

DSP


DIGITAL SIGNAL AND IMAGE PROCESSING SERIES



Bayesian Approach to Inverse Problems

Edited by Jérôme Idier

ISTE

 **WILEY**

Bayesian Approach to Inverse Problems

Bayesian Approach to Inverse Problems

Edited by
Jérôme Idier

ISTE

 **WILEY**

First published in France in 2001 by Hermes Science/Lavoisier entitled "Approche bayésienne pour les problèmes inverses"

First published in Great Britain and the United States in 2008 by ISTE Ltd and John Wiley & Sons, Inc.

Translation made by Mrs Becker

Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act 1988, this publication may only be reproduced, stored or transmitted, in any form or by any means, with the prior permission in writing of the publishers, or in the case of reprographic reproduction in accordance with the terms and licenses issued by the CLA. Enquiries concerning reproduction outside these terms should be sent to the publishers at the undermentioned address:

ISTE Ltd
6 Fitzroy Square
London W1T 5DX
UK

www.iste.co.uk

John Wiley & Sons, Inc.
111 River Street
Hoboken, NJ 07030
USA

www.wiley.com

© ISTE Ltd, 2008

© LAVOISIER, 2001

The rights of Jérôme Idier to be identified as the author of this work have been asserted by him in accordance with the Copyright, Designs and Patents Act 1988.

Library of Congress Cataloging-in-Publication Data

Bayesian approach to inverse problems / edited by Jérôme Idier.

p. cm.

Includes bibliographical references and index.

ISBN: 978-1-84821-032-5

1. Inverse problems (Differential equations) 2. Bayesian statistical decision theory I. Idier, Jérôme.

QA371.B365 2008

515'.357--dc22

2007047723

British Library Cataloguing-in-Publication Data

A CIP record for this book is available from the British Library

ISBN: 978-1-84821-032-5

Printed and bound in Great Britain by Antony Rowe Ltd, Chippenham, Wiltshire.



Mixed Sources
Product group from well-managed
forests and other controlled sources

Cert no. SGS-COC-2953
www.fsc.org
© 1996 Forest Stewardship Council

Table of Contents

Introduction	15
Jérôme IDIER	
PART I. FUNDAMENTAL PROBLEMS AND TOOLS	23
Chapter 1. Inverse Problems, Ill-posed Problems	25
Guy DEMOMENT, Jérôme IDIER	
1.1. Introduction	25
1.2. Basic example	26
1.3. Ill-posed problem	30
1.3.1. Case of discrete data	31
1.3.2. Continuous case	32
1.4. Generalized inversion	34
1.4.1. Pseudo-solutions	35
1.4.2. Generalized solutions	35
1.4.3. Example	35
1.5. Discretization and conditioning	36
1.6. Conclusion	38
1.7. Bibliography	39
Chapter 2. Main Approaches to the Regularization of Ill-posed Problems .	41
Guy DEMOMENT, Jérôme IDIER	
2.1. Regularization	41
2.1.1. Dimensionality control	42
2.1.1.1. Truncated singular value decomposition	42
2.1.1.2. Change of discretization	43
2.1.1.3. Iterative methods	43
2.1.2. Minimization of a composite criterion	44
2.1.2.1. Euclidian distances	45

2.1.2.2. Roughness measures	46
2.1.2.3. Non-quadratic penalization	47
2.1.2.4. Kullback pseudo-distance	47
2.2. Criterion descent methods	48
2.2.1. Criterion minimization for inversion	48
2.2.2. The quadratic case	49
2.2.2.1. Non-iterative techniques	49
2.2.2.2. Iterative techniques	50
2.2.3. The convex case	51
2.2.4. General case	52
2.3. Choice of regularization coefficient	53
2.3.1. Residual error energy control	53
2.3.2. “L-curve” method	53
2.3.3. Cross-validation	54
2.4. Bibliography	56
Chapter 3. Inversion within the Probabilistic Framework	59
Guy DEMOMENT, Yves GOUSSARD	
3.1. Inversion and inference	59
3.2. Statistical inference	60
3.2.1. Noise law and direct distribution for data	61
3.2.2. Maximum likelihood estimation	63
3.3. Bayesian approach to inversion	64
3.4. Links with deterministic methods	66
3.5. Choice of hyperparameters	67
3.6. <i>A priori</i> model	68
3.7. Choice of criteria	70
3.8. The linear, Gaussian case	71
3.8.1. Statistical properties of the solution	71
3.8.2. Calculation of marginal likelihood	73
3.8.3. Wiener filtering	74
3.9. Bibliography	76
PART II. DECONVOLUTION	79
Chapter 4. Inverse Filtering and Other Linear Methods	81
Guy LE BESNERAIS, Jean-François GIOVANNELLI, Guy DEMOMENT	
4.1. Introduction	81
4.2. Continuous-time deconvolution	82
4.2.1. Inverse filtering	82
4.2.2. Wiener filtering	84
4.3. Discretization of the problem	85
4.3.1. Choice of a quadrature method	85

4.3.2. Structure of observation matrix \mathbf{H}	87
4.3.3. Usual boundary conditions	89
4.3.4. Problem conditioning	89
4.3.4.1. Case of the circulant matrix	90
4.3.4.2. Case of the Toeplitz matrix	90
4.3.4.3. Opposition between resolution and conditioning	91
4.3.5. Generalized inversion	91
4.4. Batch deconvolution	92
4.4.1. Preliminary choices	92
4.4.2. Matrix form of the estimate	93
4.4.3. Hunt's method (periodic boundary hypothesis)	94
4.4.4. Exact inversion methods in the stationary case	96
4.4.5. Case of non-stationary signals	98
4.4.6. Results and discussion on examples	98
4.4.6.1. Compromise between bias and variance in 1D deconvolution	98
4.4.6.2. Results for 2D processing	100
4.5. Recursive deconvolution	102
4.5.1. Kalman filtering	102
4.5.2. Degenerate state model and recursive least squares	104
4.5.3. Autoregressive state model	105
4.5.3.1. Initialization	106
4.5.3.2. Criterion minimized by Kalman smoother	107
4.5.3.3. Example of result	108
4.5.4. Fast Kalman filtering	108
4.5.5. Asymptotic techniques in the stationary case	110
4.5.5.1. Asymptotic Kalman filtering	110
4.5.5.2. Small kernel Wiener filter	111
4.5.6. ARMA model and non-standard Kalman filtering	111
4.5.7. Case of non-stationary signals	111
4.5.8. On-line processing: 2D case	112
4.6. Conclusion	112
4.7. Bibliography	113

Chapter 5. Deconvolution of Spike Trains 117

Frédéric CHAMPAGNAT, Yves GOUSSARD, Stéphane GAUTIER, Jérôme IDIER

5.1. Introduction	117
5.2. Penalization of reflectivities, L2LP/L2Hy deconvolutions	119
5.2.1. Quadratic regularization	121
5.2.2. Non-quadratic regularization	122
5.2.3. L2LP or L2Hy deconvolution	123
5.3. Bernoulli-Gaussian deconvolution	124
5.3.1. Compound BG model	124
5.3.2. Various strategies for estimation	124

5.3.3. General expression for marginal likelihood	125
5.3.4. An iterative method for BG deconvolution	126
5.3.5. Other methods	128
5.4. Examples of processing and discussion	130
5.4.1. Nature of the solutions	130
5.4.2. Setting the parameters	132
5.4.3. Numerical complexity	133
5.5. Extensions	133
5.5.1. Generalization of structures of \mathbf{R} and \mathbf{H}	134
5.5.2. Estimation of the impulse response	134
5.6. Conclusion	136
5.7. Bibliography	137

Chapter 6. Deconvolution of Images 141

Jérôme IDIER, Laure BLANC-FÉRAUD

6.1. Introduction	141
6.2. Regularization in the Tikhonov sense	142
6.2.1. Principle	142
6.2.1.1. Case of a monovariate signal	142
6.2.1.2. Multivariate extensions	143
6.2.1.3. Discrete framework	144
6.2.2. Connection with image processing by linear PDE	144
6.2.3. Limits of Tikhonov's approach	145
6.3. Detection-estimation	148
6.3.1. Principle	148
6.3.2. Disadvantages	149
6.4. Non-quadratic approach	150
6.4.1. Detection-estimation and non-convex penalization	154
6.4.2. Anisotropic diffusion by PDE	155
6.5. Half-quadratic augmented criteria	156
6.5.1. Duality between non-quadratic criteria and HQ criteria	157
6.5.2. Minimization of HQ criteria	158
6.5.2.1. Principle of relaxation	158
6.5.2.2. Case of a convex function ϕ	159
6.5.2.3. Case of a non-convex function ϕ	159
6.6. Application in image deconvolution	159
6.6.1. Calculation of the solution	159
6.6.2. Example	161
6.7. Conclusion	164
6.8. Bibliography	165

PART III. ADVANCED PROBLEMS AND TOOLS 169**Chapter 7. Gibbs-Markov Image Models 171**

Jérôme IDIER

7.1. Introduction	171
7.2. Bayesian statistical framework	172
7.3. Gibbs-Markov fields	173
7.3.1. Gibbs fields	174
7.3.1.1. Definition	174
7.3.1.2. Trivial examples	175
7.3.1.3. Pairwise interactions, improper laws	176
7.3.1.4. Markov chains	176
7.3.1.5. Minimum cliques, non-uniqueness of potential	177
7.3.2. Gibbs-Markov equivalence	177
7.3.2.1. Neighborhood relationship	177
7.3.2.2. Definition of a Markov field	178
7.3.2.3. A Gibbs field is a Markov field	179
7.3.2.4. Hammersley-Clifford theorem	179
7.3.3. Posterior law of a GMRF	180
7.3.4. Gibbs-Markov models for images	181
7.3.4.1. Pixels with discrete values and label fields for classification	181
7.3.4.2. Gaussian GMRF	182
7.3.4.3. Edge variables, composite GMRF	183
7.3.4.4. Interactive edge variables	184
7.3.4.5. Non-Gaussian GMRFs	185
7.4. Statistical tools, stochastic sampling	185
7.4.1. Statistical tools	185
7.4.2. Stochastic sampling	188
7.4.2.1. Iterative sampling methods	189
7.4.2.2. Monte Carlo method of the MCMC kind	192
7.4.2.3. Simulated annealing	193
7.5. Conclusion	194
7.6. Bibliography	195

Chapter 8. Unsupervised Problems 197

Xavier DESCOMBES, Yves GOUSSARD

8.1. Introduction and statement of problem	197
8.2. Directly observed field	199
8.2.1. Likelihood properties	199
8.2.2. Optimization	200
8.2.2.1. Gradient descent	200
8.2.2.2. Importance sampling	200
8.2.3. Approximations	202

