvement, PhD thesis, Université Bordeaux, Paris.
9. Sabatier J, Melchior P, Oustaloup A (2005) A testing bench for fractional system education, Fifth EUROMECH Nonlinear Dynamics Conference (ENOC'05), Eindhoven, The Netherlands, August 7–12.
10. Cois O (2002) Systèmes linéaires non entiers et identification par modèle non entier: application en thermique, PhD thesis, Université Bordeaux 1, Paris.
11. Cugnet M, Melchior P, Sabatier J, Poty A, Oustaloup A. (2005) Flatness principle applied to the dynamic inversion of fractional systems, Third IEEE SSD'05, Sousse, Tunisia, March 21–24.
12. Oustaloup A, Sabatier J, Lanusse P (1999) From fractal robustness to the CRONE control, *Fract. Calcul. Appl. Anal. (FCAA): Int. J. Theory Appl.*, 2(1):1–30, January.
13. Miller KS, Ross B (1993) *An Introduction to the Fractional Calculus and Fractional Differential Equations*. Wiley, New York.
14. Podlubny I (1999) *Fractional Differential Equations*. Academic Press, San Diego.
15. Lévine J, Nguyen DV (2003) Flat output characterization for linear systems using polynomial matrices, *Syst. Controls Lett.*, 48:69–75.
16. Bitauld L, Fliess M, Lévine J (1997) A flatness based control synthesis of linear systems and applications to windshield wipers, In *Proceedings ECC'97*, Brussels, July.

17. Lévine J (2004) On necessary and sufficient conditions for differential flatness, In Proceedings of the IFAC NOLCOS 2004 Conference.
18. Samko SG, Kilbas AA, Marichev OI (1987) Integrals and Derivatives of the Fractional Order and Some of Their Applications, Nauka I Tekhnika, Minsk, Russia.
19. Oldham KB, Spanier J (1974) *The Fractional Calculus*. Academic Press, New York, London.
20. Melchior P, Cugnet M, Sabatier J, Oustaloup A (2005) Flatness control: application to a fractional thermal system, ASME, IDETC/CIE 2005, September 24–28, Long Beach, California.

ROBUSTNESS COMPARISON OF SMITH PREDICTOR-BASED CONTROL AND FRACTIONAL-ORDER CONTROL

Patrick Lanusse and Alain Oustaloup

LAPS, UMR 5131 CNRS, Bordeaux I University, ENSEIRB, 351 cours de la Libération, 33405 Talence Cedex, Tel: +33 (0)5 4000 2417, Fax: +33 (0)5 4000 66 44, E-mail: {lanusse, oustaloup}@laps.u-bordeaux1.fr

Abstract

Many modifications have been proposed to improve the Smith predictor structure used to control plant with time-delay. Some of them have been proposed to enhance the robustness of Smith predictor-based controllers. They are often based on the use of deliberately mismatched model of the plant and then the internal model control (IMC) method can be used to tune the controller. This paper compares the performance of two Smith predictor-based controllers including a mismatched model to the performance provided by a fractional-order CRONE controller which is well known for managing well the robustness and performance tradeoff. It is shown that even if it can simplify the design of (robust) controller, the use of an improved Smith predictor is not necessary to obtain good performance.

Keywords

Time-delay system, fractional-order controller, robust control, Smith predictor, IMC method.

1 Introduction

In the context of the closed-loop control of time-delay systems, Smith [1] proposed a control scheme that leads to amazing performance which is impossible to obtain using common controller. It is now well known that such performance can be obtained for perfectly modeled systems only. When a system is uncertain or perturbed, trying to obtain high-performance, for instance settling times close to or lower than the time-delay value of the system, leads to lightly damped closed-loop system and sometimes to unstable

system. Then, many authors proposed to improve the design method of Smith predictors and provided what it is often presented as robust Smith predictors [2]. Wang et al. [3] proposed a design method based on a mismatched model of the time-delay system to be controlled. Using such a model constrains to design the controller very carefully but does not really provide a meaningful degree of freedom to manage the robustness problem. Thus, Zhang and Xu [4] proposed to use the internal model control (IMC) [5] and one degree of freedom to tune the performance and the robustness of the controller. Even if one degree of freedom leads to a low order, and interesting controller, it can be thought that the performance obtained could be improved by using more degree of freedom.

Fractional-order control-system design provides such further degree of freedom [6–10]. For instance, CRONE (acronym for *Commande Robuste d'Ordre Non Entier* which means non-integer order robust control) control-system design [11–18] uses the integration fractional order which permits the use of few high-level degrees of freedom. CRONE is a frequency-domain design approach for the robust control of uncertain (or perturbed) plants. The plant uncertainties (or perturbations) are taken into account without distinction of their nature, whether they are structured or unstructured. Using frequency uncertainty domains, as in the quantitative feedback theory (QFT) approach [19] where they are called *template*, the uncertainties are taken into account in a fully structured form without overestimation, thus leading to efficient controller because as little conservative as possible [20].

Section 2 presents the classical Smith predictor design and the approaches proposed by Wang and then by Zhang.

Section 3 presents the CRONE approach and particularly its third generation.

Section 4 proposes to make uncertain a time-delay system proposed by Wang, and then to compare the robustness and performance of Wang, Zhang and CRONE controllers.

2 Smith Predictor-Based Control-Systems

The structure of the classical Smith predictor (Fig. 1) includes the nominal model G_0 of the time-delay system G and the time-delay free model P_0 .

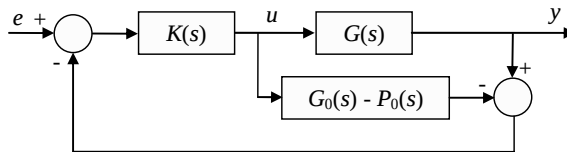


Fig. 1. Smith predictor structure.

The closed-loop transfer function y/e is

$$\frac{Y(s)}{E(s)} = \frac{K(s)G(s)}{1 + K(s)[P_0(s) + G(s) - G_0(s)]} \quad (1)$$

If G_0 models the plant G perfectly, the closed-loop stability depends on the controller K and on the delay-free model P_0 only, and any closed-loop dynamic can be obtained. As it is impossible that G_0 can model G perfectly, it has been shown that the roll-off of transfer function (1) needs to be sufficient to avoid instability. Then, it is not really important to choose a high-order an accurate model G_0 for the control of an uncertain plant G .

Wang et al. propose to replace G_0 by a deliberately mismatched model G_m , and P_0 by the first order part G_{m1} of G_m . Wang proposes a second-order system with a delay for the mismatched model

$$G_m(s) = \frac{ke^{-ts}}{(1 + \tau s)^2}, \quad (2)$$

and uses

$$G_{m1}(s) = \frac{k}{1 + \tau s}, \quad (3)$$

to design a low-order PID controller K .

Using the relation between the IMC method and the Smith predictor structure, Zhang and Xu propose an analytical way to design controller K . Figure 2 presents the Smith predictor including the IMC controller Q .

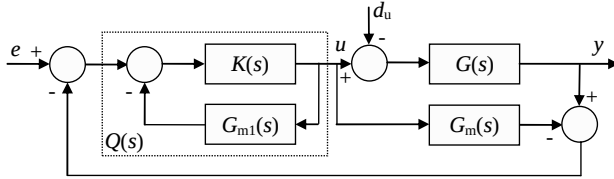


Fig. 2. Smith predictor with IMC controller.

The IMC controller Q equals:

$$Q(s) = \frac{K(s)}{1 + G_{m1}(s)K(s)} \quad (4)$$

If G_m approximates well the nominal plant G_0 , the nominal closed-loop transfer function y/e is close to the open-loop transfer function defined by:

$$J(s) = Q(s)G_m(s) \quad (5)$$

Zhang proposes to choose the user-defined transfer function J as

$$J(s) = \frac{e^{-ts}}{(1 + \lambda s)^2}, \quad (6)$$

with λ the time constant that can be tuned to achieved performance and robustness. Then, controller K is a PID controller given by:

$$K(s) = \frac{(1 + \tau s)^2}{(1 + \lambda s)^2 - (1 + \tau s)} \quad (7)$$

3 CRONE CSD Principles

3.1 Introduction

The CRONE control-system design (CSD) is based on the common unity-feedback configuration (Fig. 3). The controller or the open-loop transfer function is defined using integro-differentiation with non-integer (or fractional) order. The required robustness is that of both stability margins and performance, and particularly the robustness of the peak value M_r (called resonant peak) of the common complementary sensitivity function $T(s)$.

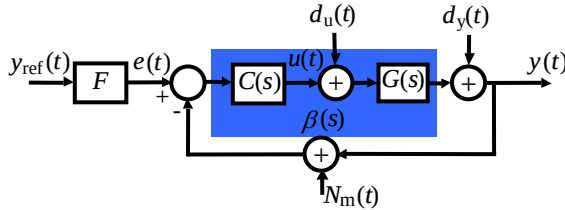


Fig. 3. Common CRONE control diagram.

Three CRONE control design methods have been developed, successively extending the application field. If CRONE design is only devoted to the closed-loop using the controller as one degree of freedom (DOF), it is obvious that a Second DOF (F , linear or not) could be added outside the loop for managing second tracking problems.

The variations of the phase margin (of a closed-loop system) come both from the parametric variations of the plant and from the controller phase variations around the frequency ω_{cg} , which can also vary. The first generation CRONE control proposes to use a controller without phase variation (fractional differentiation) around open loop gain crossover frequency ω_{cg} . Thus, the phase margin variation only results from the plant variation. This strategy has to be used when frequency ω_{cg} is within a frequency range where the plant phase is constant. In this range the plant variations are only gain like. Such a

range is often in the high frequencies, and can lead to high-level control input. So the second generation must be favored.

When the plant variations are gain like around frequency ω_{cg} , the plant phase variation (with respect to the frequency) is cancelled by those of the controller. Then there is no phase margin variation when frequency ω_{cg} varies. Such a controller produces a constant open loop phase (real fractional-order integration) whose Nichols locus is a vertical straight line named frequency template. This template ensures the robustness of phase and modulus margins and of resonant peaks of complementary sensitivity and sensitivity functions.

The third CRONE control generation must be used when the plant frequency uncertainty domains are of various types (not only gain like). The vertical template is then replaced by a generalized template always described as a straight line in the Nichols chart but of any direction (complex fractional order integration), or by a multi-template (or curvilinear template) defined by a set of generalized templates.

An optimization allows the determination of the independent parameters of the open loop transfer function. This optimization is based on the minimization of the stability degree variations, while respecting other specifications taken into account by constraints on sensitivity function magnitude. The complex fractional order permits parameterization of the open-loop transfer function with a small number of high-level parameters. The optimization of the control is thus reduced to only the search for the optimal values of these parameters. As the form of uncertainties taken into account is structured, this optimization is necessarily nonlinear. It is thus very important to limit the number of parameters to be optimized. After this optimization, the corresponding CRONE controller is synthesized as a rational fraction only for the optimal open-loop transfer function.

The third generation CRONE system design methodology, the most powerful one, is able to design controllers for plants with positive real part zeros or poles, time delay, and/or with lightly damped mode (Oustaloup et al. 1995). Associated with the w-bilinear variable change, it also permits the design of digital controllers. The CRONE control has also been extended to linear time variant systems and nonlinear systems whose nonlinear behaviors are taken into account by sets of linear equivalent behaviors [21]. For multi-input multi-output (MIMO) (multivariable) plants, two methods have been development [22]. The choice of the method is done through an analysis of the coupling rate of the plant. When this rate is reasonable, one can opt for the simplicity of the multi single-input single-output (SISO) approach.

3.2 Third generation CRONE methodology

Within a frequency range $[\omega_A, \omega_B]$ around open-loop gain-crossover frequency ν_{cg} , the Nichols locus of a third generation CRONE open-loop is defined by a any-angle straight line-segment, called a generalized template (Fig. 4).

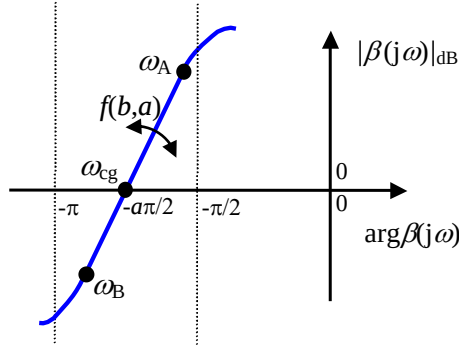


Fig. 4. Generalized template in the Nichols plane.

The generalized template can be defined by an integrator of complex fractional order n whose real part determines its phase location at frequency ω_{cg} , that is $-\text{Re}_i(n)\pi/2$, and whose imaginary part then determines its angle to the vertical (Fig. 5).

The transfer function including complex fractional-order integration is:

$$\beta(s) = \left(\cosh\left(b\frac{\pi}{2}\right) \right)^{\text{sign}(b)} \left(\frac{\omega_{cg}}{s} \right)^a \left(\text{Re}_i \left(\left(\frac{\omega_{cg}}{s} \right)^{ib} \right) \right)^{-\text{sign}(b)} \quad (8)$$

with $n = a + ib \in \Re_i$ and $w \in \Re_j$, and where $\Re_i$ and $\Re_j$ are respectively time-domain and frequency-domain complex planes.

The definition of the open-loop transfer function including the nominal plant must take into account:

- The accuracy specifications at low frequencies
- The generalized template around frequency ω_{cg}
- The plant behavior at high frequencies while respecting the control effort specifications at these frequencies

Thus, the open-loop transfer function is defined by a transfer function using band-limited complex fractional-order integration:

$$\beta(s) = \beta_l(s) \beta_m(s) \beta_h(s), \quad (9)$$

where $\beta(s)$ is a set of band-limited generalized templates :

$$\beta_m(s) = \prod_{k=-N^-}^{N^+} \beta_k(s), \quad (10)$$

with:

$$\beta_k(s) = C_k^{\text{sign}(b_k)} \left(\alpha_k \frac{1+s/\omega_{k+1}}{1+s/\omega_k} \right)^{a_k} \left(\Re e_{/i} \left\{ \left(\alpha_k \frac{1+s/\omega_{k+1}}{1+s/\omega_k} \right)^{ib_k} \right\} \right)^{-q_k \text{sign}(b_k)}$$

$$\alpha_k = (\omega_{k+1}/\omega_k)^{1/2} \quad \text{for } k \neq 0 \quad \text{and} \quad \alpha_0 = \left(1 + \left(\frac{\omega_r}{\omega_0} \right)^2 \right)^{1/2} / \left(1 + \left(\frac{\omega_r}{\omega_1} \right)^2 \right)^{1/2} \quad (11)$$

where $\beta_l(s)$ is an integer order n_l proportional integrator:

$$\beta_l(s) = C_l \left(\frac{\omega_{-N^-}}{s} + 1 \right)^{n_l} \quad (12)$$

where $\beta_h(s)$ is a low-pass filter of integer order n_h :

$$\beta_h(s) = \frac{C_h}{\left(\frac{s}{\omega_{N^+}} + 1 \right)^{n_h}} \quad (13)$$

The optimal open-loop transfer function is obtained by the minimization of the robustness cost function

$$J = \sup_{\omega, G} |T(j\omega)| - M_{r0}, \quad (14)$$

where M_{r0} is the resonant peak set for the nominal parametric state of the plant, while respecting the following set of inequality constraints for all plants (or parametric states of the plant) and for $\omega \in \nabla^+$:

$$\inf_G |T(j\omega)| \geq T_l(\omega) \quad \text{and} \quad \sup_G |T(j\omega)| \leq T_u(\omega), \quad (15)$$

$$\sup_G |S(j\omega)| \leq S_u(\omega), \quad \sup_G |CS(j\omega)| \leq CS_u(\omega) \quad \text{and} \quad \sup_G |GS(j\omega)| \leq GS_u(\omega), \quad (16)$$

with

$$\begin{cases} T(s) = \frac{C(s)G(s)}{1+C(s)G(s)} & S(s) = \frac{1}{1+C(s)G(s)} \\ CS(s) = \frac{C(s)}{1+C(s)G(s)} & GS(s) = \frac{G(s)}{1+C(s)G(s)} \end{cases} \quad (17)$$

As the uncertainties are taken into account by the least conservative method, a nonlinear optimization method must be used to find the optimal

values of the four independent parameters. The parameterization of the open-loop transfer function by complex fractional order of integration, then simplifies the optimization considerably. During optimization the complex order has, alone, the same function as many parameters found in common rational controllers.

When the optimal nominal open-loop transfer is determined, the fractional controller $C_F(s)$ is defined by its frequency response:

$$C_F(j\omega) = \frac{\beta(j\omega)}{G_0(j\omega)}, \quad (18)$$

where $G_0(j\omega)$ is the nominal frequency response of the plant.

The synthesis of the rational controller $C_R(s)$, consists in identifying ideal frequency response $C_F(j\omega)$ by that of a low-order transfer function. The parameters of a transfer function with a predefined structure are adapted to frequency response $C_F(j\omega)$. The rational integer model on which the parametric estimation is based, is given by:

$$C_R(s) = \frac{B(s)}{A(s)}, \quad (19)$$

where $B(s)$ and $A(s)$ are polynomials of specified integer degrees n_B and n_A . All the frequency-domain system-identification techniques can be used. An advantage of this design method is that whatever the complexity of the control problem, it is easy to find satisfactory values of n_B and n_A generally about 6 without performance reduction.

3.3 CRONE control of nonminimum-phase and time-delay plants

Let G be a plant whose nominal transfer function is:

$$G_0(s) = G_{mp}(s) e^{-\tau s} \prod_{i=1}^{n_z} \left(1 - \frac{s}{z_i} \right), \quad (20)$$

where: $G_{mp}(s)$ is its minimum-phase part; z_i is one of the its n_z right half-plane zeros; τ is a time-delay.

If $\beta(s)$ remains defined by (9), the use of (18) leads to an unstable controller (whose right half-plane poles are the n_z right half-plane zeros of the nominal plant) with a predictive part $e^{+\tau s}$. Taking into account, internal stability for the nominal plant, stability for the perturbed plants and achievability of the controller, it is obvious that such a controller cannot be used. Thus, the definition of $\beta(s)$ needs to be modified by including the nominal right half-plane zeros and the nominal time-delay:

$$\beta(s) = \beta_1(s)\beta_m(s)\beta_h(s)C_z e^{-\tau s} \prod_{i=1}^{n_z} \left(1 - \frac{s}{z_i}\right), \quad (21)$$

where C_z ensures the unitary magnitude of $\beta(s)$ at frequency ω_{cg} .

As frequency ω_{cg} must be smaller than the smallest modulus of the right half-plane zeros [23–25], and in a range where the effect of the time-delay is weak, the modification of $\beta(s)$ does not reduce the efficiency around ω_{cg} of the optimizing parameters during the constrained minimization.

5 Illustrative Example

A nonminimum and time-delay plant defined in [3] is used to compare the performance of Wang and Zhang controllers (both based on the Smith predictor structure), and CRONE controller. To assess the robustness of the controllers, a $\pm 20\%$ uncertainty is associated with each plant parameters. Then, the uncertain plant is defined by

$$G(s) = \frac{g(zs+1)}{(ps+1)^5} e^{-ds}, \quad (22)$$

with: $g \in [0.8, 1.2]$, $z \in [-1.2, -0.8]$, $p \in [0.8, 1.2]$ and $d \in [1.6, 2.4]$. Its nominal value given by Wang is defined by $g = -z = p = 1$ and $d = 2$.

To approximate the nominal plant, Wang proposes the mismatched model

$$G_m(s) = \frac{1e^{-5.07s}}{(0.999 + 1.46s)^2}, \quad (23)$$

and then the first-order transfer function

$$G_{m1}(s) = \frac{1}{0.999 + 1.46s}, \quad (24)$$

to design a low-order PID controller K_W

$$K_W(s) = \frac{5.85s+4}{s(s+2.83)} \quad (25)$$

Figure 5 presents the response of the output y for a set of parametric states of the plant to a unit step variation at $t = 0$ s of the reference signal e and to a 0.1 step disturbance d_u on the plant input at $t = 60$ s. Even if the nominal percent overshoot is 1.28%, it reaches 40% for a parametric state of the plant. The 90% response time is 9.09 s for the nominal plant and can reach about 15 s for another plant-parametric state.

As the responses presented by Fig. 5 show that the closed-loop responses can be very lightly damped, it is possible to use degree of freedom λ of the Zhang methodology to tune a robust controller that leads to an overshoot about 10% at least.

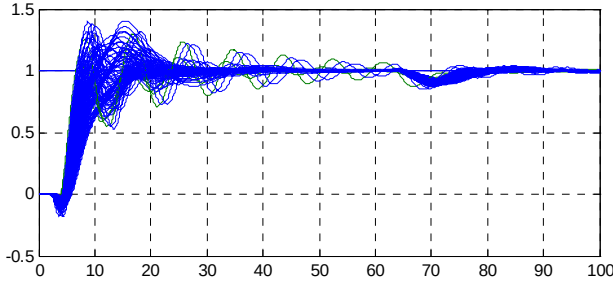


Fig. 5. Response $y(t)$ of the plant with the Wang controller for possible values of g , z , p , and d .

The controller defined by (7) is

$$K_Z(s) = \frac{(0.999 + 1.46s)^2}{\lambda^2 s^2 + (2\lambda - 1.46)s} \quad (26)$$

Taking into account the closed-loop time response obtained by time-domain simulations for all the possible parametric states of the plant, an iterative tuning leads to the optimal value $\lambda = 5.1$. Figure 6 presents the response of the plant controlled by the optimal robust Zhang controller.

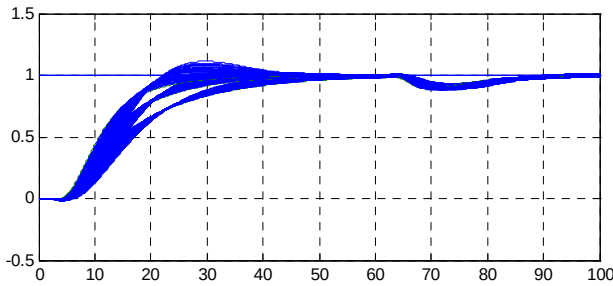


Fig. 6. Response $y(t)$ of the plant with a robust Zhang controller for possible values of g , z , p , and d .

The nominal and greatest values of the overshoot are respectively 0.07% and 11.3%. The nominal and greatest values of the 90% response time are respectively 24.9 s and 37.45 s.

Then, a third generation CRONE controller is designed to control the uncertain plant without using a Smith predictor structure. Figure 7 and Fig. 8 present the Bode and Nichols diagrams of the uncertain plant.

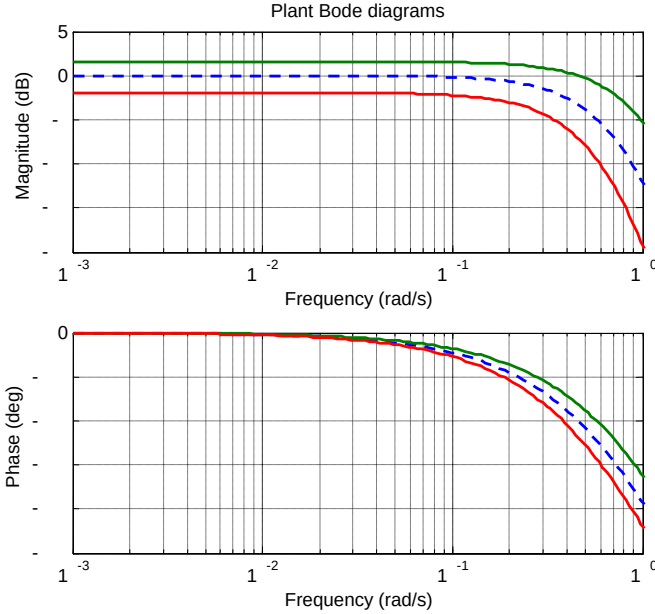


Fig. 7. Bode diagram of the nominal plant G_0 (---) and lower and greatest magnitude and phase of the uncertain plant G (—).

The time-delay and right half-plane zero of G_0 (22) are respectively 2 and 1. As plant low-frequency order is 0, order n_l of (12) equals 1 to reject any constant input disturbance d_u . As the plant relative degree is 4 and as the nominal open-loop transfer function β needs to include the plant right half-plane zero, order n_h of (13) equals 6 to obtain a strictly proper controller.

To be sure to have enough parameters to be tuned, orders N^- and N^+ of (11) are set to 1.