8.2.3.1. Encoding methods	202
8.2.3.2. Pseudo-likelihood	203
8.2.3.3. Mean field	204
8.3. Indirectly observed field	205
8.3.1. Statement of problem	205
8.3.2. EM algorithm	206
8.3.3. Application to estimation of the parameters of a GMRF	207
8.3.4. EM algorithm and gradient	208
8.3.5. Linear GMRF relative to hyperparameters	210
8.3.6. Extensions and approximations	212
8.3.6.1. Generalized maximum likelihood	212
8.3.6.2. Full Bayesian approach	213
8.4. Conclusion	215
8.5. Bibliography	216

PART IV. SOME APPLICATIONS 219

Chapter 9. Deconvolution Applied to Ultrasonic Non-destructive Evaluation 221

Stéphane GAUTIER, Frédéric CHAMPAGNAT, Jérôme IDIER

9.1. Introduction	221
9.2. Example of evaluation and difficulties of interpretation	222
9.2.1. Description of the part to be inspected	222
9.2.2. Evaluation principle	222
9.2.3. Evaluation results and interpretation	223
9.2.4. Help with interpretation by restoration of discontinuities	224
9.3. Definition of direct convolution model	225
9.4. Blind deconvolution	226
9.4.1. Overview of approaches for blind deconvolution	226
9.4.1.1. Predictive deconvolution	226
9.4.1.2. Minimum entropy deconvolution	228
9.4.1.3. Deconvolution by “multipulse” technique	228
9.4.1.4. Sequential estimation: estimation of the kernel, then the input	228
9.4.1.5. Joint estimation of kernel and input	229
9.4.2. DL2Hy/DBG deconvolution	230
9.4.2.1. Improved direct model	230
9.4.2.2. Prior information on double reflectivity	230
9.4.2.3. Double Bernoulli-Gaussian (DBG) deconvolution	230
9.4.2.4. Double hyperbolic (DL2Hy) deconvolution	231
9.4.2.5. Behavior of DL2Hy/DBG deconvolution methods	231
9.4.3. Blind DL2Hy/DBG deconvolution	232
9.5. Processing real data	232
9.5.1. Processing by blind deconvolution	233
9.5.2. Deconvolution with a measured wave	234

9.5.3. Comparison between DL2Hy and DBG	237
9.5.4. Summary	240
9.6. Conclusion	240
9.7. Bibliography	241

Chapter 10. Inversion in Optical Imaging through Atmospheric Turbulence 243

Laurent MUGNIER, Guy LE BESNERAIS, Serge MEIMON

10.1. Optical imaging through turbulence	243
10.1.1. Introduction	243
10.1.2. Image formation	244
10.1.2.1. Diffraction	244
10.1.2.2. Principle of optical interferometry	245
10.1.3. Effect of turbulence on image formation	246
10.1.3.1. Turbulence and phase	246
10.1.3.2. Long-exposure imaging	247
10.1.3.3. Short-exposure imaging	247
10.1.3.4. Case of a long-baseline interferometer	248
10.1.4. Imaging techniques	249
10.1.4.1. Speckle techniques	249
10.1.4.2. Deconvolution from wavefront sensing (DWFS)	250
10.1.4.3. Adaptive optics	251
10.1.4.4. Optical interferometry	251
10.2. Inversion approach and regularization criteria used	253
10.3. Measurement of aberrations	254
10.3.1. Introduction	254
10.3.2. Hartmann-Shack sensor	255
10.3.3. Phase retrieval and phase diversity	257
10.4. Myopic restoration in imaging	258
10.4.1. Motivation and noise statistic	258
10.4.2. Data processing in deconvolution from wavefront sensing	259
10.4.2.1. Conventional processing of short-exposure images	259
10.4.2.2. Myopic deconvolution of short-exposure images	260
10.4.2.3. Simulations	261
10.4.2.4. Experimental results	262
10.4.3. Restoration of images corrected by adaptive optics	263
10.4.3.1. Myopic deconvolution of images corrected by adaptive optics	263
10.4.3.2. Experimental results	265
10.4.4. Conclusion	267
10.5. Image reconstruction in optical interferometry (OI)	268
10.5.1. Observation model	268
10.5.2. Traditional Bayesian approach	271
10.5.3. Myopic modeling	272
10.5.4. Results	274

10.5.4.1. Processing of synthetic data	274
10.5.4.2. Processing of experimental data	276
10.6. Bibliography	277

Chapter 11. Spectral Characterization in Ultrasonic Doppler Velocimetry 285

Jean-François GIOVANNELLI, Alain HERMENT

11.1. Velocity measurement in medical imaging	285
11.1.1. Principle of velocity measurement in ultrasound imaging	286
11.1.2. Information carried by Doppler signals	286
11.1.3. Some characteristics and limitations	288
11.1.4. Data and problems treated	288
11.2. Adaptive spectral analysis	290
11.2.1. Least squares and traditional extensions	290
11.2.2. Long AR models – spectral smoothness – spatial continuity	291
11.2.2.1. Spatial regularity	291
11.2.2.2. Spectral smoothness	292
11.2.2.3. Regularized least squares	292
11.2.2.4. Optimization	293
11.2.3. Kalman smoothing	293
11.2.3.1. State and observation equations	293
11.2.3.2. Equivalence between parameterizations	294
11.2.4. Estimation of hyperparameters	294
11.2.5. Processing results and comparisons	296
11.2.5.1. Hyperparameter tuning	296
11.2.5.2. Qualitative comparison	296
11.3. Tracking spectral moments	297
11.3.1. Proposed method	298
11.3.1.1. Likelihood	298
11.3.1.2. Amplitudes: prior distribution and marginalization	298
11.3.1.3. Frequencies: prior law and posterior law	300
11.3.1.4. Viterbi algorithm	302
11.3.2. Likelihood of the hyperparameters	302
11.3.2.1. Forward-Backward algorithm	302
11.3.2.2. Likelihood gradient	303
11.3.3. Processing results and comparisons	304
11.3.3.1. Tuning the hyperparameters	304
11.3.3.2. Qualitative comparison	305
11.4. Conclusion	306
11.5. Bibliography	307

Chapter 12. Tomographic Reconstruction from Few Projections 311

Ali MOHAMMAD-DJAFARI, Jean-Marc DINTEN

12.1. Introduction	311
------------------------------	-----

12.2. Projection generation model	312
12.3. 2D analytical methods	313
12.4. 3D analytical methods	317
12.5. Limitations of analytical methods	317
12.6. Discrete approach to reconstruction	319
12.7. Choice of criterion and reconstruction methods	321
12.8. Reconstruction algorithms	323
12.8.1. Optimization algorithms for convex criteria	323
12.8.1.1. Gradient algorithms	324
12.8.1.2. SIRT (Simultaneous Iterative Relaxation Techniques)	325
12.8.1.3. ART (Algebraic Reconstruction Technique)	325
12.8.1.4. ART by blocks	326
12.8.1.5. ICD (Iterative Coordinate Descent) algorithms	326
12.8.1.6. Richardson-Lucy algorithm	326
12.8.2. Optimization or integration algorithms	327
12.9. Specific models for binary objects	328
12.10. Illustrations	328
12.10.1. 2D reconstruction	328
12.10.2. 3D reconstruction	329
12.11. Conclusions	331
12.12. Bibliography	332
Chapter 13. Diffraction Tomography	335
Hervé CARFANTAN, Ali MOHAMMAD-DJAFARI	
13.1. Introduction	335
13.2. Modeling the problem	336
13.2.1. Examples of diffraction tomography applications	336
13.2.1.1. Microwave imaging	337
13.2.1.2. Non-destructive evaluation of conducting materials using eddy currents	337
13.2.1.3. Geophysical exploration	338
13.2.2. Modeling the direct problem	338
13.2.2.1. Equations of propagation in an inhomogeneous medium	338
13.2.2.2. Integral modeling of the direct problem	339
13.3. Discretization of the direct problem	340
13.3.1. Choice of algebraic framework	340
13.3.2. Method of moments	341
13.3.3. Discretization by the method of moments	342
13.4. Construction of criteria for solving the inverse problem	343
13.4.1. First formulation: estimation of \mathbf{x}	344
13.4.2. Second formulation: simultaneous estimation of \mathbf{x} and ϕ	345
13.4.3. Properties of the criteria	347
13.5. Solving the inverse problem	347

13.5.1. Successive linearizations	348
13.5.1.1. Approximations	348
13.5.1.2. Regularization	349
13.5.1.3. Interpretation	349
13.5.2. Joint minimization	350
13.5.3. Minimizing MAP criterion	351
13.6. Conclusion	353
13.7. Bibliography	354
Chapter 14. Imaging from Low-intensity Data	357
Ken SAUER, Jean-Baptiste THIBAUT	
14.1. Introduction	357
14.2. Statistical properties of common low-intensity image data	359
14.2.1. Likelihood functions and limiting behavior	359
14.2.2. Purely Poisson measurements	360
14.2.3. Inclusion of background counting noise	362
14.2.4. Compound noise models with Poisson information	362
14.3. Quantum-limited measurements in inverse problems	363
14.3.1. Maximum likelihood properties	363
14.3.2. Bayesian estimation	366
14.4. Implementation and calculation of Bayesian estimates	368
14.4.1. Implementation for pure Poisson model	368
14.4.2. Bayesian implementation for a compound data model	370
14.5. Conclusion	372
14.6. Bibliography	372
List of Authors	375
Index	377

Introduction

When a physical quantity is not directly accessible for measurement, it is common to proceed by observing other quantities that are connected with it by physical laws. The notion of an *inverse problem* corresponds to the idea of inverting these physical laws to gain indirect access to the quantity we are interested in.

For example, in electromagnetism, calculating the electric field induced by a known distribution of electric charges is a *direct problem*, i.e., a problem posed in “the natural direction” of physics as we are used to practising and controlling it. Deducing the distribution of the electric charges from measurements of the field is, on the other hand, an *inverse problem*.

Similarly, in signal processing, modeling a transmission channel that introduces distortion, interference and parasitic signals corresponds to solving a direct problem. Reconstructing the shape of a signal input to the channel from measurements made at the output is an inverse problem.

The situation where the quantity of interest is directly accessible for measurement is obviously more favorable. Nevertheless, *direct measurement* does not signify *perfect measurement*: the *instrumental response* of a piece of measuring apparatus and the various error sources connected with the observation process (systematic error, fluctuations connected with the physical sensors or electronic components, quantization, etc.) are degradations that can also be encompassed in the question of inversion.