As a small overshoot is required, the nominal resonant peak used in the objective function (14) is $M_{r0} = 0.2$ dB.

Taking into account the five sensitivity function constraints (15–16) presented by Fig. 10, the 10 independent optimal parameters leads to the open-loop definition: $K = 0.56$, $\omega_1 = 0.0075$, $a_{-1} = 0.98$, $b_{-1} = 0.016$, $q_{-1} = 1$, $\omega_0 = 0.20$, $a_0 = -9.14$, $b_0 = 1.72$, $q_0 = 2$, $\omega_1 = 0.38$, $a_1 = 2.51$, $b_1 = 1.31$, $q_1 = 4$, $\omega_2 = 2.09$, and $\omega_r = 0.0507$, $Y_r = 6.03$ dB. Figure 9 presents the optimal open-loop Nichols locus.

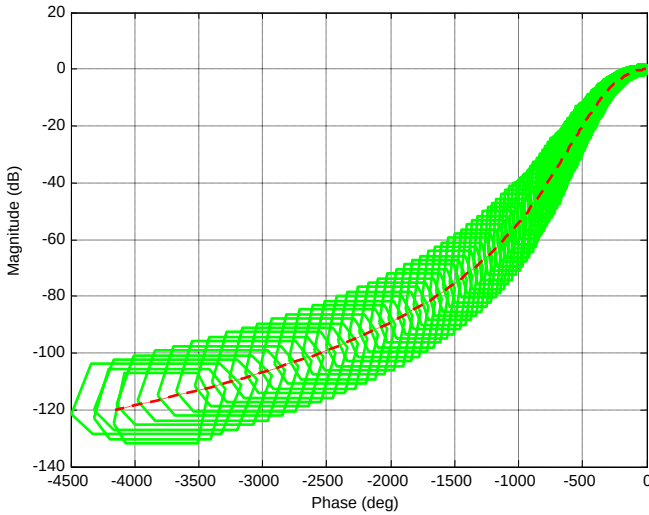


Fig. 8. Nominal plant Nichols locus (---) and uncertainty domains (—).

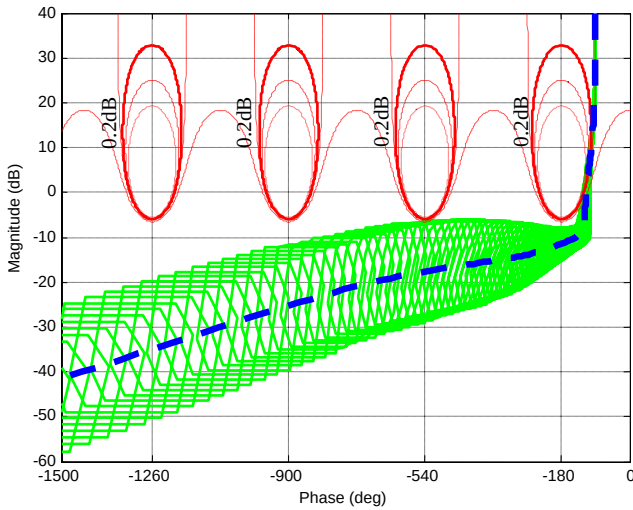


Fig. 9. Nominal open-loop Nichols locus (---), uncertainty domains (—) and 0.2dB M-contour.

By minimizing the cost function ($J_{\text{opt}} = 0.75\text{dB}$), the optimal template positions the uncertainty domains so that they overlap the 0.2dB M-contour as little as possible. The sensitivity functions met almost the constraints (Fig. 10). Only T_1 is exceeded of 0.23dB around 0.1rad/s. Using zeros and poles, the rational controller $C_R(s)$ is now synthesized from (18):

$$C_R(s) = \frac{127.85s^6 + 335.05s^5 + 413.94s^4 + 262.85s^3 + 75.828s^2 + 8.6076s + 0.19602}{0.11595s^7 + 3.0704s^6 + 30.739s^5 + 143.43s^4 + 282.11s^3 + 30.855s^2 + s} \quad (30)$$

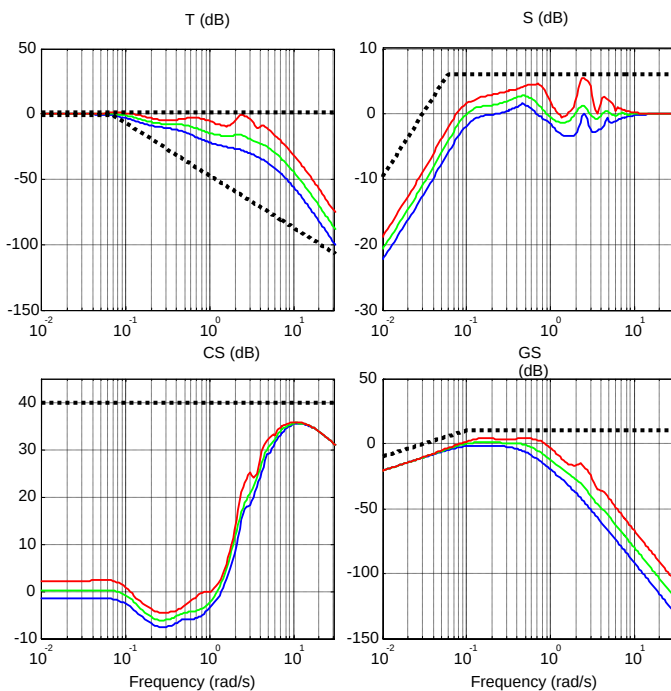


Fig. 10. Nominal and extreme closed-loop sensitivity function (—) and sensitivity function constraints (....).

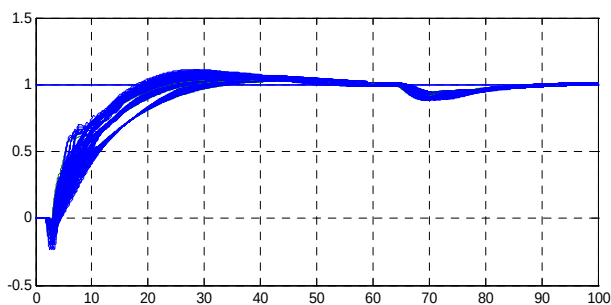


Fig. 11. Response $y(t)$ of the plant with the CRONE controller for possible values of g , z , p , and d .

The nominal and greatest values of the overshoot are respectively 5.7% and 11.4% (Fig. 11). The nominal and greatest values of the 90% response time are respectively 19.3 s and 24.7 s. Table 1 compares the performance (90% response time) and robustness (percent overshoot variation) obtained with the 3 controllers.

Table 1. 90% response time $t_{90\%}$ and percent overshoot O obtained with the 3 controllers.

Controller	$t_{90\% \text{ nom.}}$	$t_{90\% \text{ max.}}$	$O_{\text{nom.}}$	$O_{\text{max.}}$
Wang	9.09 s	15 s	1.28%	40%
Zhang	24.9 s	37.45 s	0.07%	11.3%
CRONE	19.3 s	24.7 s	5.7%	11.4%

Even if the Wang controller provides short 90% response times, it also provides very long settling times (Fig. 5) and great overshoots. The optimized robust Zhang controller provides greater 90% response times but shorter settling times (Fig. 6) and small variations of the overshoot. The CRONE controller provides small variations of the overshoot also, and shorter 90% response times than provided by the Zhang controller.

6 Conclusion

In the context of the control of time-delay systems, many modifications have been proposed to enhance the performance and robustness of control-systems based on Smith predictor structure. This paper has proposed to compare the performance of two Smith predictor-based controllers (Wang and Zhang controllers) including a mismatched model of the time-delay system, to the performance provided by a CRONE controller. For that comparison, a nonminimum phase plant with a time-delay is chosen. To assess the robustness of the controllers some uncertainty is added on each plant parameters. Even if it is more secure than a classical Smith predictor, the Wang controller reveals not to be robust enough. Based on the IMC method, the Zhang controller has been optimized using one degree of freedom correlated to the settling time of the closed-loop system. The time-domain optimization succeeds and provides a robust Zhang controller which provides perfectly acceptable performance. Using more high-level degree of freedom, a CRONE controller has been designed with a frequency domain methodology. As the genuine plant uncertainty is taken into account without any overestimation, the CRONE controller reveals to be both robust and with higher performance.

Then, it can be concluded that even if it can simplify the design of (robust) controller for time-delay system, the use of an improved Smith predictor is not necessary to obtain good performance.

References

1. Smith OJM (1957) Closer control of loops with dead time, *Chem. Eng. Progr.*, 53(5):217–219.
2. Normey-Rico JE, Camacho EF (1999) Smith predictor and modifications: a comparative study, *European Control Conference 1999 (ECC'99)*, Karlsruhe, Germany, August 31–September 3.
3. Wang QG, Bi Q, Zhang Y (2000) Re-design of Smith predictor systems for performance enhancement, *ISA Trans.*, 39:79–92.
4. Zhang WD, Xu XM (2001) Analytical design and analysis of mismatched Smith predictor, *ISA Trans.*, 40:133–138.
5. Morari M, Zafiriou E (1989) *Robust Process Control*, Prentice-Hall, Englewood Cliffs, NJ.
6. Manabe S (2003) Early development of fractional order control, *DETC'2003, 2003 ASME Design Engineering Technical Conferences*, Chicago, Illinois, September 2–6.
7. Podlubny I (1999) Fractional-order systems and PID-controllers, *IEEE Trans. Auto. Control*, 44(1):208–214.
8. Vinagre B, Chen YQ (2002) Lecture notes on fractional calculus applications in control and robotics, in: Vinagre Blas, YangQuan Chen, (ed.) *The 41st IEEE CDC2002 Tutorial Workshop 2*, pp. 1–310 http://mechatronics.ece.usu.edu/foc/cdc02_tw2_in.pdf, Las Vegas, Nevada, December 9.
9. Chen YQ, Vinagre BM, Podlubny I (2004) Fractional order disturbance observer for vibration suppression, *Nonlinear Dynamics*, Kluwer, 38(1–4): 355–367, December.
10. Monje CA, Vinagre BM, Chen YQ, Feliu V, Lanusse P, Sabatier J (2004) Proposals for fractional $PI^{\lambda}D^{\mu}$ tuning, *1st IFAC Workshop on Fractional Differentiation and its Applications*, FDA'04, Bordeaux, France, July 19–20.
11. Oustaloup A (1981) Linear feedback control systems of fractional order between 1 and 2, *IEEE Int. Symp. Circ. Syst.* Chicago, Illinois, April 27–29.
12. Oustaloup A (1981) *Systèmes asservis linéaires d'ordre fractionnaire*, PhD thesis, Bordeaux I University, France.
13. Oustaloup A (1983) *Systèmes asservis linéaires d'ordre fractionnaire*, Masson, Paris.
14. Oustaloup A, Ballouk A, Melchior P, Lanusse P, Elyagoubi A (1990) Un nouveau régulateur CRONE fondé sur la dérivation non entière complexe, *GR Automatique CNRS Meeting*, Bordeaux, France, March 29–30.
15. Oustaloup A (1991) The CRONE control, *ECC'91*, Grenoble, France, July 2–5.
16. Oustaloup A, Mathieu B, Lanusse P (1995) The CRONE control of resonant plants: application to a flexible transmission, *Eur. J. Control*, 1(2).
17. Oustaloup A (1999) *La Commande CRONE*, 2nd edition. Editions HERMES, Paris.
18. Lanusse P (1994) De la commande CRONE de première génération à la commande CRONE de troisième génération, PhD thesis, Bordeaux I University, France.

19. Horowitz IM (1993) *Quantitative Feedback Design Theory* – QFT, QFT Publications, Boulder, Colorado.
20. Landau ID, Rey D, Karimi A, Voda A, Franco A (1995) A flexible transmission system as a benchmark for digital control, *Eur. J. Control*, 1(2).
21. Pommier V, Sabatier J, Lanusse P, Oustaloup A (2002) CRONE control of a nonlinear hydraulic actuator, *Contr. Eng. Practi.*, 10(4):391–402.
22. Lanusse P, Oustaloup A, Mathieu B (2000) Robust control of LTI square MIMO plants using two CRONE control design approaches, *IFAC Symposium on Robust Control Design “ROCOND 2000”*, Prague, Czech Republic, June 21–23.
23. Francis BA, Zames G (1974) On H_∞ optimal sensitivity theory for SISO feedback systems, *IEEE Trans. Auto. Control*, 29:9–16.
24. Freudenberg JS, Looze JS (1985) Right half plane poles and zeros and design tradeoffs in feedback systems, *IEEE Trans. Auto. Control*, 30:555–565.
25. Kwakernaak H (1984) La commande robuste: optimisation à sensibilité mixte, Chapter 2 of *La robustesse*, coordinated by Oustaloup A, Editions HERMES, Paris.

ROBUST DESIGN OF AN ANTI-WINDUP COMPENSATED 3RD-GENERATION CRONE CONTROLLER

Patrick Lanusse, Alain Oustaloup, and Jocelyn Sabatier

LAPS, UMR 5131 CNRS, Bordeaux I University, ENSEIRB, 351 cours de la Libération, 33405 Talence Cedex; Tel: +33 (0)5 4000 2417 – Fax: +33 (0)5 4000 66 44, E-mail: {lanusse, oustaloup, sabatier}@laps.u-bordeaux1.fr

Abstract

Based on an electromechanical system to be digitally controlled, this paper shows how to add an anti-windup feature to a third-generation CRONE controller. First, the plant perturbed model is analyzed to build uncertainty domains. Then, a robust fractional controller is designed by taking into account small-level signal specification. A rational controller is synthesized from its required frequency response. Finally, to manage windup problem, a nonlinear controller is designed by splitting in two parts the optimal linear controller.

Keywords

Robust control, CRONE control, fractional-order control system, computer aided control-system Design.

1 Introduction

CRONE control methodology [1] is one of the most developed approaches to design robust and fractional-order controllers. Depending on the plant nature and on the required performance, one of the three generations of the CRONE (frequency-domain) methodology can be used. First and second generations are really easy to be used, the third one a little less but more performing.

The first generation is particularly adapted to control plants with a frequency response whose magnitude only is perturbed around the required closed-loop cutoff frequency and whose phase is constant with respect to the frequency around this cutoff frequency. Thus, the CRONE controller is defined

by a fractional-order n transfer function that can be considered as that of a fractional PID^n controller.

The second generation is also adapted to plant with a perturbed magnitude around the cutoff frequency but can deal with variable plant phase with respect to the frequency around this cutoff frequency. In the second CRONE generation, it is now the open-loop transfer function that is defined from a fractional-order n integrator. Then, as for the first generation, the rational controller can be obtained by using the well-known Oustaloup approximation method [2].

Unluckily, these two generations are not always sufficient to handle: more general plant perturbations than gain-like, nonminimum phase plants, time-delay or unstable plants, plants with bending modes, very various and hard-to-meet specification, etc. For the third generation [3, 4, 5], the nominal open-loop transfer function is defined from a band-limited complex fractional-order integrator and its few high-level parameters are optimized to minimize the sensitivity of the closed-loop stability degree to the perturbed plant parameters, and to permit the respect of the closed-loop required performance. Before designing the robust and performing control-system, it is sometimes difficult to translate the initial (time-domain) requirements to frequency-domain design specification and to set some of the open-loop parameters. As the robust controller is designed only taking into account small-level exogenous signals, an anti-windup system often needs to be included [6].

Using a laboratory plant digitally controlled as illustration example, this paper proposes to explain in detail: how the uncertainty of plant parameters is taken into account; how the digital implementation way could be taken into account correctly; how magnitude bounds are defined from specification to constraint the four common closed-loop sensitivity functions; how some open-loop transfer function parameters are set and how the others are optimized; how the rational robust controller is synthesized; how an anti-windup system is included; and finally how the controller is implemented. All these different steps will be illustrated using the CRONE control-system design toolbox [7] developed for Matlab/Simulink.

2 Introduction to Crone Control-System Design

The CRONE control-system design is based on the common unity-feedback configuration (Fig. 1). The controller or the open-loop transfer function is defined using integro-differentiation with non-integer (or fractional) order. The required robustness is that of both stability margins and performance, and particularly the robustness of the peak value M_r (called resonant peak) of the common complementary sensitivity function $T(s)$.

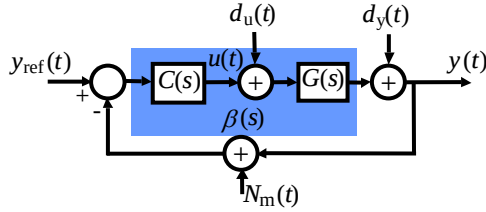


Fig. 1. Common CRONE control diagram.

Three CRONE control design methods have been developed, successively extending the application field. To design controller C , the third CRONE control generation must be used when the plant frequency uncertainty domains are of various types (not only gain like). The vertical template used in the second generation of the CRONE methodology is then replaced by a generalized template always described as a straight line in the Nichols chart but of any direction (complex fractional order integration), or by a multi-template (or curvilinear template) defined by a set of generalized templates.

An optimization allows the determination of the independent parameters of the open-loop transfer function. This optimization is based on the minimization of the stability degree variations, while respecting other specifications taken into account by constraints on sensitivity function magnitude. The complex fractional order permits parameterization of the open-loop transfer function with a small number of high-level parameters. The optimization of the control is thus reduced to only the search for the optimal values of these parameters. As the form of uncertainties taken into account is structured, this optimization is necessarily nonlinear. It is thus very important to limit the number of parameters to be optimized. After this optimization, the corresponding CRONE controller is synthesized as a rational fraction only for the optimal open-loop transfer function.

The third-generation CRONE system-design methodology, the most powerful one, is able to manage the robustness/performance tradeoff. It is also able to design controllers for plants with positive real part zeros or poles, time delay, and/or with lightly damped mode [8]. Associated with the w-bilinear variable change, it also permits the design of digital controllers. The CRONE control has also been extended to linear time variant systems and nonlinear systems whose nonlinear behaviors are taken into account by sets of linear equivalent behaviors [9]. For multi-input multi-output (MIMO) (multivariable) plants, two methods have been developed [10]. The choice of the method is done through an analysis of the coupling rate of the plant. when this rate is reasonable, one can opt for the simplicity of the multi single-input single-output (SISO) approach.

If CRONE design is only devoted to the closed-loop using the controller as one degree of freedom (DOF), it is obvious that a second DOF (F , linear or not) could be added outside the loop for managing pure tracking problems [11, 12]. Another solution is to implement the linear controller in a nonlinear way

that provides an anti-windup system to manage the saturation effect that appears for large variation of the closed-loop reference input.

3 Electromechanical System to be Controlled

The application example used here deals with the digital control of the angular position of a direct current (DC) motor (driven by a servo amplifier including a current control-loop) rigidly linked to another identical DC motor. This second motor can be used to achieve any type of load (see figure 2).

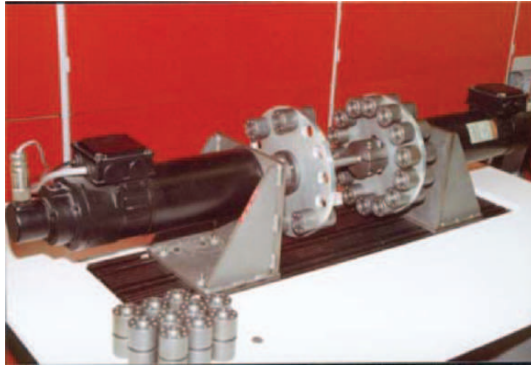


Fig. 2. Electromechanical system to be controlled.

Plant input u is provided by a 12 bit/ ± 10 V DAC and output y is measured by a 10 K counts per turn (CPT) incremental encoder. The control is digital and the sampling period T_s equals 2 ms. Counter, DAC and control law are managed by a homemade real-time software.

Twenty-eight individual payloads, 600 g each, can be added to modify the global inertia payload.

For all possible parametric states, the control system must satisfy the following performance specifications:

- Obtain a step response to a 5 turns variation of y_{ref} (50 K CPT) with a first overshoot about 3% as possible and a settling time as small as possible
- Limit the solicitation level of the plant input response u to ± 10 for a disturbance d_m of $\pm 1/2$
- Reduce the steady-state effect of constant disturbance d_u to zero

The transfer function which models the plant is:

$$G(s) = \frac{9092}{s(1 + \tau_m s)(1 + \tau_e s)}, \quad (1)$$

whose time-constant τ_e which equals 0.0047 s models the electrical high-frequency dynamics, and whose mechanical time-constant τ_m is within the interval [12, 96] when the inertia payload varies. For the nominal parametric state of the plant, τ_m equals 34 s ($\approx (12 * 96)^{1/2}$).

As part of a CRONE synthesis which is a continuous-time frequency approach, the initial digital control design problem is transformed into a pseudo-continuous-time problem by using the bilinear w variable change defined by:

$$z^{-1} = \frac{1-w}{1+w} \quad \text{or} \quad w = \frac{1-z^{-1}}{1+z^{-1}},$$

with $w = jv$ and $v = \tan\left(\frac{\omega T_s}{2}\right)$,

(2)

where v is a pseudo-frequency.

A zero-order hold is included in the calculation of the z -transform $G(z)$ of the continuous plant model. Due to the zero-order hold, $G(w)$ is a non-minimum-phase plant with two right half-plane zeros which equal +1 and +1.84. The uncertainty domains are computed for 120 pseudo-frequency v within the range $[10^{-4}, 10^2]$ and for nine log-spaced values of τ within the interval [12, 96]. Figure 3 presents the nominal Nichols locus of the plant and its uncertainty domains [4]. Even if the uncertainty domains of the plant to be controlled are almost vertical (gain like perturbation), it is interesting to use the third-generation CRONE methodology to manage not only the robustness but also the performance problem.

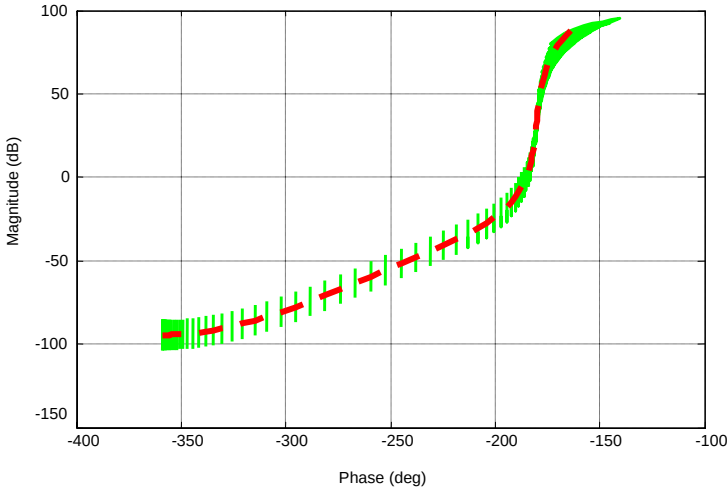


Fig. 3. Nominal plant Nichols locus (---) and uncertainty domains (—).

4 Third Generation Crone Methodology Application

Within a frequency range $[v_A, v_B]$ around open-loop gain-crossover pseudo frequency v_{cg} , the Nichols locus of a third-generation CRONE open-loop is defined by a any-angle straight line segment, called a generalized template (Fig. 4).

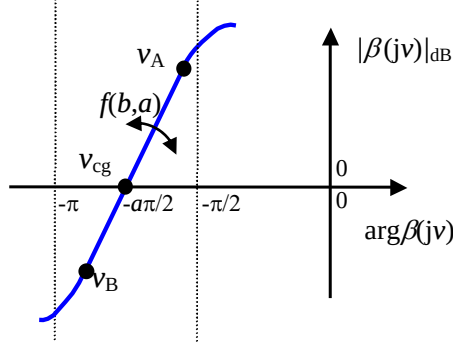


Fig. 4. Generalized template in the Nichols plane.

The generalized template can be defined by an integrator of complex fractional order n whose real part determines its phase location at frequency v_{cg} , that is $-\text{Re}_i(n)\pi/2$, and whose imaginary part then determines its angle to the vertical (Fig. 5).

The transfer function including complex fractional-order integration is:

$$\beta(w) = \left(\cosh\left(b \frac{\pi}{2}\right) \right)^{\text{sign}(b)} \left(\frac{v_{cg}}{w} \right)^a \left(\text{Re}_i \left(\left(\frac{v_{cg}}{w} \right)^{ib} \right) \right)^{-\text{sign}(b)} \quad (3)$$

with $n = a + ib \in \Re_i$ and $w \in \Re_j$, and where $\Re_i$ and $\Re_j$ are respectively time-domain and frequency-domain complex planes.