When all is said and done, the concept of inverse problems underlies the processing of experimental data in its broadest sense. In the experience of the authors of this book, making it explicit that a data processing chain actually carries out an inversion is also often a very worthwhile exercise. It brings to light *ad hoc* hypotheses and arbitrary

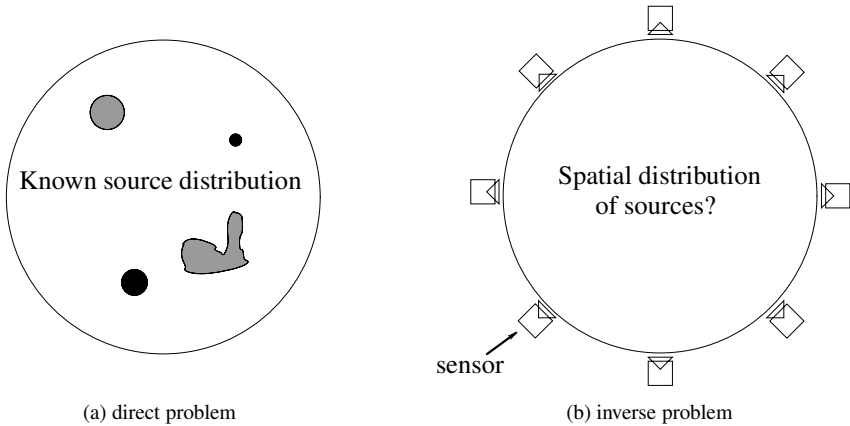


Figure 1. A simple example of the direct problem/inverse problem pair in electromagnetism: the direct problem consists of finding the field at the circular boundary of the domain from the distribution of the sources; the inverse problem consists of deducing the distribution of the sources from measurements of the field at the boundary

constraints, and provides a rational, modular framework for designing data processing methods and analyzing their efficiency and their faults.

When approached with no special precautions, the inversion problems we meet with in practice, unlike direct problems, have a nasty tendency to be “naturally unstable”: if there are errors, however tiny, on the data, the behavior of “naive” inversion methods is not robust.

Let us take the example of *inverse filtering* (or *deconvolution*), a classic in signal processing. Figure 2 proposes two digital experiments:

- The first consists of inverting a discrete convolution relationship, $y = h \star x$. The input signal x is triangular, of length $M = 101$ (Figure 2a), and the impulse response (IR) h is a discretized, truncated Gaussian of length $L = 31$ (Figure 2b). The output y , of length $N = M + L - 1 = 131$, is calculated using the Matlab language in the form `y=conv(h,x)` (Figure 2c). Matlab also offers a deconvolution method (by polynomial division) that, here, faithfully gives x again from y , in the form `deconv(y,h)` (Figure 2d).

- Now let us suppose that output y was measured imperfectly, e.g. with no measurement error but uniformly quantized to 10 bits, i.e., 1,024 levels: z is the quantized output (Figure 2e), calculated in Matlab in the form `z=round(y*2^10)/2^10`. The difference between y and z is imperceptible; and yet, `deconv(z,h)` (Figure 2f) is very different from x . Note in passing that the on-line help provided by Matlab (version 7.5) for `deconv` contains no warning of the unstable nature of this operation.

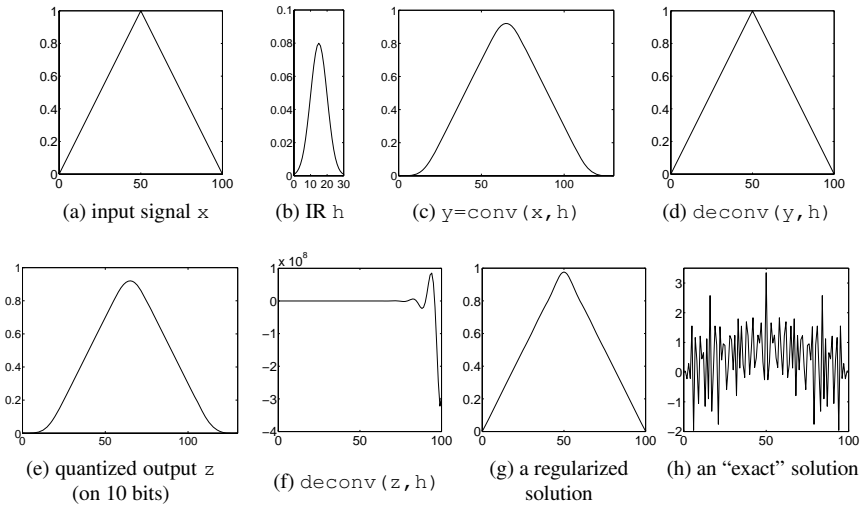


Figure 2. An example of a “naturally unstable” inverse problem: deconvolution. A tiny difference between z and $\text{conv}(x, h)$ is enough to make $\text{deconv}(z, h)$ very different from x . Like “conv”, “deconv” is an instruction in Matlab, version 7.5; it uses a non-regularized deconvolution method which is thus not robust. In comparison, (g) corresponds to a robust solution, obtained by Tikhonov regularization. Finally, (h) is an “exact” solution in the same way as x , in the sense that it reproduces z exactly by quantized convolution

In the early 20th century, Hadamard characterized these unstable problems mathematically, qualifying them as *ill posed*, in the sense that they did not lend themselves to being satisfactorily solved mathematically (and physically) [HAD 01]. Chapter 1 of this book develops the notion of ill posed problems and makes the non-robust behavior of “naive” inversion methods mathematically explicit.

In the 1960s, the Russian mathematician Tikhonov laid down the theoretical basis of modern inversion methods by introducing the concept of *regularized solutions* [TIK 63]. These solutions result from two ingredients being brought together. They are faithful to the data but this ingredient is not discriminating enough if the problem is ill posed. Among other solutions that are faithful to the data, they are the most regular, in a pragmatic sense that depends on the context. Tikhonov formalizes this trade-off between fidelity to the data and regularity by defining regularized solutions as those that minimize a composite criterion. He shows that the problem thus reformulated is *well-posed*. The principle of regularization in Tikhonov’s sense is one of the main subjects of Chapter 2.

Figure 2g illustrates the use of a method, regularized in Tikhonov’s sense, that reproduces input x very acceptably from imperfect data z . The calculation of this

type of solution is dealt with in Chapters 3 and 4. The only imperfection visible on Figure 2g concerns the point of the triangle, which is slightly blunt. This is a logical effect of the regularity imposed uniformly on a signal which is, in fact, locally irregular at its midpoint (its derivative is discontinuous). The case of signals or images that are globally regular but have localized irregularities is a very important one in practice. It is treated in the context of deconvolution in Chapter 5 for irregularities that are “bright spots” and in Chapter 6 for irregularities that are borders between homogeneous areas. The latter situation is precisely the one we have in Figure 2.

The problem of Figure 2 could also be solved by positioning straight lines through the minimization of a least squares criterion. A *parametric* approach of this kind, which regularizes the problem by *dimension control* (Chapter 2, section 2.1.1), can be considered as the oldest of the inversion methods as it was invented around the same time as the least squares method in the late 18th century. One of the first times the principle was put into practice was when Gauss estimated the coefficient of ellipticity of the Earth from arc length measurements, having modeled the Earth’s profile in the form of an ellipse [STI 81]. This was indeed a case of an inverse problem being solved by a parametric approach, even though the concept actually appeared much later.

Finally, we could think that the inversion would naturally become stable if the direct problem could be modeled with no errors. If this were true, it is a more detailed description of the direct problem that would lead to stabilization of the inversion. Let us take the example of Figure 2, for which the exact mathematical relation linking x and z includes quantization: the inversion of this exact relation remains unstable. In fact, x is only one of an infinite number of solutions, some of which stay remarkably far from x ; Figure 2h is an example.

Figure 2 is instructive but simplistic. In a more realistic situation, attaining a description of a direct problem – including the measuring system – in a mathematically perfect form is, in any case, more than we can hope for. In the inversion field, it is widely accepted that a credible inverse method must possess a minimum of robustness with respect to imperfect modeling of the direct problem. Adding pseudo-random noise to simulated data is a way of testing this robustness. Testing inversion only with “exact” simulated data is sometimes called the *inverse crime*.

Regularization in the Tikhonov sense, dimensionality control, adoption of such or such a parametric model, etc., there is no universal way of stabilizing an ill-posed inversion problem. The regularity of the solution must be defined case by case, in a form that may therefore appear subjective. Due to this, the concept of regularization is sometimes criticized or misunderstood. In fact, regularization is part of an application-oriented approach: it is not a question of inverting abstract problems that can be characterized by an *input-output* equation, but of solving *real problems*, where there is always advantage to be drawn from a few general characteristics of the quantity we are interested in, which may have been neglected in the initial formulation.

The practical success of Tikhonov's regularized approach and its subsequent evolutions has demonstrated that the approach is well founded. It is now accepted that an ill-posed inverse problem cannot be satisfactorily solved without some prior information, and this prior information is often qualitative or partial. For example, in an image restoration problem, it is desirable to take into consideration the fact that an image is generally composed of *homogeneous* regions, but this characteristic is qualitative and does not directly correspond to a mathematical model.

The encoding of uncertain or partial information can be envisaged within a *Bayesian probabilistic* framework. Work published a considerable time ago [FRA 70] showed that Tikhonov's contribution could be interpreted in this framework. For direct problems formulated deterministically, the handling of probabilistic rules for the inversion gives rise to comprehension difficulties. It has to be understood that these probabilistic rules are inference rules: they enable states of knowledge to be quantified and their evolution, through measurements, is itself uncertain because of errors. So we are not judging the fundamentally deterministic or random nature of the observed phenomena or even of the measurement errors. In this respect, Jaynes' work, brought together in [JAY 03], provides a reference for understanding the Bayesian approach in the data processing field. Chapter 3 of this book takes its inspiration from this work.

Modern methods for solving inverse problems have been arousing increasing interest since their beginnings in the 1960s. The question of inversion concerns a variety of sectors of activity, such as Earth and space sciences, meteorology, the aerospace industry, medical imaging, the nuclear electricity industry and civil engineering. Added up over all these sectors, its scientific and economic impact is enormous.

As far as the structure of inverse problems is concerned, very different fields may have very similar needs. However, the compartmentalization of scientific disciplines makes it difficult for ideas and methods to circulate. In this respect, it is the role of the signal processing community to respond to needs common to other disciplines in terms of data processing methods and algorithms. It was with this in mind that this book was written. Its 14 chapters are grouped together in four parts.

Part I is devoted to introducing the problems and the basic inversion tools and is the most abstract. It comprises three chapters:

- Chapter 1 introduces the problem of inversion as a whole, in a structured mathematical framework. It gives the characteristics of inverse problems posed in a continuous or discrete framework, and of ill-posed problems. It introduces the ideas of the pseudo-solutions and generalized inverse;
- Chapter 2 introduces the essential characteristics of regularization theory, and describes Tikhonov's approach and its subsequent evolutions together with other approaches to regularization. Reminders are then given of methods for minimizing criteria, followed by techniques for estimating the regularization parameter, so that an automatic choice can be made for the trade-off between fidelity to data and regularity;

– Chapter 3 deals with solving inverse problems in the framework of statistical inference. It becomes apparent that the conventional estimation technique, known as *maximum likelihood*, corresponds to a non-regularized solution, whereas a Tikhonov regularized approach finds a natural interpretation in the framework of Bayesian estimation. A number of questions are then re-examined in this context: the automatic choice of parameters, and the building of models and criteria. The end of the chapter is devoted to the Gaussian linear framework, which constitutes a fundamental special case.