The definition of the open-loop transfer function including the nominal plant must take into account:

- The accuracy specifications at low frequencies
- The generalized template around pseudo frequency v_{cg}
- The plant behavior at high frequencies while respecting the control effort specifications at these frequencies

Thus, the open-loop transfer function is defined by a transfer function using band-limited complex fractional order integration:

$$\beta(w) = \beta_l(w) \beta_m(w) \beta_h(w), \quad (4)$$

where $\beta_m(w)$ is a set of band-limited generalized templates:

$$\beta_m(w) = \prod_{k=-N^-}^{N^+} \beta_k(w), \quad (5)$$

with:

$$\beta_k(w) = C_k^{\text{sign}(b_k)} \left(\alpha_k \frac{1+w/v_{k+1}}{1+w/v_k} \right)^{a_k} \left(\Re e_{/i} \left\{ \left(\alpha_k \frac{1+w/v_{k+1}}{1+w/v_k} \right)^{ib_k} \right\} \right)^{-q_k \text{sign}(b_k)}$$

$$\alpha_k = (v_{k+1}/v_k)^{1/2} \text{ for } k \neq 0 \text{ and } \alpha_0 = \left(1 + \left(\frac{v_r}{v_0} \right)^2 \right) / \left(1 + \left(\frac{v_r}{v_1} \right)^2 \right)^{1/2} \quad (6)$$

where $\beta_l(w)$ is an integer order n_l proportional integrator:

$$\beta_l(w) = C_l \left(\frac{v_{-N^-}}{w} + 1 \right)^{n_l} \quad (7)$$

where $\beta_h(w)$ is a low-pass filter of integer order n_h :

$$\beta_h(w) = \frac{C_h}{\left(\frac{w}{v_{N^+}} + 1 \right)^{n_h}} \quad (8)$$

The optimal open-loop transfer function is obtained by the minimization of the robustness cost function

$$J = \sup_{v, G} |T(jv)| - M_{r0}, \quad (9)$$

where M_{r0} is the resonant peak set for the nominal parametric state of the plant, while respecting the following set of inequality constraints for all plants (or parametric states of the plant) and for $v \in \nabla^+$:

$$\inf_G |T(jv)| \geq T_l(v) \text{ and } \sup_G |T(jv)| \leq T_u(v), \quad (10)$$

$$\sup_G |S(jv)| \leq S_u(v), \quad \sup_G |CS(jv)| \leq CS_u(v) \text{ and } \sup_G |GS(jv)| \leq GS_u(v), \quad (11)$$

$$\text{with } \begin{cases} T(w) = \frac{C(w)G(w)}{1+C(w)G(w)} & S(w) = \frac{1}{1+C(w)G(w)} \\ CS(w) = \frac{C(w)}{1+C(w)G(w)} & GS(w) = \frac{G(w)}{1+C(w)G(w)} \end{cases} \quad (12)$$

As the uncertainties are taken into account by the least conservative method, a nonlinear optimization method must be used to find the optimal values of the four independent parameters. The parameterization of the open-loop transfer function by complex fractional order of integration, then simplifies the optimization considerably. During optimization, the complex order modifies, alone, the shape of the open-loop frequency response as many parameters of common rational controllers could do it.

By taking $n_1 = 3$, an efficient integrator is introduced in the transfer function of the controller to nullify static error. As the nominal open-loop needs to include the two right half-plane plant zeros (1 and 1.84), the controller gain will decrease with $n_1 = 3$. To be sure to have enough parameters to be tuned, orders N^- and N^+ of (5) are set to 1. Nevertheless, to limit the number of parameters to be optimized, it is possible to set $b_{-1} = b_{+1} = 0$ [13]. So, the open-loop transfer function to be optimized is:

$$\beta(w) = K \left(\frac{v_{-1} + 1}{w} \right)^3 \prod_{k=1}^{k=1} \left(\frac{1 + \frac{w}{v_{k+1}}}{1 + \frac{w}{v_k}} \right)^{a_k} \left(\Re e_{/i} \left\{ \left(\alpha_0 \frac{1 + \frac{w}{v_1}}{1 + \frac{w}{v_0}} \right)^{ib_0} \right\} \right)^{-q_0 \text{sign}(b_0)} \frac{(1-w)(1-0.54w)}{\left(1 + \frac{w}{v_2} \right)^3} \quad (13)$$

The nominal resonance peak M_{r0} of the objective function (9) is set at 2.3 dB.

The five constraints of inequalities (11–12) are presented by Fig. 9 and are defined from the specifications:

- Up to $v = 3E-3$, T_u equals 1dB to limit the sluggishness of the responses $y(t)$ to step signals $e(t)$ and $d_y(t)$. Then, up to $v = 0.1$ T_u equals +5dB to limit the resonance peak M_r and then T_u is defined by a -20dB/decade slope to limit the effect of measurement noise $d_m(t)$.
- Up to $v = 1E-2$ T_1 equals -1dB to limit the sluggishness and the lowest value of the bandwidth. T_1 then is very small (-150dB).
- Up to $v = 1E-3$, S_u is defined by a $+60\text{dB/dec. slope}$. Up to $v = 1E-2$, it is defined by a $+20\text{dB/dec. slope}$ and S_u then equals $+6\text{dB}$ to limit the lowest value of the modulus margin to 0.5.
- From specifications, the greatest admissible magnitude of the control effort sensitivity function equals 20 (10:1/2). Thus, CS_u equals $+26\text{dB}$ up to $v = 2$, and is then defined by a -20dB/dec. slope .
- Up to $v = 2E-3$, GS_u is defined by a $+20\text{dB/dec. slope}$ to ensure the rejection of a step disturbance $d_u(t)$ modeling a Coulomb friction torque. GS_u then equals $+20\text{dB}$.

The eight parameters we choose to optimize are: the corner frequencies v_{-1} , v_0 , v_1 , and v_2 ; the real non-integer orders a_{-1} and a_1 ; the nominal resonance frequency v_r and the ordinate Y_r of the tangency point to the desired M contour for each set of these eight parameters, a_0 , b_0 , q_0 and K are computed so that $\beta(jv)$ tangents the M_{r0} M contour at ordinate Y_r and frequency v_r . The

optimization is achieved using the *fmincon* Matlab function. The optimized parameter values are: $Y_r = 3.83\text{dB}$; $v_r = 0.0192$; $a_{-1} = 0.90$; $a_1 = 0.80$; $v_{-1} = 0.00117$, $v_0 = 0.0115$, $v_1 = 0.0727$ and $v_2 = 0.204$. Thus, $a_0 = 1.57$, $b_0 = 0.66$, $q_0 = 1$, and $C = 33.3$. The final value of the cost function is null and all the constraints are verified. Figure 5 and 6 present the optimized open loop Nichols locus and sensitivity functions. The very good management of the robustness/performance tradeoff is proved by the perfect robustness of the stability degree (null cost function) and by the sensitivity functions very close to the performance constraints.

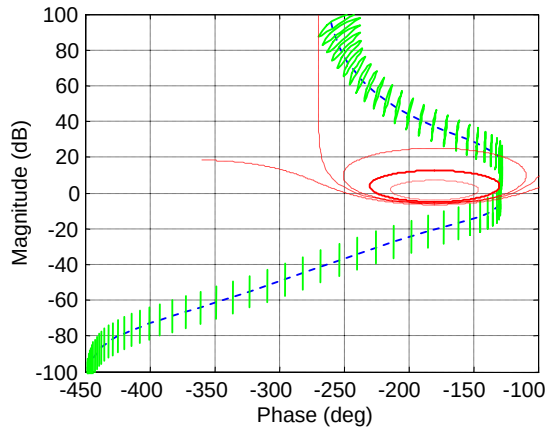


Fig. 5. Optimal nominal open-loop locus (- -) and uncertainty domains.

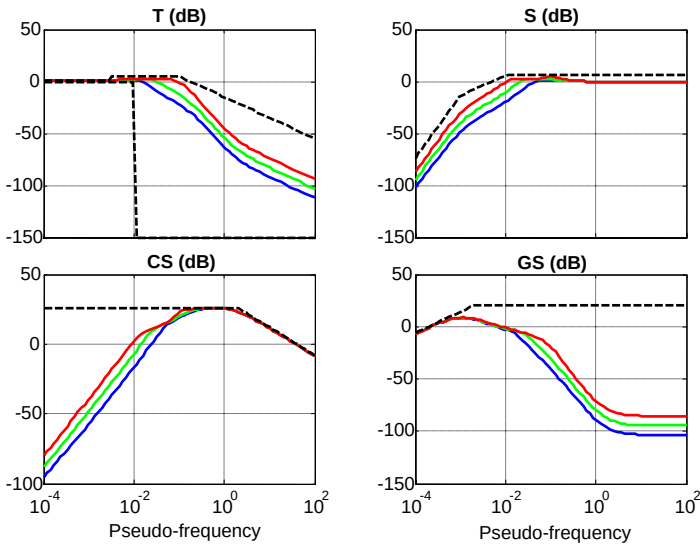


Fig. 6. Closed-loop sensitivity functions (nominal and extreme) compared to performance constraints (- -).

From the optimal nominal open-loop transfer, the fractional controller $C_F(w)$ is defined by its frequency response:

$$C_F(jv) = \frac{\beta(jv)}{G_0(jv)}, \quad (14)$$

where $G_0(jv)$ is the nominal frequency response of the plant. Then, the rational transfer function $C_R(w)$ of the controller can be synthesized by the approximation of the frequency response given by (14). The rational controller C_R is in the following form:

$$C_R(w) = C_0 w^{N_{\text{diff}}} \prod_{i=1}^n f(v_i, \zeta_i, o_i), \quad (15)$$

where, C_0 is a gain, N_{diff} , n , and o_i are integer orders, v_i are corner frequencies, and ζ_i are damping coefficients. When the order o_i is different from ± 2 , the function f is in the following form:

$$f(v_i, \zeta_i, o_i) = \left(1 + \frac{w}{v_i}\right)^{o_i} \quad (16)$$

When o_i equals ± 2 , f is in the following form:

$$f(v_i, \zeta_i, o_i) = \left(1 + \frac{2\zeta_i w}{v_i} + \frac{w^2}{v_i^2}\right)^{\text{sign}(o_i)} \quad (17)$$

Numerator and denominator of C_R are respectively order 4 and 5 polynomials: $N_{\text{diff}} = 1$; $C_0 = 0.000235$; $o_1 = 2$, $v_1 = 0.00114$, and $d_1 = 0.935$; $o_2 = -1$ and $v_2 = 0.0115$; $o_3 = 1$ and $v_3 = 0.0299$; $o_4 = 1$ and $v_4 = 0.0801$; $o_5 = -1$ and $v_5 = 0.129$; $o_6 = -1$ and $v_6 = 0.202$; $o_7 = -1$ and $v_7 = 1.63$.

Using the inverse w variable change, the transfer function $C_R(z^{-1})$ of the digital controller is obtained from the pseudo-continuous time-transfer function $C_R(w)$. Figure 7 shows the time responses obtained for a small 200 CPT (1/50th turn) step variation of y_{ref} . For greater step variations of y_{ref} , it is obvious that the control effort u will be saturated at ± 2048 , which certainly would lead to a windup problem.

5 Anti-Windup System Design

5.1 Principle

Whereas it is very useful to apply a plant input greater or equal to the saturation level ($\pm u_M = \pm 2048$) to ensure short settling times, it is very

important to be able to go out quickly of this saturation functioning mode. This could be achieved if the controller output is remained close to the saturation level.

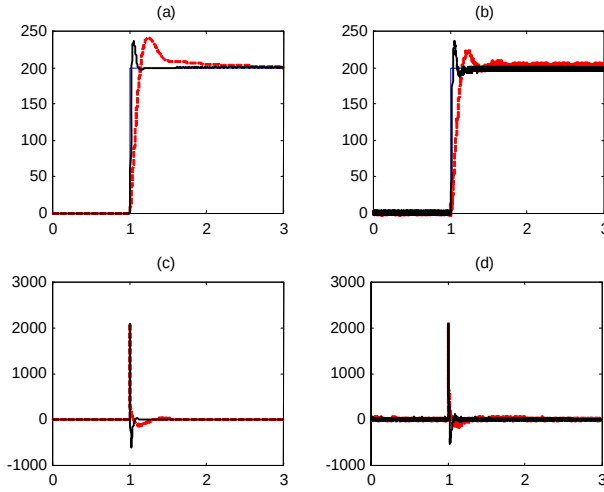


Fig. 7. Simulated (a and c) and actual (b and d) time responses of y (a and b) and u (c and d) for lowest and greatest payloads.

Thus, as presented in [6], taking into account the model of the plant nonlinearity, the plant linear model and the previous controller C_R optimized for low-level signals, it is possible to split the controller so that:

- The linear behavior of the new controller remains the same as before
- The output of the controller tracks its saturated value obtained by using a model of the plant saturation

The solution based on an inner loop which feedbacks a part of the controller is presented by Fig. 8.

For small signals ($|u(t)| \leq u_M$), the couple (K_y, K_u) must ensures that:

$$\frac{U(w)}{\varepsilon(w)} = \frac{K_y(w)}{1 + K_u(w)} = C_R(w) \quad (18)$$

C_R , K_y and K_u are respectively written:

$$C_R(w) = \frac{N_C(w) D(w)}{D_C(w) D(w)}, K_y(w) = \frac{N_y(w)}{\Delta(w)} \text{ and } K_u(w) = \frac{N_u(w)}{\Delta(w)} \quad (19)$$

where N_C , N_y , N_u , D_C , and Δ are polynomials, and where the polynomial D is with left half-plane zeros.

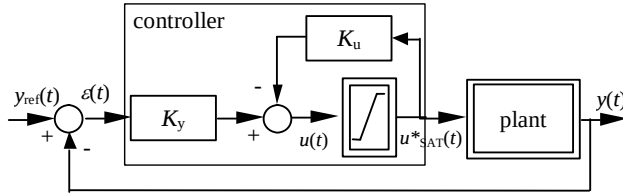


Fig. 8. New structure with anti-windup system.

Thus, (18) and (19) lead to

$$N_y(w) = N_C(w)D(w) \text{ and } N_u(w) = D_C(w)D(w) - \Delta(w) \quad (20)$$

The linear part of the open-loop transfer function L_{NL} is now

$$L_{NL}(w) = -\frac{U(w)}{U_{SAT}(w)} = K_u(w) + K_y(w)G(w) = -1 + \frac{D_C(w)D(w)}{\Delta(w)}(1 + L(w)) \quad (21)$$

As at low-frequency ($v \ll v_{cg} \Rightarrow |L(jv)| \gg 1$), this open-loop transfer function is such that:

$$L_{NL}(w) \approx \frac{D_C(w)D(w)}{\Delta(w)} L(w) \quad (22)$$

The transfer function $D_C D / \Delta$ looks like a compensator that permits the modification of the open-loop. Taking into account the describing function $N(u_1)$ of the saturation model, Fig. 9 shows that it is possible to increase the stability domain defined by u_{10} ($u'_{10} > u_{10}$ for L_{0NL1}) and sometimes to make disappear the stability problem (L_{0NL2}).

To ensure that the open-loop $L_{NL}(jv)$ equals $L(jv)$ at high-frequency ($v > v_{cg}$), D and Δ should be such that:

$$\lim_{v \rightarrow \infty} D_C(jv)D(jv) = \lim_{v \rightarrow \infty} \Delta(jv) \quad (23)$$

Thus it is useful that Δ includes the high-frequency part of D_C (i.e., all the zeros whose modulus are greater than v_{cg}). The other parts of Δ and D are determined to:

- Shape well the frequency response of the perturbed open-loop L_{NL}
- Ensure the degree condition:

$$\text{degree } D_C(w) + \text{degree } D(w) = \text{degree } \Delta(w) \quad (24)$$

Then polynomials N_u and N_y can be determined from polynomials D and Δ . Nevertheless, the global controller (K_u , K_y and saturation model) needs to be implementable (no algebraic loop) and thus the discrete-time filter $K_u(z)$ need to be strictly proper. It can be if

$$\lim_{z \rightarrow \infty} K_u(z) = \lim_{z \rightarrow \infty} \frac{D_C(z)D(z)}{\Delta(z)} - 1 = 0 \quad (25)$$

As $w = (z-1)/(z+1)$, (31) can be ensured if

$$\lim_{w \rightarrow 1} \frac{D_C(w)D(w)}{\Delta(w)} = 1, \quad (26)$$

which replaces the constraint given by (23).

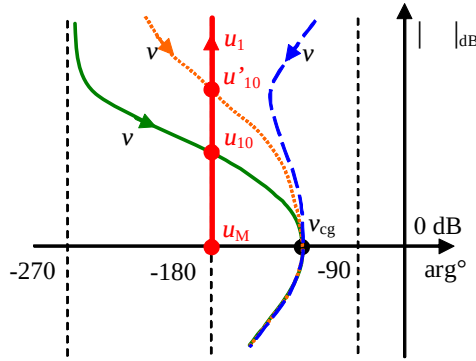


Fig. 9. Nichols plot of the negative inverse describing function $-1/N(u_1)$ and of nominal open-loop frequency responses: — $L_0(j\omega)$; $L_{0NL1}(j\omega)$ and - - - $L_{0NL2}(j\omega)$.

5.2 Design of the anti-windup system

The objective is to limit the overshoot of y at about 3% for step variations of y_{ref} lower than 50 K CPT (5 turns). It is obvious that the settling time needs to be minimized.

Two roots of D_C are lower than the gain crossover frequency v_{cg} : 0 (the integrator) and -0.0115 . Its three others roots are included in Δ : -0.129 , -0.202 and -1.63 . To add one more degree of freedom, D is chosen as a first degree polynomial. Condition (30) imposes to include 3 further roots in Δ . They are determined by taking into account the frequency response of the perturbed open-loop L_{NL} and the overshoot and 90% response time of the response y to the 50 K CPT step variation of y_{ref} . The 3 chosen roots are -0.0001 , -0.0006 and -0.0048 . Relation (26) leads to $D(w) = 1 + w/251.10^{-10}$. Using the inverse w variable change, the digital transfer functions of filters K_u and K_y are:

$$\begin{aligned}
K_u(z) &= \frac{0.01188z^{-1} - 0.037976z^{-2} + 0.042304z^{-3} - 0.016737z^{-4} - 0.00093439z^{-5} + 0.001464z^{-6}}{1 - 4.1844z^{-1} + 6.7183z^{-2} - 4.9259z^{-3} + 1.3125z^{-4} + 0.20127z^{-5} - 0.1218z^{-6}} \\
K_y(z) &= \frac{11.384 - 43.138z^{-1} + 49.833z^{-2} + 4.5813z^{-3} - 52.124z^{-4} + 38.557z^{-5} - 9.0934z^{-6}}{1 - 4.1844z^{-1} + 6.7183z^{-2} - 4.9259z^{-3} + 1.3125z^{-4} + 0.20127z^{-5} - 0.1218z^{-6}}
\end{aligned} \tag{27}$$

Figure 10 shows the Nichols locus of the nonlinear uncertain open loop. The chosen roots of Δ permit that the nonlinear open-loop does not cross the negative inverse $-1/N(u_1)$ of the describing function.

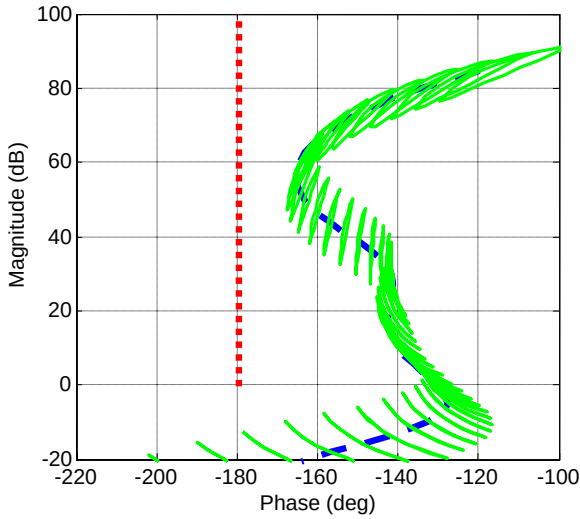


Fig. 10. Nonlinear open-loop Nichols locus (— — —), uncertainty domains (—) and $-1/N(u_1)$ (....).

Figure 11 shows the variation of the plant input and control effort for the 50 K CPT step variation of y_{ref} with or without anti-windup system. Using the anti-windup system, the greatest overshoot and the 90% response time are 2.7% and 0.77 s. For a double y_{ref} , the greatest overshoot and 90% response time are 10.5% and 1.03 s.

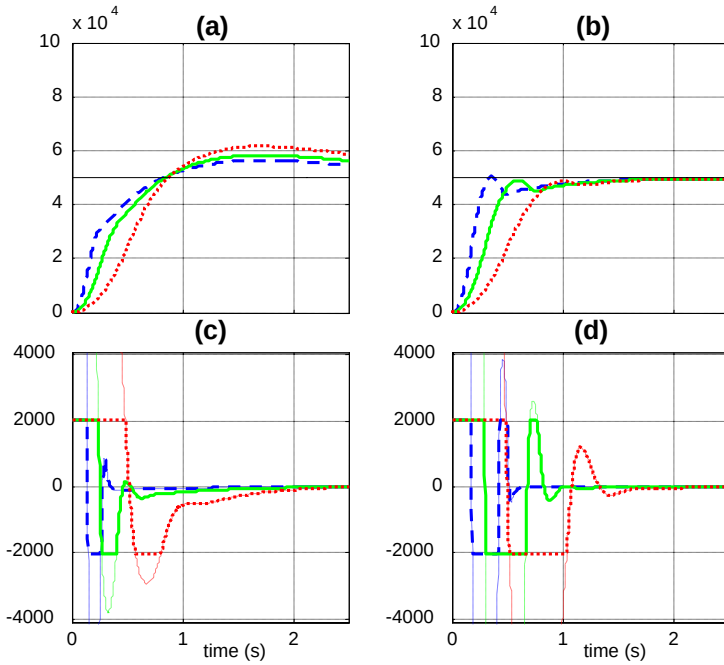


Fig. 11. Plant output (a and b) and control effort (c and d) without (a and c) and with (b and d) anti-windup system for nominal (—), lowest (---) and greatest (....) payloads.

6 Conclusion

As the system had to be digitally controlled, the CRONE control methodology has been applied in the pseudo-continuous time domain. Even if the plant perturbation is gain-like, the third generation methodology has been used to manage both robustness and performance. The obtained optimal linear controller is robust and met the specification translated in sensitivity function constraints. To avoid the windup problem that commonly appears for large variations of the reference signal, a nonlinear controller has been designed. It included both the linear controller split in two parts, and the model of the saturation of the plant. As the linear behavior of the nonlinear controller remains the same than that of the linear optimal controller, the closed-loop robustness remains ensured, and the sensitivities to small-level disturbances and to measurement noise remain optimal. The third-generation CRONE control application has been achieved by using the CRONE control-system design toolbox developed for Matlab/Simulink.

References

1. Oustaloup A, Mathieu B, Lanusse P, Sabatier J (1999) *La commande CRONE*, 2nd edition. Editions HERMES, Paris.
2. Oustaloup A, Levron F, Nanot F, Mathieu B (2000) Frequency-band complex non integer differentiator: characterization and synthesis, IEEE Trans. Circ. Syst., 47(1):25–40.
3. Oustaloup A (1991) The CRONE control, ECC'91, Grenoble, France.
4. Lanusse P (1994) De la commande CRONE de première génération à la commande CRONE de troisième génération, PhD thesis, Bordeaux I University, France.
5. Vinagre B, Chen YO (2002) Lecture notes on fractional calculus applications in control and robotics; in: Vinagre Blas, YangQuan Chen, (eds.) The 41st IEEE CDC2002 Tutorial Workshop #2, pp. 1–310 http://mechatronics.ece.usu.edu/foc/cdc02_tw2_in.pdf, Las Vegas, Nevada.
6. Lanusse P, Oustaloup A (2004) Windup compensation system for fractional controller; 1st IFAC Workshop on Fractional Differentiation and its Applications, FDA'04, Bordeaux, France.
7. Melchior P, Petit N, Lanusse P, Aoun M, Levron F, Oustaloup A (2004) Matlab based crone toolbox for fractional systems, 1st IFAC Workshop on Fractional Differentiation and its Applications, FDA'04, Bordeaux, France.
8. Oustaloup A, Mathieu B, Lanusse P (1995) The CRONE control of resonant plants: application to a flexible transmission, Eur. J. Control, 1(2).
9. Pommier V, Sabatier J, Lanusse P, Oustaloup A (2002) CRONE control of a nonlinear Hydraulic Actuator, Control Eng. Pract. 10(4):391–402.
10. Lanusse P, Oustaloup A, Mathieu B (2000) Robust control of LTI square MIMO plants using two CRONE control design approaches, IFAC Symposium on Robust Control Design “ROCOND 2000”, Prague, Czech Republic.
11. Melchior P, Poty A, Oustaloup A (2005) Path tracking design by frequency band-limited fractional differentiator prefilter; ENOC-2005, Eindhoven, Netherlands.
12. Orsoni B, Melchior P, Oustaloup A (2001) Davidson-Cole transfer function in path tracking, 6th IEEE European Control Conference ECC'2001, Porto, Portugal.
13. Sutter D (1997) La commande CRONE multiscalaire: application à des systèmes mécaniques articulés; PhD thesis, Bordeaux I University, France.