Part II is made up of Chapters 4, 5 and 6, and is entirely given over to deconvolution, as a case that is very widespread in practice and also as a very instructive case where many of the tools introduced in these chapters can be adapted to the inversion of problems that are structured differently:

– Chapter 4 deals with deconvolution methods yielding solutions that are linear functions of the data. It first studies the general properties of the solutions, then the various algorithm structures that enable them to be calculated. Traditional signal processing tools, such as Wiener and Kalman filters, figure among these structures;

– Chapter 5 looks at the more specific problem of deconvolution when the signal of interest is a series of pulses. This situation is very common in numerous domains such as non-destructive evaluation and medical imaging. Taking the pulse character of the input signal into account leads us to two classes of nonlinear solutions according to the data. One follows a *detection-estimation* approach and the other uses the minimization of convex criteria and *robust estimation*;

– Chapter 6 is devoted to deconvolution when the unknown signal is “regular almost everywhere” and, in particular, takes this characteristic into consideration for images rather than monovariate signals. As in the previous chapter, we find two classes of solutions, according to whether the problem is approached in terms of detection of edges or as a problem of image restoration in robust form.

Part III groups together two chapters introducing “advanced tools” specific to the Bayesian framework presented in Chapter 3:

– Chapter 7 is concerned with imaging from a probabilistic point of view. The composite criteria of Chapter 6 are reinterpreted in this framework, which leads us to the Gibbs-Markov models. Various sub-classes are introduced as models for images. We next look at the statistical aspects connected with calculating the estimators and evaluating their performance. Finally, the principle of iterative methods for random sampling of the Gibbs-Markov models is presented;

– Chapter 8 is entirely devoted to the problem of *non-supervised* inversion, i.e., inversion without a regularization parameter fixed by the user. This question, already mentioned in Chapters 2 and 3, is of considerable theoretical and practical interest but brings together several types of methodological and algorithmic difficulties. Chapter 8 studies the case of a linear Markov penalization function with respect to the parameters

to be estimated and proposes, in particular, deterministic or stochastic techniques for implementing *maximum likelihood*, exact or approximate estimators.

Part IV, in six chapters, presents some inverse problems in their applications. This is by no means a complete review of all the domains involving inversion; several important fields such as heat, mechanics and geophysics are not covered¹. For the applications that are mentioned, we do not give an overall synthesis of the inversion problems encountered but rather some typical examples chosen by the authors as concrete illustrations of how regularized solutions are implemented in a particular domain. This last part also contains some important methodology extensions – myopic deconvolution (Chapters 9 and 10), Fourier synthesis (Chapter 10), spectral estimation and handling of *hidden Markov chains* (Chapter 11), and tomography problem solving (Chapters 12, 13 and 14):

- Chapter 9 concerns *industrial non-destructive evaluation using ultrasound*. It compares the implementation of the spike train deconvolution methods presented in Chapter 5. The problem of an impulse response that is poorly known or that introduces deformation is specifically studied. The result of this is some extended versions of the algorithms looked at in Chapter 5;

- Chapter 10 is about the inversion problems encountered in optical imaging in astronomy and, more specifically, for ground-based telescopes. In this case, atmospheric turbulence considerably reduces the resolution of the images. Various configurations intended to limit this degradation are considered. On the one hand, it is possible to approach the image restoration problem through myopic deconvolution; on the other, the effects of turbulence can be partially compensated by a technique known as *adaptive optics*, where the deconvolution of the images thus acquired remains a helpful step. The end of Chapter 10 is devoted to optical interferometry, which leads to a Fourier synthesis problem complicated by aberration due to atmospheric turbulence;

- Chapter 11 is devoted to *Doppler ultrasound velocimetry*, an imaging technique that is widespread in medicine. Two data inversion problems are seen to arise: *time-frequency analysis* and *frequency tracking*. These are *spectral characterization* problems that are particularly difficult for two reasons: firstly, the number of observed data points is very small and, secondly, the Shannon sampling conditions are not always respected. Chapter 11 covers these two problems in the regularization framework in order to compensate, at least partially, for the lack of information in the data;

- Chapter 12 considers the problem of reconstruction in X-ray tomography using a small number of projections. The Radon transform is introduced, and after a brief reminder of the various conventional approaches for its inversion, the algebraic and probabilistic methods are developed more specifically. In fact these are the only methods that can be used effectively in cases where the projections are limited in number and contain noise;

1. See [BEC 85, BUI 94, DES 90] as respective entry points to these fields.

– Chapter 13 looks at how the Bayesian approach can be used to solve the problem of diffraction tomography. For this type of problem, the measurements collected are the waves scattered by an object and depend nonlinearly on the physical parameters we are trying to image. This chapter deals with diffraction tomography without the usual linear approximations, whose domains of validity do not cover all the situations encountered in practice. The wave propagation equations provide a integral direct model in the form of two coupled equations. These are discretized by the *method of moments*. The Bayesian approach then allows the inversion to be approached by minimizing a penalized criterion. However, the nonlinearity of the direct model leads to non-convexity of this criterion. In particularly difficult situations where local minima exist, the use of global optimization techniques is recommended;

– Chapter 14 studies, in the framework of medical imaging techniques such as positron emission tomography, the situations in which the corpuscular character of the counting measurements needs to be taken into account. Poisson's law serves as the reference statistical distribution here to define the likelihood of the observations. There are also composite cases, in which Gaussian noise is added to the Poisson variables. It is sometimes possible to approximate the Poisson likelihood by a Gaussian law. When this is not the case, the properties of the log-likelihood criterion are studied and algorithms are put forward for various situations: emission or transmission tomography and composite cases.

Bibliography

- [BEC 85] BECK J., BLACKWELL B., ST. CLAIR C., *Inverse Heat Conduction, Ill-posed Problems*, Wiley Interscience, New York, NY, 1985.
- [BUI 94] BUI H. D., *Inverse Problems in the Mechanics of Materials: An Introduction*, CRC Press, Boca Raton, FL, 1994.
- [DES 90] DESAUBIES Y., TARANTOLA A., ZINN-JUSTIN J., Eds., *Oceanographic and Geophysical Tomography*, Amsterdam, The Netherlands, Elsevier Science Publishers (North-Holland), 1990.
- [FRA 70] FRANKLIN J. N., "Well-posed stochastic extensions of ill-posed linear problems", *J. Math. Anal. Appl.*, vol. 31, p. 682-716, 1970.
- [HAD 01] HADAMARD J., "Sur les problèmes aux dérivées partielles et leur signification physique", *Princeton University Bull.*, vol. 13, 1901.
- [JAY 03] JAYNES E. T., *Probability Theory – The Logic of Science*, Cambridge University Press, Apr. 2003.
- [STI 81] STIGLER S. M., "Gauss and the invention of least squares", *Annals Statist.*, vol. 9, num. 3, p. 465-474, 1981.
- [TIK 63] TIKHONOV A., "Regularization of incorrectly posed problems", *Soviet. Math. Dokl.*, vol. 4, p. 1624-1627, 1963.

PART I

Fundamental Problems and Tools

Chapter 1

Inverse Problems, Ill-posed Problems

1.1. Introduction

In many fields of applied physics, such as optics, radar, heat, spectroscopy, geophysics, acoustics, radioastronomy, non-destructive evaluation, biomedical engineering, instrumentation and imaging in general, we are faced with the problem of determining the spatial distribution of a scalar or vector quantity – we often talk about an *object* – from direct measurements – called an *image* – or indirect measurements – called *projections* in the case of tomography, for example – of this object. Solving such imaging problems can habitually be broken down into three stages [HER 87, KAK 88]:

- a *direct problem* where, knowing the object and the observation mechanism, we establish a mathematical description of the data observed. This model needs to be accurate enough to provide a correct description of the physical observation phenomenon and yet simple enough to lend itself to subsequent digital processing;
- an *instrumentation problem* in which the most informative data possible must be acquired so that the imaging problem can be solved in the best conditions;
- an *inverse problem* where the object has to be estimated from the preceding model and data.

Obtaining a good estimate of the object obviously requires these three sub-problems to be studied in a coordinated way. However, the characteristic that these image reconstruction or restoration problems have in common is that they are often ill-posed or ill-conditioned. Higher level problems that are found in computer vision, such as

image segmentation, optical flow processing and shape reconstruction from shading, are also inverse problems and suffer from the same difficulties [AND 77, BER 88, MAR 87]. In the same way, a problem such as spectral analysis, which has similarities with the Fourier synthesis used in radio-astronomy, for example, and which is not usually treated as an inverse problem, can gain from being approached this way, as we will see later.

Schematically, there are two broad communities that are interested in these inverse problems from a methodological point of view:

- the *mathematical physics* community, with the seminal works of Phillips, Twomey and Tikhonov in the 1960s [PHI 62, TIK 63, TWO 62]. Sabatier was one of the pioneers in France [SAB 78]. A representative journal is *Inverse Problems*;

- the *statistical data processing* community, which can be linked to the work of Franklin in the late 1960s [FRA 70], although the ideas involved – the basis of Wiener filtering – had been bubbling beneath the surface in many works for several years [FOS 61]. The Geman brothers gave a major boost to image processing about twenty years ago [GEM 84] A representative journal is *IEEE Transactions on Image Processing*.

A very rough distinction can be made between these two communities by saying that the former deals with the problem in an infinite dimension, with the questions of existence, uniqueness and stability, which become very complicated for nonlinear direct problems, and solves it numerically in finite dimensions, while the latter starts with a problem for which the discretization has already been performed and is not called into question, and takes advantage of the finite nature of the problem to introduce prior information built up from probabilistic models.

In this chapter, we propose to use a basic example to point out the difficulties that arise when we try to solve these inverse problems.

1.2. Basic example

We will now illustrate the basic concepts introduced in this chapter by an artificial example that mixes the essential characteristics of several types of inverse problems.