ROBUSTNESS OF FRACTIONAL-ORDER BOUNDARY CONTROL OF TIME FRACTIONAL WAVE EQUATIONS WITH DELAYED BOUNDARY MEASUREMENT USING THE SMITH PREDICTOR

Jinsong Liang¹, Weiwei Zhang², YangQuan Chen¹, and Igor Podlubny³

¹ Center for Self-Organizing and Intelligent Systems (CSOIS), Department of Electrical and Computer Engineering, Utah State University, 4120 Old Main Hill, Logan, UT 84322-4120;

E-mail: {jsliang,yqchen}@ieee.org

² Department of Mathematics, Michigan State University, East Lansing, MI 48824-1027;

E-mail: zhangwe6@msu.edu

³ Department of Information and Control of Processes, Technical University of Kosice, B. Nemcovej 3, 04200 Kosice, Slovak Republic;

E-mail: igor.podlubny@tuke.sk

Abstract

In this paper, we analyse the robustness of the fractional wave equation with a fractional-order boundary controller subject to delayed boundary measurement. Conditions are given to guarantee stability when the delay is small. For large delays, the Smith predictor is applied to solve the instability problem and the scheme is proved to be robust against a small difference between the assumed delay and the actual delay. The analysis shows that fractional-order controllers are better than integer order controllers in the robustness against delays in the boundary measurement.

Keywords

Fractional wave equation, fractional-order boundary control, measurement delay, Smith predictor.

1 Introduction

In recent years, boundary control of flexible systems has become an active research area, due to the increasing demand on high-precision control of many mechanical systems, such as spacecraft with flexible attachments or robots with flexible links, which are governed by PDEs (partial differential equations) rather than ODEs (ordinary differential equations) [1, 2, 3, 4, 5, 6, 7, 8, 9].

In this research area, the robustness of controllers against delays is an important topic and has been studied by many researchers [10, 11, 12, 13, 14, 15], due to the fact that delays are unavoidable in practical engineering.

Fractional diffusion and wave equations are obtained from the classical diffusion and wave equations by replacing the first- and second-order time derivative term by a fractional derivative of an order satisfying $0 < \alpha \leq 1$ and $1 < \alpha \leq 2$, respectively. Since many of the universal phenomena can be modelled accurately using the fractional diffusion and wave equations (see [16]), there has been a growing interest in investigating the solutions and properties of these evolution equations. Compared with the publications on control of integer order PDEs, results on control of fractional wave equations are relatively few [17, 18]. To the best of the authors' knowledge, there is still no publication on robust stabilization of fractional wave equations subject to delayed boundary measurement.

In this paper, we will investigate two robust stabilization problems of the fractional wave equations subject to delayed boundary measurement. First, under what conditions a very small delay in boundary measurement will not cause instability problems. Second, how to stabilize the system when the delay is large enough and makes the system unstable.

The paper is organized as follows. In section 2, the mathematical formulation is given. The robustness of boundary stabilization of fractional wave equation subject to a small delay in boundary measure is analysed in section 3. Section 4 investigates the large delay case and the corresponding compensation scheme. Finally, section 5 concludes this paper.

2 Problem Formulation

We consider a cable made with special smart materials governed by the fractional wave equation, fixed at one end, and stabilized by a boundary controller at the other end. Omitting the mass of the cable, the system can be represented by

$$\frac{\partial^\alpha u}{\partial t^\alpha} = \frac{\partial^2 u}{\partial x^2}, \quad 1 < \alpha \leq 2, \quad x \in [0, 1], \quad t \geq 0 \quad (1)$$

$$u(0, t) = 0, \quad (2)$$

$$u_x(1, t) = f(t), \quad (3)$$

$$u(x, 0) = u_0(x), \quad (4)$$

$$u_t(x, 0) = v_0(x), \quad (5)$$

where $u(x, t)$ is the displacement of the cable at $x \in [0, 1]$ and $t \geq 0$, $f(t)$ is the boundary control force at the free end of the cable, $u_0(x)$ and $v_0(x)$ are the initial conditions of displacement and velocity, respectively.

The control objective is to stabilize $u(x, t)$, given the initial conditions (4) and (5).

We adopt the following Caputo definition for fractional derivative of order α of any function $f(t)$, because the Laplace transform of the Caputo derivative allows utilization of initial values of classical integer-order derivatives with known physical interpretations [19, 20]

$$\frac{d^\alpha f(t)}{dt^\alpha} = \frac{1}{\Gamma(\alpha - n)} \int_0^t \frac{f^{(n)}(\tau) d\tau}{(t - \tau)^{\alpha+1-n}}, \quad (6)$$

where n is an integer satisfying $n - 1 < \alpha \leq n$ and Γ is the Euler's **gamma** function.

In this paper, we study the robustness of the controllers in the following format:

$$f(t) = -k \frac{d^\mu u(1, t)}{dt^\mu}, \quad 0 < \mu \leq 1 \quad (7)$$

where k is the controller gain, μ is the order of fractional derivative of the displacement at the free end of the cable.

Based on the definition (6), the Laplace transform of the fractional derivative is [19, 20]:

$$\left\{ \frac{d^\alpha f}{dt^\alpha} \right\} = s^\alpha F(s) - \sum_{k=0}^{n-1} f^{(k)}(0^+) s^{\alpha-1-k} \quad (8)$$

In the following, the transfer function from the boundary controller $f(t)$ to the tip end displacement will be derived for later use.

Assuming zero initial conditions of $u(x, 0)$ and $u_t(x, 0)$, take the Laplace transform of (1), (2), and (3) with respect to t , making use of (8), the original PDE of $u(x, t)$ with initial and boundary conditions can be transformed into the following ODE of $U(x, s)$ with boundary conditions.

$$\frac{d^2 U(x, s)}{dx^2} - s^\alpha U(x, s) = 0, \quad (9)$$

$$U(0, s) = 0, \quad (10)$$

$$U_x(1, s) = F(s), \quad (11)$$

where $U(x, s)$ is the Laplace transform of $u(x, t)$ and $F(s)$ is the Laplace transform of $f(t)$.

Solving the ODE (9), we have the following solution of $U(x, s)$ with two arbitrary constants C_1 and C_2 (s can be treated as a constant in this step).

$$U(x, s) = C_1 e^{xs^{\frac{\alpha}{2}}} + C_2 e^{-xs^{\frac{\alpha}{2}}}. \quad (12)$$

Substitute (12) into (10) and (11), we have the following two equations.

$$C_1 + C_2 = 0, \quad (13)$$

$$s^{\frac{\alpha}{2}}(C_1 e^{s^{\frac{\alpha}{2}}} - C_2 e^{-s^{\frac{\alpha}{2}}}) = F(s). \quad (14)$$

Solving (13) and (14) simultaneously, we can obtain the exact value of C_1 and C_2

$$C_1 = -C_2 = \frac{F(s)e^{s^{\frac{\alpha}{2}}}}{s^{\frac{\alpha}{2}}(e^{2s^{\frac{\alpha}{2}}} + 1)}. \quad (15)$$

Now we have obtained the solution of $U(x, s)$. Substituting $x = 1$ into $U(x, s)$ and divide $U(x, s)$ by $F(s)$, we obtain the following transfer function of the fractional wave equation $P(s)$:

$$P(s) = \frac{U(1, s)}{F(s)} = \frac{1 - e^{-2s^{\frac{\alpha}{2}}}}{s^{\frac{\alpha}{2}}(1 + e^{-2s^{\frac{\alpha}{2}}})}. \quad (16)$$

3 Robustness of Boundary Stabilization Subject to A Small Delay in Boundary Measurement

We consider the presence of a very small time delay θ in boundary measurement, shown as follows

$$f(t) = -ku_t^{(\mu)}(1, t - \theta), \quad (17)$$

where θ is the time delay.

The situation is also illustrated in Fig. 1, where $P(s)$ is the transfer function of the plant and $C(s)$ is the Laplace transform of the controller. In our case, $P(s)$ is (16) and $C(s)$ is

$$C(s) = k s^{\mu} \quad (18)$$

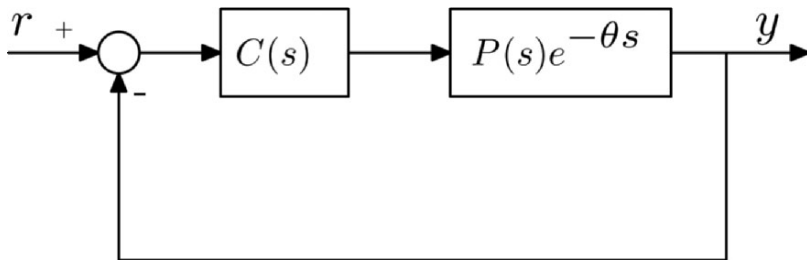


Fig. 1. A feedback control system with a time delay.

In [10, 11, 12, 13], it was shown that an arbitrarily small delay in boundary measurement causes the instability problem in boundary control of wave

equations using integer-order controllers $f(t) = -ku_t(1, t)$. Does this problem exist in boundary control of the fractional wave equation? Since fractional-order controllers are chosen in this paper, will this additional tuning knob bring us any benefits of robustness against the small delay? To answer these questions, we will first introduce a theorem presented in [13, 12].

Theorem 1. *Let $H(s)$ be the open-loop transfer function as illustrated in Fig. 2 and $\mathfrak{D}_H$ the set of all its poles. Define two closed-loop transfer functions $G_0(s)$ and $G_\epsilon(s)$ as*

$$G_0(s) = \frac{H(s)}{1 + H(s)},$$

and

$$G_\epsilon(s) = \frac{H(s)}{1 + e^{-\epsilon s} H(s)}.$$

Define again

$$\mathbb{C}_0 = \{s \in \mathbb{C} | \Re(s) > 0\},$$

and

$$\gamma(H(s)) = \limsup_{|s| \rightarrow \infty, s \in \mathbb{C}_0 \setminus \mathfrak{D}_H} |H(s)|.$$

Suppose G_0 is L^2 -stable. If $\gamma(H) < 1$, then there exists ϵ^* such that G_ϵ is L^2 -stable for all $\epsilon \in (0, \epsilon^*)$.

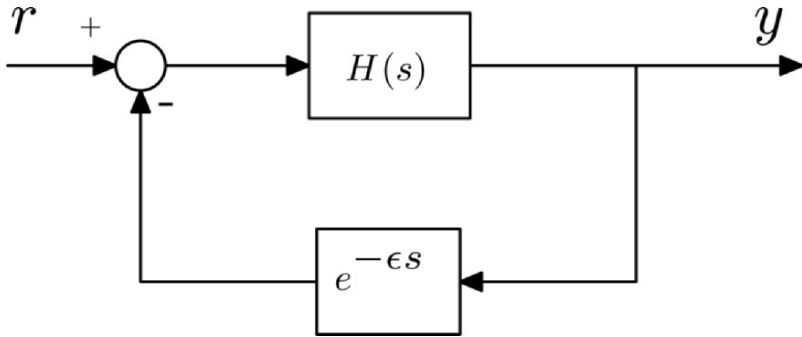


Fig. 2. Feedback system with delay.

The underlying idea of the above theorem is that the robustness of the closed-loop transfer function $G_0(s)$ against a small unknown delay can be determined by studying the open-loop transfer function $H(s)$. Notice that $H(s) = C(s)P(s)$ in our case.