We are looking for a *spectrum*, the square of the modulus of a function $\widehat{x}(\nu)$, $\nu \in \mathbb{R}$ but, because of the experimental constraints, we only have access to the dual domain of the variable ν , through the function $x(t)$ of which $\widehat{x}(\nu)$ is the Fourier transform (FT). What is more, imperfections in the apparatus mean that the function $x(t)$ is only observable as weighted by a “window” $h(t)$, which gives the observable function $y(t)$:

$$y(t) = h(t) x(t). \quad (1.1)$$

To make our ideas clear, let us think of a visible optical interferometry device like that by Michelson. To have access to the emission spectrum of the light source, we measure an energy flux as a function of the phase difference between two optical paths. The interferogram obtained is, ignoring the additional constant, the Fourier transform of the function we are looking for but the limitations of the apparatus make the interferogram observable only in a limited area of space, which is equivalent to its being modulated by a weighting function $h(t)$. This is assumed to be known but the experimental data that is actually available is made up of a finite number of regularly-spaced samples of the function $y(t)$, which inevitably contain measuring errors that we assume to be additive. If we take a unit sampling step, we can write:

$$y_n = h_n x_n + b_n, \quad n = 1, \dots, N, \quad (1.2)$$

where y_n designates the available data, h_n and x_n the samples of the functions $h(t)$ and $x(t)$ respectively, and b_n the measurement “noise”. This is a special case of a system of linear equations of the form:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b} \quad (1.3)$$

that we will find repeatedly throughout this book. Here we have a diagonal matrix \mathbf{A} which, at first glance, appears to be a simple situation.

A first difficulty appears, however, independently of the presence of the weighting $h(t)$: the discrete nature of the data means that we only have information on $\widehat{x}_1(\nu)$, $\nu \in [0, 1]$, a 1-periodic function deduced from $\widehat{x}(\nu)$ by the periodization due to the sampling, since we have:

$$x_n = \int_0^1 \widehat{x}_1(\nu) \exp\{2j\pi\nu n\} d\nu. \quad (1.4)$$

The samples x_n are in fact the Fourier series development coefficients of \widehat{x}_1 . To have any hope of accessing \widehat{x} , it is necessary for $\widehat{x}(\nu)$ to have limited support and for the sampling step to be such that there is no aliasing. Observation model (1.2) can thus be written indifferently:

$$y_n = \int_0^1 \widehat{h} \star \widehat{x}_1(\nu) \exp\{2j\pi\nu n\} d\nu + b_n, \quad (1.5)$$

where $\widehat{h}(\nu)$ is the FT of $h(t)$. The presence of this convolution core expresses the loss of resolving power of the instrument due to the weighting by $h(t)$.

A simulated example is presented in Figure 1.1. Signal $x(t)$ is composed of three sine waves, the spectrum of which is marked by the circles in Figure 1.1a. Two have frequencies that are close together (relative frequency difference less than 0.008). Response $\widehat{h}(\nu)$ is a Gaussian of standard deviation $\sigma_{\widehat{h}} = 0.0094$ intentionally chosen

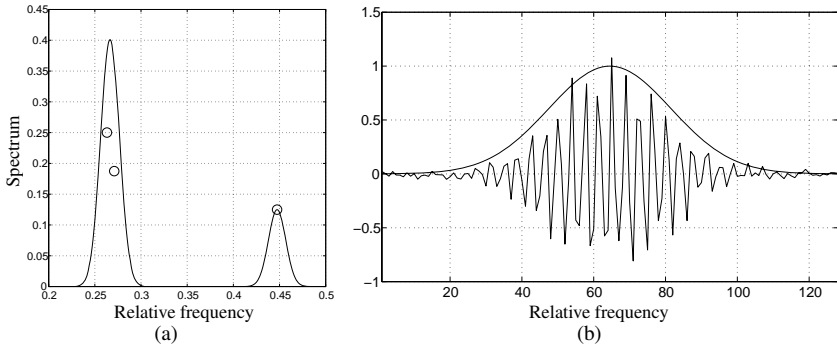


Figure 1.1. (a) Spectrum \hat{x} of a linear combination of three sine waves, indicated by circles, and $\hat{h} \star \hat{x}$, where \hat{h} is a Gaussian of standard deviation close to 0.01 in relative frequency; (b) 128 data points y simulating an interferogram that contains noise and is quantified, corresponding to model (1.5)

high to point out clearly the difficulties of inversion. The “non-resolved” spectrum $\hat{h} \star \hat{x}$ is also represented in Figure 1.1a. Figure 1.1b superposes the $N = 128$ simulated data y_n and the series of weighting coefficients h_n , which also have a Gaussian form (of standard deviation $1/2\pi\sigma_{\hat{h}} = 17$).

A second difficulty comes from the impossibility of inverting equation (1.5) in a mathematically exact way, i.e., of finding the “true” function \hat{x}_1 among other candidate functions, even in the absence of noise. Consider, for example, the FT $\hat{\hat{x}}$ of a stable series $\{\hat{x}_n\}_{\mathbb{Z}}$ such that:

$$\hat{\hat{x}}_n = y_n/h_n \quad \text{if } n \in \{1, \dots, N\} \quad \text{and} \quad h_n \neq 0. \quad (1.6)$$

Since this series is only defined for N values at most, there is an infinite number of solutions $\hat{\hat{x}}$ that satisfy constraints (1.6), and are equivalent considering the data. The problem is therefore *indeterminate*. In this respect, the periodogram of the data:

$$\Gamma(\nu) \triangleq \frac{1}{N} \left| \sum_{n=1}^N y_n \exp \{-2j\pi\nu n\} \right|^2, \quad \nu \in [0, 1], \quad (1.7)$$

calculable by fast discrete FT on a fine, regularly spaced grid, is a particular solution for h_n close to 1 (i.e., \hat{h} close to a Dirac). It is obtained by extending $\hat{\hat{x}}_n = y_n$ with zeros on either side of the observation window.

The small number of data points and the spread of the instrument response \hat{h} give the periodogram very low resolving power (see Figure 1.2a, curve (P₁)). We can try to get around the need to have \hat{h} by calculating the periodogram associated with y_n/h_n

or, in other words, by making a spectral estimator \widehat{x} from a time series extrapolating y_n/h_n with zeros. It is also worth noting that this is none other than the trivial solution to the problem of finding a series $\{x_n\}_{\mathbb{Z}}$ that is stable and has a minimal norm, and which minimizes the least squares criterion – even reducing it to zero in this case:

$$\sum_{n=1}^N (y_n - h_n x_n)^2.$$

The result is disappointing (see Figure 1.2a, curve (P_2)). In fact, this is not really surprising as the series y_n/h_n contains aberrant values at its extremities because of the measurement noise and quantification. These error terms, which are amplified by $1/h_n$ when h_n is small, make a contribution to the estimated spectrum that completely masks the peaks of the theoretical spectrum.

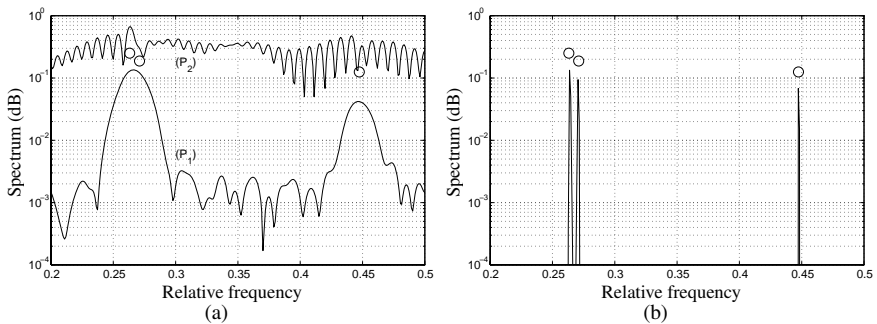


Figure 1.2. (a) Curve (P_1) is the periodogram of the data y_n represented in dB; the lack of resolution is a result of the lack of data but also of the spread response of the instrument. Curve (P_2) is the periodogram associated with y_n/h_n ; (b) spectral estimate obtained as the minimizer of criterion (1.8), calculated by approximation on a discrete grid of 1,024 points, for “well chosen” values of λ and τ

These negative results could lead us to think that the data is too poor to be used. This is not the case, as shown by the spectral estimate whose modulus is represented in Figure 1.2b, and which is obtained as the function \widehat{x} that minimizes the *regularized* criterion:

$$\sum_{n=1}^N (y_n - h_n x_n)^2 + \lambda \int_0^1 \sqrt{\tau^2 + |\widehat{x}(\nu)|^2} d\nu, \quad (1.8)$$

where x_n is connected to \widehat{x} by (1.4) for “well chosen” values of *hyperparameters* λ and τ . As the FT \widehat{x} is discretized on 1,024 points, this process is strictly equivalent to extrapolating the series of 128 observed x_n by 896 values that are not necessarily zero, unlike in the periodogram.

This example is typical of the difficulties that arise in the solving of numerous inverse problems [AND 77, BER 88, HER 87]. Certain conventional signal processing

tools prove to be unsuitable whereas others provide qualitatively and quantitatively exploitable solutions. The improvement obtained with this regularized criterion (1.8) is striking and several questions immediately come to mind: why do we need to penalize the least squares criterion in this way? How do we obtain the argument of the minimum of such a criterion? How do we choose the values of the hyperparameters that are part of it? It can be said that the main part of this book is devoted to just that: the construction and use of regularized criteria. However, it is important to understand the nature of the difficulties encountered during inversion before studying the regularized solutions that allow them to be solved.

1.3. Ill-posed problem

The aim of this section is to correct the false impression that the difficulties encountered in solving an inverse problem come from the discrete nature of the data and its finite amount and that, if we had access to a *continuum* of values, i.e., the function $y(t)$ in the example above, everything would be fine. Often unsuspected difficulties are already present at this level. They are proper to problems known as *ill-posed* problems. When the problem is inevitably discretized as in the previous example, some of these difficulties paradoxically disappear, but the problem most often remains *ill-conditioned*.

Hadamard has defined three conditions for a mathematical problem to be *well-posed* [AND 80, HAD 01, NAS 81, TIK 77] (by default, it will be called *ill-posed*):

- (a) for each item of data y in a defined class \mathcal{Y} , there exists a solution x in a stipulated class \mathcal{X} (*existence*);
- (b) the solution is unique in \mathcal{X} (*uniqueness*);
- (c) the dependence of x on y is continuous, i.e., when the error δy on data item y tends towards zero, the error δx induced on the solution x also tends towards zero (*continuity*).

The continuity requirement is connected to that of *stability* or *robustness* of the solution (with respect to the errors that inevitably occur on the data). Continuity is, however, a necessary but not a sufficient condition for robustness [COU 62]. A well-posed problem can be *ill-conditioned*, which makes its solution non-robust, as we shall see in section 1.5.

All the traditional problems of mathematical physics, such as the Dirichlet problem for elliptical equations or the Cauchy problem for hyperbolic equations, are well-posed in Hadamard's sense [AND 80]. However, the “inverse” problems obtained from “direct” problems by exchanging the roles of the solution and data are generally *not* well-posed.

The example of section 1.2 clearly comes into this category of ill-posed problems since, with a finite number of discrete data items, a solution $\widehat{x}(\nu)$ exists, but it is not unique. It is interesting to note that this same problem, before discretization, is a special case of the general problem of solving a Fredholm integral equation of the first kind:

$$y(s) = \int k(s, r) x(r) dr, \quad (1.9)$$

where $y(s)$, $x(r)$ and $k(s, r)$ are replaced by $y(t)$, $\widehat{x}(\nu)$ and $h(t) \exp \{2j\pi\nu t\}$ respectively. Later in this book we will find the same type of integral equations for other forms of kernel $k(s, r)$, for deconvolution, tomographic reconstruction and Fourier synthesis.

As the data is uncertain or noisy, we cannot hope to solve this equation exactly and the solution needs to be approached from a certain direction. The concept of distance between functions is thus a natural way of evaluating the quality of an approximation, which explains why x and y are often assumed to belong to Hilbert spaces. Problem (1.9) can thus be rewritten as:

$$y = Ax, \quad x \in \mathcal{X}, \quad y \in \mathcal{Y}, \quad (1.10)$$

where x and y are now elements of functional spaces of infinite dimension \mathcal{X} and \mathcal{Y} , respectively, and where $A: \mathcal{X} \rightarrow \mathcal{Y}$ is the linear operator corresponding to (1.9). The necessary and sufficient conditions for the existence, uniqueness and continuity of the solution can thus be written respectively [NAS 81]:

$$\mathcal{Y} = \text{Im } A, \quad \text{Ker } A = \{0\}, \quad \text{Im } A = \overline{\text{Im } A}, \quad (1.11)$$

where $\text{Im } A$ is the *image* of A (i.e., the set of y that are images of an $x \in \mathcal{X}$), $\text{Ker } A$ its *kernel* (i.e., the set of solutions to the equation $Ax = 0$) and $\overline{\text{Im } A}$ the *closure* of $\text{Im } A$ [BRE 83].

The manner in which conditions (1.11) are stated gives rise to several comments. On the one hand, $\mathcal{Y} = \text{Im } A$ implies $\text{Im } A = \overline{\text{Im } A}$ (a Hilbert space is closed upon itself). In other words, the very existence of a solution to problem (1.9) $\forall y \in \mathcal{Y}$ implies the continuity of this solution. *In contrast*, if the existence condition $\mathcal{Y} = \text{Im } A$ is not verified, the continuity condition seems to become pointless; in fact, it applies to *pseudo-solutions*, which will be defined in section 1.4.1 as minimizing the norm $\|Ax - y\|_{\mathcal{Y}}$ (without systematically reducing it to zero).

1.3.1. Case of discrete data

When the data is discrete, y is a vector of dimension N in a Euclidian space. Ignoring errors on the data, a linear inverse problem with discrete data can be stated

as follows. Given a set $\{F_n(x)\}_{n=1}^N$ of linear functionals defined on \mathcal{X} and a set $\{y_n\}_{n=1}^N$ of numbers, find a function $x \in \mathcal{X}$ such that:

$$y_n = F_n(x), \quad n = 1, \dots, N.$$

In particular, when functionals F_n are continuous on \mathcal{X} , Riesz theorem [BRE 83] states that functions ψ_1, \dots, ψ_N exist such that:

$$F_n(x) = \langle x, \psi_n \rangle_{\mathcal{X}},$$

where the notation $\langle \cdot, \cdot \rangle_{\mathcal{X}}$ designates the scalar product used in space \mathcal{X} . The example of equation (1.9) takes this form when $y(s)$ is measured on a finite number of points s_1, \dots, s_N , and \mathcal{X} is an L^2 space. In this case we have:

$$\psi_n(r) = k(s_n, r).$$

This problem is a particular case of that of equation (1.10) if we define an operator A of \mathcal{X} in \mathcal{Y} by the relation:

$$(Ax)_n = \langle x, \psi_n \rangle_{\mathcal{X}} \quad n = 1, \dots, N.$$

Operator A is not injective: $\text{Ker } A$ is the closed subspace of infinite dimension of all the functions x orthogonal to the subspace engendered by the functions ψ_n . Conversely, the image of A , $\text{Im } A$ is closed: $\text{Im } A$ is simply \mathcal{Y} when the functions ψ_n are linearly independent; otherwise it is a subspace of dimension $N' < N$. We thus see clearly why the example of section 1.2 is an ill-posed problem: the difficulty does not lie in a lack of continuity but in a lack of uniqueness.

1.3.2. Continuous case

Let us now assume that x and y belong to the same Hilbert space and that k is square integrable, a condition fulfilled by many imaging systems – it would be the situation if our example of section 1.2 was modified so that the function $\widehat{y}(\nu) = \widehat{h} \star \widehat{x}_1(\nu)$ was continuously observed. The direct problem is thus well-posed: a small error δx on the data entails a small error δy on the solution. This condition is not, however, fulfilled in the corresponding inverse problem, where it is object x that must be calculated from response y : $x = A^{-1}y$. In fact, when kernel k is square integrable – which would be the case for a Gaussian kernel in our example – the Riesz-Fréchet theorem indicates that operator A is bounded and *compact* [BRE 83]. However, the image of a compact operator is not closed (except in the degenerate case where its dimension is finite). This signifies that the inverse operator A^{-1} is not bounded, or stable, its image is not closed and the third of Hadamard's conditions is not satisfied for the inverse problem [NAS 81].

To get a better grasp of these abstract ideas, it is handy to use the spectral properties of compact operators in Hilbert spaces. The most remarkable property of these operators is that they can be *decomposed into singular values*, like matrices (the famous *singular value decomposition*, or SVD). The singular system of a compact operator is defined as the set of solutions of the coupled equations:

$$A u_n = \sigma_n v_n \quad \text{and} \quad A^* v_n = \sigma_n u_n, \quad (1.12)$$

where the *singular values* σ_n are positive numbers, where the *singular functions* u_n and v_n are elements of \mathcal{X} and \mathcal{Y} respectively, and where A^* is the *adjoint operator* of A , which exists since A is continuous and therefore such that: $\langle Ax, y \rangle_{\mathcal{Y}} = \langle x, A^* y \rangle_{\mathcal{X}}$ for any $x \in \mathcal{X}$ and $y \in \mathcal{Y}^1$.

When A is compact, it always possesses a singular system $\{u_n, v_n; \sigma_n\}$ with the following properties [NAS 81]:

- σ_n being ordered and counted with their multiplicity (which is finite): $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq \dots \geq 0$, σ_n tends towards 0 when $n \rightarrow \infty$ and either the limit is reached for $n = n_0$ (in which case operator A is degenerate), or it is not reached for any finite value of n ;

- functions u_n form an orthonormal basis of $(\text{Ker } A)^\perp$, the orthogonal complement of $\text{Ker } A$ in the decomposition: $\mathcal{X} = \text{Ker } A \oplus (\text{Ker } A)^\perp$ and the functions v_n form an orthonormal basis of $(\text{Ker } (A^*))^\perp$, i.e., $\overline{\text{Im } A}$, orthogonal complement of $\text{Ker } (A^*)$ in the decomposition $\mathcal{Y} = \text{Ker } (A^*) \oplus (\text{Ker } (A^*))^\perp$.

Let $E \subseteq \mathbb{N}$ be the set of indices n such that $\sigma_n \neq 0$. The Picard criterion [NAS 81] ensures that a function $y \in \mathcal{Y}$ is in $\text{Im } A$ if and only if:

$$y \in (\text{Ker } (A^*))^\perp \quad \text{and} \quad \sum_{n \in E} \sigma_n^{-2} \langle y, v_n \rangle^2 < +\infty. \quad (1.13)$$

For the second condition (1.13) to be satisfied, it is necessary, when operator A is not degenerate ($E \equiv \mathbb{N}$), for the components $\langle y, u_n \rangle$ of the development of image y on the set of eigenfunctions $\{v_n\}$ to tend towards zero faster than the eigenvalues σ_n^2 when $n \rightarrow \infty$. This strict condition has no reason to be satisfied by an arbitrary function of $(\text{Ker } (A^*))^\perp$. Note, however, that it is naturally satisfied if $y = Ax$ is the perfect image resulting from an object x of finite energy. The solution is thus written:

$$x = \sum_{n \in E} \sigma_n^{-1} \langle y, v_n \rangle u_n. \quad (1.14)$$

1. Note that the self-adjoint operator A^*A , which appears in the symmetrized problem $A^*y = A^*Ax$, verifies: $A^*Av_n = \sigma_n^2 v_n$. It is thus defined as non-negative since its eigenvalues are σ_n^2 (which are also those of AA^*). This property will be used in section 2.1.1.

However, this solution, when it exists, is *unstable*: a small additive perturbation $\delta y = \varepsilon v_N$, for example, on the perfect data y leads to a perturbation δx on the solution calculated with the data $y + \delta y$:

$$\delta x = \sigma_N^{-1} \langle \delta y, v_N \rangle u_N = \sigma_N^{-1} \varepsilon u_N. \quad (1.15)$$

The ratio $\|\delta x\| / \|\delta y\|$ equals σ_N^{-1} , which can be arbitrarily large. The inverse linear operator $A^{-1} : \mathcal{Y} \rightarrow \mathcal{X}$, defined by (1.14), is thus not *bounded* as it is not possible to find a constant C such that, for all $y \in \mathcal{Y}$, we have $\|A^{-1}y\|_{\mathcal{X}} \leq C \|y\|_{\mathcal{Y}}$, which is a necessary and sufficient condition for A^{-1} to be *continuous* [BRE 83]. The ill-posed nature of the problem stems this time from the lack of continuity and not from the lack of uniqueness.

The need for a deeper understanding of these problems that are not mathematically well-posed but are of great interest in engineering sciences is at the origin of two recent branches of analysis: *generalized inversion* theory [NAS 76], which is summarized below, and *regularization* theory, which will be the subject of the next chapter.

1.4. Generalized inversion

Let us suppose that the equation $Ax = 0$ has non-trivial solutions. The set $\text{Ker } A \neq \{0\}$ of these solutions is a closed subspace of \mathcal{X} . It is the set of “invisible objects” as they produce an image y that is zero. Let us also suppose that $\text{Im } A$ is a closed subspace of \mathcal{Y} . An example is provided by the integral operator corresponding to an ideal low-pass filter of cut-off pulsation Ω [BER 87]:

$$(Ax)(r) = \int_{-\infty}^{+\infty} \frac{\sin \Omega(r-r')}{\pi(r-r')} x(r') dr'. \quad (1.16)$$

If we choose $\mathcal{X} = \mathcal{Y} = L^2_{\mathbb{R}}$, the kernel is the set of all the functions x whose FT is zero in the band $[-\Omega, +\Omega]$, while the image of A is the set of functions having a limited band in the same interval, which is a closed subspace of $L^2_{\mathbb{R}}$.