CLAIM:

If the derivative order μ of controller (7) and the fractional-order α in the fractional wave equation (1) satisfy

$$\mu < \frac{\alpha}{2}, \quad (19)$$

then the system is stable for a small enough delay θ in boundary measurement.

Proof:

For $s \in \mathbb{C}_0$,

$$\begin{aligned} |H(s)| &= |C(s)P(s)| \\ &= \left| \frac{ks^\mu(1 - e^{-2s^{\frac{\alpha}{2}}})}{s^{\frac{\alpha}{2}}(1 + e^{-2s^{\frac{\alpha}{2}}})} \right| \\ &= \left| \frac{k(1 - e^{-2s^{\frac{\alpha}{2}}})}{s^{(\frac{\alpha}{2}-\mu)}(1 + e^{-2s^{\frac{\alpha}{2}}})} \right| \\ &\leq \frac{k|1 - e^{-2s^{\frac{\alpha}{2}}}|}{|s^{(\frac{\alpha}{2}-\mu)}||1 + e^{-2s^{\frac{\alpha}{2}}}|} \end{aligned} \quad (20)$$

Since $\frac{\alpha}{2} > \mu$, $|s^{(\frac{\alpha}{2}-\mu)}| \rightarrow \infty$ for $|s| \rightarrow \infty$.

Since $\frac{1}{2} < \frac{\alpha}{2} < 1$, for $|s|$ large enough, $|1 - e^{-2s^{\frac{\alpha}{2}}}|$ is bounded and $|1 - e^{-2s^{\frac{\alpha}{2}}}| > \eta > 0$, where η is a positive number.

So

$$\limsup_{|s| \rightarrow \infty, s \in \mathbb{C}_0} |H(s)| = 0 < 1.$$

Following the above proof, it can be easily proved that an integer-order controller $f(t) = -ku_t(1, t)$ is not robust against an arbitrarily small delay.

4 Compensation of Large Delays in Boundary Measurement Using the Smith Predictor

In the last section, it is shown that an fractional-order controller is robust against a small delay under the condition (19). In this section, we investigate the problem that what if the delay is large and makes the system unstable? We will apply the Smith predictor to solve this problem.

4.1 A brief introduction to the Smith predictor

The Smith predictor was proposed by Smith in [21] and is probably the most famous method for control of systems with time delays [22, 23]. Consider a typical feedback control system with a time delay in Fig. 1, where $C(s)$ is the controller; $P(s)e^{-\theta s}$ is the plant with a time delay θ .

With the presence of the time delay, the transfer function of the closed-loop system relating the output $y(s)$ to the reference $r(s)$ becomes

$$\frac{y(s)}{r(s)} = \frac{C(s)P(s)e^{-\theta s}}{1 + C(s)P(s)e^{-\theta s}}. \quad (21)$$

Obviously, the time delay θ directly changes the closed-loop poles. Usually, the time delay reduces the stability margin of the control system, or more seriously, destabilizes the system.

The classical configuration of a system containing a Smith predictor is depicted in Fig. 3, where $\hat{P}_0(s)$ is the assumed model of $P_0(s)$ and $\hat{\theta}$ is the assumed delay. The block $C(s)$ combined with the block $\hat{P}(s) - \hat{P}(s)e^{-\hat{\theta}s}$ is called “the Smith predictor”. If we assume the perfect model matching, i.e., $\hat{P}_0(s) = P_0(s)$ and $\theta = \hat{\theta}$, the closed-loop transfer function becomes

$$\frac{y(s)}{r(s)} = \frac{C(s)P(s)e^{-\theta s}}{1 + C(s)P(s)}. \quad (22)$$

Now, it is clear what the underlying idea of the Smith predictor is. With the perfect model matching, the time delay can be removed from the denominator of the transfer function, making the closed-loop stability irrelevant to the time delay.

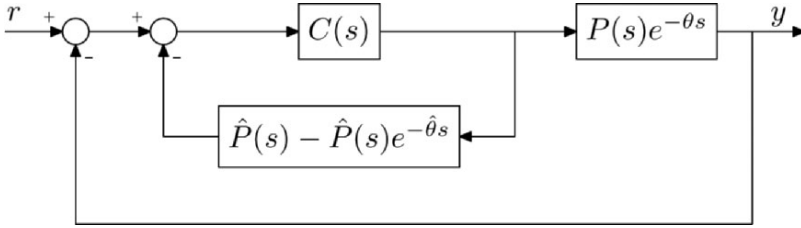


Fig. 3. The Smith predictor.

Based on the controller (18) as $C(s)$, we have the following expression of the boundary controller (the Smith predictor), denoted as $C_{sp}(s)$:

$$C_{sp}(s) = \frac{ks^\mu}{1 + ks^\mu P(s)(1 - e^{-\hat{\theta}s})} \quad (23)$$

4.2 Robustness analysis of the Smith predictor

In section 4.1, it is shown that if the assumed delay is equal to the actual delay, the Smith predictor removes the delay term completely from the denominator of the closed-loop. However, the actual delay is not exactly known. In this

section, we will investigate what if an unknown small difference ϵ between the assumed delay and the actual delay is introduced to the system, as shown in Fig. 4.

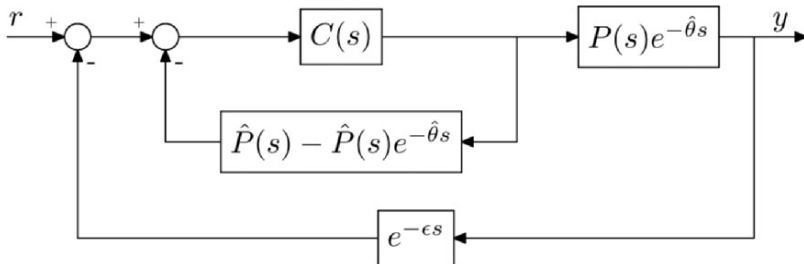


Fig. 4. System with mismatched delays.

CLAIM:

If $\hat{\theta}$ is chosen as the minimum value of the possible delay and μ is chosen to satisfy (19), then the controller (23) is robust against a small difference ϵ between the assumed delay $\hat{\theta}$ and the actual delay $\theta = \hat{\theta} + \epsilon$.

Proof:

For $s \in \mathbb{C}_0$,

$$\begin{aligned}
 |H(s)| &= \left| \frac{ks^\mu P(s)e^{-\hat{\theta}s}}{1 + ks^\mu P(s)(1 - e^{-\hat{\theta}s})} \right| \\
 &\leq \frac{k|1 - e^{-2s^{\frac{\alpha}{2}}}| |e^{-\theta s}|}{|s^{(\frac{\alpha}{2}-\mu)}(1 + e^{-2s^{\frac{\alpha}{2}}}) + k(1 - e^{-2s^{\frac{\alpha}{2}}})(1 - e^{-\theta s})|} \\
 &< \frac{k|1 - e^{-2s^{\frac{\alpha}{2}}}|}{\left| |s^{(\frac{\alpha}{2}-\mu)}(1 + e^{-2s^{\frac{\alpha}{2}}})| - k|(1 - e^{-2s^{\frac{\alpha}{2}}})(1 - e^{-\theta s})| \right|}
 \end{aligned}$$

When $|s| \rightarrow \infty$,

$$|s^{(\frac{\alpha}{2}-\mu)}(1 + e^{-2s^{\frac{\alpha}{2}}})| \rightarrow \infty,$$

while both $|1 - e^{-2s^{\frac{\alpha}{2}}}|$ and $|(1 - e^{-2s^{\frac{\alpha}{2}}})(1 - e^{-\theta s})|$ are bounded.

So

$$\limsup_{|s| \rightarrow \infty, s \in \mathbb{C}_0} |H(s)| = 0 < 1.$$

Remarks:

In *Theorem 1*, ϵ is positive. To satisfy this condition, $\hat{\theta}$ should be chosen as the minimal value of the possible delay.

5 Concluding Remarks

In boundary stabilization of the fractional wave equation, well-designed fractional-order controllers are robust against a small delay in boundary measurement; while the integer-order controller is unstable with an arbitrarily small delay introduced in boundary measurement. For large delays which makes the system unstable, the fractional-order controller combined with the Smith predictor is able to compensate the time delay and robust against a small difference between the assumed delay and the actual delay.

Acknowledgment

We acknowledge that this paper is a modified version of a paper published in the Proceedings of IDETC/CIE 2005 (Paper# DETC2005-85299). We would like to thank the ASME for granting us permission in written form to publish a modified version of IDETC/CIE 2005 (Paper# DETC2005-85299) as a chapter in the book entitled *Advances in Fractional Calculus: Theoretical Developments and Applications in Physics and Engineering* edited by Professors Machado, Sabatier, and Agrawal (Springer).

References

1. Morgül Ö (2002) An exponential stability result for the wave equation, *Automatica*, 38:731–735.
2. Morgül Ö (2001) Stabilization and disturbance rejection for the beam equation, *IEEE Trans. Auto. Control*, 46(12):1913–1918.
3. Conrad F, Morgül Ö (1998) On the stability of a flexible beam with a tip mass, *SIAM J. Control Optim.* 36(6):1962–1986.
4. Morgül Ö (1998) Stabilization and disturbance rejection for the wave equation, *IEEE Trans. Auto. Control*, 43(1):89–95.
5. Guo B-Z (2001) Riesz basis approach to the stabilization of a flexible beam with a tip mass, *SIAM J. Control Optim.*, 39(6):1736–1747.
6. Guo B-Z (2002) Riesz basis property and exponential stability of controlled euler-bernoulli beam equations with variable coefficients, *SIAM J. Control Optim.*, 40(6):1905–1923.
7. Chen G (1979) Energy decay estimates and exact boundary value controllability for the wave equation in a bounded domain, *J. Math. Pure. Appl.*, 58:249–273.
8. Chen G, Delfour MC, Krall AM, Payre G (1987) Modelling, stabilization and control of serially connected beams, *SIAM J. Contr. Optimiz.*, 25:526–546.
9. Morgül Ö (2002) On the boundary control of beam equation, In: *Proceedings of the 15-th IFAC World Congress on Automatic Control*.
10. Datko R, Lagnese J, Polis MP (1986) An example on the effect of time delays in boundary feedback stabilization of wave equations, *SIAM J. Control Optim.*, 24:152–156.
11. Datko R (1993) Two examples of ill-posedness with respect to small time delays in stabilized elastic systems, *IEEE Trans. Auto. Control*, 38(1):163–166.

12. Logemann H, Rebarber R (1998) PDEs with distributed control and delay in the loop: transfer function poles, exponential modes and robustness of stability, *Eur. J. Control*, 4(4):333–344.
13. Logemann H, Rebarber R, Weiss G (1996) Conditions for robustness and nonrobustness of the stability of feedback systems with respect to small delays in the feedback loop, *SIAM J. Control Optim.*, 34(2):572–600.
14. Morgül Ö (1995) On the stabilization and stability robustness against small delays of some damped wave equations, *IEEE Trans. Auto. Control*, 40(9):1626–1623.
15. Liang J, Zhang W, Chen Y (2005) Robustness of boundary control of damped wave equations with large delays at boundary measurement, In: *The 16th IFAC World Congress*.
16. Nigmatullin RR (1986) Realization of the generalized transfer equation in a medium with fractal geometry, *Phys. Stat. Sol. (b)*, 133:425–430.
17. Matignon D, d'Andréa–Novel B (1995) Spectral and time-domain consequences of an integro-differential perturbation of the wave PDE, In: *SIAM proceedings of the third international conference on mathematical and numerical aspects of wave propagation phenomena*, France, pp. 769–771.
18. Liang J, Chen YQ, Fullmer R (2003) Simulation studies on the boundary stabilization and disturbance rejection for fractional diffusion-wave equation, In: *2004 IEEE American Control Conference*.
19. Caputo M (1967) Linear models of dissipation whose q is almost frequency independent-II, *Geophys. J. R. Astronom. Soc.*, 13:529–539.
20. Podlubny I (1999) *Fractional Differential Equations*. Academic Press, San Diego, CA.
21. Smith OJM (1957) Closer control of loops with dead time, *Chem. Eng. Progress*, 53(5):217–219.
22. Levine W (ed.) (1996) *The Control Handbook*. CRC Press, Boca Raton, FL, pp. 224–237.
23. Wang Q-G, Lee TH, Tan KK (1999) *Finite Spectrum Assignment for Time-delay Systems*. Springer, London.