A means of re-establishing the existence and the uniqueness of the solution in the above conditions is to redefine both the space \mathcal{X} of the solutions and the space \mathcal{Y} of the data. If we choose a new space \mathcal{X}' which is the set of all the functions orthogonal to $\text{Ker } A$ (in the case of equation (1.16), \mathcal{X}' is the set of functions with summable squares and band limited to the interval $[-\Omega, +\Omega]$), and if y is restrained to a new data space $\mathcal{Y}' = \text{Im } A$ (which is, once again, in the case of equation (1.16), the set of functions with summable squares and band limited to the interval $[-\Omega, +\Omega]$), thus, for any $y \in \mathcal{Y}'$, there exists a unique $x \in \mathcal{X}'$ such that $Ax = y$ (in our example (1.16), the solution is even trivial: $x = y$) and the new problem is thus well-posed.

It is often possible to choose the spaces \mathcal{X} and \mathcal{Y} so that the problem becomes well-posed but the practical interest of the choice is limited because it is generally the

intended application that imposes the appropriate spaces. Another means that could be envisaged is to change the notion of a solution itself.

1.4.1. *Pseudo-solutions*

Let us first consider the case where A is injective ($\text{Ker } A = \{0\}$) but not surjective ($\text{Im } A \neq \mathcal{Y}$). The set of functions x that are solutions of the variational problem:

$$x \in \mathcal{X} \text{ minimizes } \|Ax - y\|_{\mathcal{Y}}, \quad (1.17)$$

where $\|\cdot\|_{\mathcal{Y}}$ designates the norm in \mathcal{Y} , are called pseudosolutions or least squares solutions of the problem (1.10). If $\text{Im } A$ is closed, (1.17) always has a solution, but it is not unique if the kernel $\text{Ker } A$ is not trivial. When it is, as we assume here, it can be said that the well-posed character has been restored by reformulating the problem in the form (1.17).

By making the first variation of the function minimized in (1.17) zero, we obtain Euler's equation:

$$A^*Ax = A^*y, \quad (1.18)$$

which brings in the self-adjoint operator A^*A , the eigensystem of which can be deduced from the singular system of A .

1.4.2. *Generalized solutions*

Let us now consider the case where the uniqueness condition is not satisfied ($\text{Ker } A \neq \{0\}$, the problem is indeterminate). The set of solutions of (1.18) being a convex, closed subset of \mathcal{X} , it contains a single element with a minimal norm, noted x^\dagger or \hat{x}^{Gi} and called the *generalized solution* of (1.10). As x^\dagger is orthogonal to $\text{Ker } A$, this way of defining the solution is equivalent to choosing $\mathcal{X}' = (\text{Ker } A)^\perp$. In other words, the generalized solution is a least squares solution having the minimal norm among these solutions. As there is a single x^\dagger for every $y \in \mathcal{Y}$, a linear application A^\dagger of \mathcal{Y} in \mathcal{X} is defined by:

$$A^\dagger y = x^\dagger = \hat{x}^{\text{Gi}}. \quad (1.19)$$

The operator A^\dagger is called the *generalized inverse* of A and is continuous [NAS 76].

1.4.3. *Example*

To illustrate the idea of generalized inversion, let us go back to our example of section 1.2 and, first of all, neglect the weighting $h(t)$. To impose a unique solution in

the class of possible solutions prolonging the series of known x_n , we can choose the generalized inverse solution of the initial problem (1.10):

$$\widehat{x}^{\text{GI}}(\nu) = \arg \min_{\widehat{x}_1 \in L^2_{\mathbb{C}}[0,1]} \int_0^1 |\widehat{x}_1(\nu)|^2 d\nu \quad \text{subject to (s. t.) } x_n = y_n, \quad n = 1, \dots, N.$$

The Plancherel and Parseval relations show us that this is equivalent to finding coefficients:

$$\widehat{x}_n = \int_0^1 \widehat{\widehat{x}}(\nu) \exp\{2j\pi n\nu\} d\nu, \quad n \in \mathbb{Z},$$

such that:

$$\widehat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \ell^2_{\mathbb{C}}} \sum_{n \in \mathbb{Z}} |x_n|^2, \quad \text{s. t. } x_n = y_n, \quad n = 1, \dots, N.$$

The solution is trivial since the problem is separable:

$$\widehat{x}_n = \begin{cases} y_n & \text{if } n \in \{1, 2, \dots, N\}, \\ 0 & \text{otherwise,} \end{cases} \quad \implies \quad \widehat{\widehat{x}}^{\text{GI}}(\nu) = \sum_{n=1}^N y_n e^{-2j\pi n\nu}, \quad (1.20)$$

whose squared modulus, with just the difference of a coefficient, gives the *Schuster periodogram* of equation (1.7). The case of weighting by $h(t)$ is treated in the same way, by replacing y_n by y_n/h_n . It can thus be seen that the periodogram is the generalized inverse solution of a spectral analysis problem that is ill-posed because the number of data points is finite.

1.5. Discretization and conditioning

A first description of a direct problem often brings in functions of real variables (time, frequency, space variables, etc.), representing the physical quantities involved: quantities accessible for measurement and quantities of interest that are unknown. The analysis of the problem at this level of description has provided an explanation for the difficulties that arise during inversion, by situating us in functional spaces of infinite dimension. We have thus seen that, in the case of a direct problem described by an integral equation of the first kind, the inversion is often an ill-posed problem as it is unstable.

This analysis is, however, insufficient. The available experimental data are almost always composed of measurements of physical quantities accessible at a necessarily finite number of points in the domain of definition of their variables. They are thus naturally discrete and we group them together in the vector \mathbf{y} as we did in section 1.4 above. However, the unknown object is also discretized, either right from the start, or during the process of solution (as in the example of the periodogram above), by

decomposition over a finite number of functions. If these are basic elements of the space to which the object belongs, the decomposition is necessarily truncated. In imaging, for example, in the vast majority of cases, pixel indicators or cardinal sines are used as basic functions, according to whether the object is implicitly assumed to have a limited support or a limited spectrum. Basic wavelets or wavelet packets are also coming into use [STA 02, KAL 03]. The starting point is thus composed of a *model parametrized* by the vector \mathbf{x} of the decomposition coefficients, in other words by a set of exclusive hypotheses, each of which is indexed by the value of the coefficients. This hypothesis space is thus the set of possible values of these unknown parameters, $\mathcal{H} = \{x_i\}$. The choice of these basic functions obviously forms part of the inversion problem, even if it is not often touched upon.

In the discrete case (or more exactly the “discrete-discrete” case), the problem changes noticeably as \mathbf{x} and \mathbf{y} belong to spaces of finite dimensions and the linear operator A becomes a matrix \mathbf{A} . Equation (1.10) has a unique solution with minimal norm $\hat{\mathbf{x}}^{\text{Gi}} = \mathbf{A}^\dagger \mathbf{y}$ which depends continuously on \mathbf{y} since the generalized inverse \mathbf{A}^\dagger is then always bounded [NAS 76]. The problem is thus always well-posed in Hadamard’s sense. However, even in this framework, the inversion problem still has an unstable nature, this time from a numerical point of view. The spectral decomposition (1.15) is still valid, the only difference being that the number of singular values of matrix \mathbf{A} is now finite. These singular values can rarely be calculated explicitly [KLE 80]. From this point of view, the example in section 1.2 is not representative because, if we choose to decompose the solution over $M > N$ complex exponentials of the Fourier basis:

$$\hat{x}(\nu) = \sum_{m=1}^M x_m \exp \{-2j\pi m\nu\},$$

model (1.2) can be written $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$ in a matrix-vector notation, where \mathbf{A} is a rectangular $N \times M$ matrix composed of the diagonal matrix $\text{diag}\{h_n\}$, juxtaposed with the zero matrix of size $N \times (M - N)$. Its singular values are thus $\sigma_n = h_n$ for $n = 1, 2, \dots, N$ and $\sigma_n = 0$ otherwise. Even if we exclude all the zero singular values, as by using \mathbf{A}^\dagger , there are always some singular values close to zero with the weighting $h(t)$ of our example. Matrix \mathbf{A} is thus *ill-conditioned*. The coefficients $\sigma_n^{-1} \langle \delta \mathbf{y}, \mathbf{u}_n \rangle$ in equation (1.15) become very large for the σ_n that are close to zero, even if $\delta \mathbf{y}$ is small.

Generally speaking, whether we have the discrete case or not, let us assume that $\text{Im } A$ is closed so that the generalized inverse A^\dagger exists $\forall y \in \mathcal{Y}$ (and is continuous). Let us designate an error on the data y as δy and the error induced on the generalized inverse solution x^\dagger as δx^\dagger . The linearity of (1.19) leads to $\delta x^\dagger = A^\dagger \delta y$, which implies

$$\|\delta x^\dagger\|_{\mathcal{X}} \leq \|A^\dagger\| \|\delta y\|_{\mathcal{Y}},$$

where $\|A^\dagger\|$ designates the norm of the continuous operator A^\dagger , that is to say the quantity: $\sup_{y \in \mathcal{Y}} \|A^\dagger y\|_{\mathcal{X}} / \|y\|_{\mathcal{Y}}$ [BRE 83]. In a similar way, (1.10) implies:

$$\|y\|_{\mathcal{Y}} \leq \|A\| \|x^\dagger\|_{\mathcal{X}},$$

where $\|A\| = \sup_{x \in \mathcal{X}} \|Ax\|_{\mathcal{Y}} / \|x\|_{\mathcal{X}}$. By combining these two relations, we obtain the inequality:

$$\frac{\|\delta x^\dagger\|_{\mathcal{X}}}{\|x^\dagger\|_{\mathcal{X}}} \leq \|A\| \|A^\dagger\| \frac{\|\delta y\|_{\mathcal{Y}}}{\|y\|_{\mathcal{Y}}}. \quad (1.21)$$

It is important to note that this inequality is precise in a certain sense. When A is a matrix of dimensions $(N \times M)$ or corresponds to an inverse problem with discrete data, the inequality can become an equality for certain $(y, \delta y)$ pairs. When A is an operator on spaces of infinite dimension, it can only be established that the left hand side of inequality (1.21) can be arbitrarily close to the right hand side. The quantity:

$$c = \|A\| \|A^\dagger\| \geq 1 \quad (1.22)$$

is called the *condition number* of the problem. When c is close to one, the problem is said to be *well-conditioned*, whereas when it is considerably larger than one, the problem is said to be *ill-conditioned*.

In practice, it is useful to have an estimate of the condition number, which gives an idea of the numerical stability of the problem. When $A = \mathbf{A}$ is a matrix of dimensions $(N \times M)$, $\|\mathbf{A}\|$ is the square root of the largest of the eigenvalues of the positive semidefinite symmetric matrix $\mathbf{A}^* \mathbf{A}$, of dimensions $(M \times M)$ (the positive eigenvalues of this matrix coincide with those of the matrix $\mathbf{A} \mathbf{A}^*$) and $\|\mathbf{A}^\dagger\|$ is the inverse of the square root of the smallest of these eigenvalues:

$$c = \sqrt{\lambda_{\max} / \lambda_{\min}}.$$

In our example in section 1.2, we obtain $c = |h|_{\max} / |h|_{\min}$ and we understand why the weighting by $h(t)$ can degrade the conditioning of the generalized inversion problem which is otherwise well-posed.

1.6. Conclusion

To sum up the above, when we have a simple situation where we are dealing with a direct, linear problem in an infinite dimension, bringing in an operator $A: \mathcal{X} \rightarrow \mathcal{Y}$ defined in the Hilbert spaces \mathcal{X} and \mathcal{Y} , we have three main situations:

- if A is continuous and injective (the only solution to the equation $Ax = 0$ is the trivial solution $x = 0$, thus $\text{Ker } A = \{0\}$) and its image is closed and given by $\text{Im } A = \mathcal{Y}$, the inverse problem is well-posed, since the inverse operator is continuous;

- if A is not injective, but $\text{Im } A$ is closed, then, if we look for a pseudosolution, the inverse problem becomes well-posed in as far as the generalized inverse is continuous;
- if the image $\text{Im } A$ is not closed, using a pseudo-solution cannot, in itself, guarantee the existence and the continuity of the inverse solution.

When we are dealing with a linear operator defined in spaces of finite dimension \mathbb{R}^N and \mathbb{R}^M and of the type $A: \mathbb{R}^M \rightarrow \mathbb{R}^N$, we again have three main situations:

- if p is the rank of the matrix associated with the operator A and if $p = N = M$, then A is bijective. A solution always exists, it is unique, and the inverse problem is well defined;
- if $p < M$, then the uniqueness is not certain but can be established by considering a generalized inversion;
- if $p < N$, then the existence is not certain for any given data but can be ensured by again considering a generalized inversion.

To conclude this chapter, we see that an inverse problem is often ill-posed or ill-conditioned, and that generalized inversion does not, in general, provide a satisfactory solution. In the next chapter we will see that another development of modern analysis, regularization, allows us to get around these difficulties and gives a generic framework for inversion.

1.7. Bibliography

- [AND 77] ANDREWS H. C., HUNT B. R., *Digital Image Restoration*, Prentice-Hall, Englewood Cliffs, NJ, 1977.
- [AND 80] ANDERSSON R. S., DE HOOG F. R., LUKAS M. A., *The Application and Numerical Solution of Integral Equations*, Sijthoff and Noordhoff, Alphen aan den Rijn, 1980.
- [BER 87] BERTERO M., POGGIO T., TORRE V., Ill-posed Problems in Early Vision, Memo 924, MIT, May 1987.
- [BER 88] BERTERO M., DE MOL C., PIKE E. R., “Linear inverse problems with discrete data: II. Stability and regularization”, *Inverse Problems*, vol. 4, p. 573-594, 1988.
- [BRE 83] BREZIS H., *Analyse fonctionnelle : théorie et applications*, Masson, Paris, 1983.
- [COU 62] COURANT R., HILBERT D., *Methods of Mathematical Physics*, Interscience, London, 1962.
- [FOS 61] FOSTER M., “An application of the Wiener-Kolmogorov smoothing theory to matrix inversion”, *J. Soc. Indust. Appl. Math.*, vol. 9, p. 387-392, 1961.
- [FRA 70] FRANKLIN J. N., “Well-posed stochastic extensions of ill-posed linear problems”, *J. Math. Anal. Appl.*, vol. 31, p. 682-716, 1970.

- [GEM 84] GEMAN S., GEMAN D., “Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images”, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, num. 6, p. 721-741, Nov. 1984.
- [HAD 01] HADAMARD J., “Sur les problèmes aux dérivées partielles et leur signification physique”, *Princeton University Bull.*, vol. 13, 1901.
- [HER 87] HERMAN G. T., TUY H. K., LANGENBERG K. J., SABATIER P. C., *Basic Methods of Tomography and Inverse Problems*, Adam Hilger, Bristol, UK, 1987.
- [KAK 88] KAK A. C., SLANEY M., *Principles of Computerized Tomographic Imaging*, IEEE Press, New York, NY, 1988.
- [KAL 03] KALIFA J., MALLAT S., ROUGÉ B., “Deconvolution by thresholding in mirror wavelet bases”, *IEEE Trans. Image Processing*, vol. 12, num. 4, p. 446-457, Apr. 2003.
- [KLE 80] KLEMA V. C., LAUB A. J., “The singular value decomposition: its computation and some applications”, *IEEE Trans. Automat. Contr.*, vol. AC-25, p. 164-176, 1980.
- [MAR 87] MARROQUIN J. L., MITTER S. K., POGGIO T. A., “Probabilistic solution of ill-posed problems in computational vision”, *J. Amer. Stat. Assoc.*, vol. 82, p. 76-89, 1987.
- [NAS 76] NASHED M. Z., *Generalized Inverses and Applications*, Academic Press, New York, 1976.
- [NAS 81] NASHED M. Z., “Operator-theoretic and computational approaches to ill-posed problems with applications to antenna theory”, *IEEE Trans. Ant. Propag.*, vol. 29, p. 220-231, 1981.
- [PHI 62] PHILLIPS D. L., “A technique for the numerical solution of certain integral equation of the first kind”, *J. Ass. Comput. Mach.*, vol. 9, p. 84-97, 1962.
- [SAB 78] SABATIER P. C., “Introduction to applied inverse problems”, in SABATIER P. C. (Ed.), *Applied Inverse Problems*, p. 2-26, Springer Verlag, Berlin, Germany, 1978.
- [STA 02] STARK J.-L., PANTIN E., MURTAGH F., “Deconvolution in astronomy: a review”, *Publ. Astr. Soc. Pac.*, vol. 114, p. 1051-1069, 2002.
- [TIK 63] TIKHONOV A., “Regularization of incorrectly posed problems”, *Soviet. Math. Dokl.*, vol. 4, p. 1624-1627, 1963.
- [TIK 77] TIKHONOV A., ARSENIN V., *Solutions of Ill-Posed Problems*, Winston, Washington, DC, 1977.
- [TWO 62] TWOMEY S., “On the numerical solution of Fredholm integral equations of the first kind by the inversion of the linear system produced by quadrature”, *J. Assoc. Comp. Mach.*, vol. 10, p. 97-101, 1962.

Chapter 2

Main Approaches to the Regularization of Ill-posed Problems

In the previous chapter, we saw that, when the image $\text{Im } A$ of a linear operator we want to invert is not closed, then the inverse A^{-1} , or the generalized inverse A^\dagger , is not defined everywhere in the data space \mathcal{Y} and is not continuous. This is the case, for example, of compact, non-degenerate (or non-finite rank) operators and it is easy to see that the condition number of the problem is infinite. Suitable solving techniques are thus required.

We also saw that, in a finite dimension, the inverse or the generalized inverse is always continuous. In consequence, the use of a generalized inversion is sufficient to guarantee that the problem is well posed in this case. However, it must not be forgotten that a problem that is well posed but severely ill-conditioned behaves in practice like an ill-posed problem and has to be treated with the same regularization methods, which we present below.

2.1. Regularization

In a finite or infinite dimension, a *regularizer* of equation (1.10) $y = Ax$ is a family of operators $\{R_\alpha; \alpha \in \Lambda\}$ such that [NAS 81, TIK 63]:

$$\begin{cases} \forall \alpha \in \Lambda, & R_\alpha \text{ is a continuous operator of } \mathcal{Y} \text{ in } \mathcal{X}; \\ \forall y \in \text{Im } A, & \lim_{\alpha \rightarrow 0} R_\alpha y = A^\dagger y. \end{cases} \quad (2.1)$$

In other words, since the inverse operator A^{-1} does not have the required continuity or stability properties, we construct a family of continuous operators, indexed by a regulating parameter α (called the *regularization coefficient*) and including A^\dagger as a limit case. Applied to perfect data y , R_α gives an approximation of x^\dagger that is all the better as $\alpha \rightarrow 0$. However, when R_α is applied to data $y_\varepsilon = Ax + b$ that inevitably contain noise, b , we obtain an *approximate* solution $x_\varepsilon = R_\alpha y_\varepsilon$ and we have:

$$R_\alpha y_\varepsilon = R_\alpha y + R_\alpha b. \quad (2.2)$$

The second term diverges when $\alpha \rightarrow 0$. It follows that a trade-off has to be made between two opposing terms, the approximation error (first term) and the error due to noise (second term). This can be done, within a given family of operators R_α , by adjusting the value of the coefficient of regularization α .

Most of the methods that have been put forward for solving and stabilizing ill-posed problems in the past 30 years fall into this general scheme in one way or another. They can be divided into two broad families: those that proceed by dimensionality control – Λ is thus a discrete set – and those that work by minimization of a composite criterion or by optimization under constraint – Λ is thus \mathbb{R}_+ . In what follows, we will mainly concern ourselves with the second family of regularization methods.

2.1.1. Dimensionality control

In the case of an ill-posed or ill-conditioned problem, the methods of regularization by dimensionality control get around the difficulty in two ways:

- by minimizing the criterion $\|y - Ax\|$ (or, more generally, $\mathcal{G}(y - Ax)$) in a subspace of reduced dimension, after an appropriate change of basis if necessary;
- by minimizing the criterion $\mathcal{G}(y - Ax)$ in the space initially chosen but by an iterative method in which the number of iterations is limited.

2.1.1.1. Truncated singular value decomposition

A typical example of methods of the first family can be found by examining equation (1.15): to suppress the ill-conditioned nature of the problem, we just truncate the development, keeping the components corresponding to singular values that are large enough for error terms of the form $\sigma_n^{-1} \langle \delta y, v_n \rangle u_n$ to remain small. This is *truncated singular value decomposition*, or TSVD [AND 77, NAS 81]. It is very effective for ensuring numerical stability. However, the problem arises of choosing the truncation order, which plays the role, here, of the inverse of a regularization coefficient. However, the main failing of this approach is that we give up the possibility of re-establishing the spectral components that have been too degraded by the imaging device. As for the definition of the Rayleigh resolution criterion in optics, we use no information about the object sought other than the fact that its energy is finite although,

most of the time, we know that it is, for example, positive, or that it contains regions of smooth spatial variation separated by sharp boundaries, or that it has bounded values, or bounded support, etc. If we want to go beyond Rayleigh resolution, it is indispensable to be able to take this type of prior information into account.

2.1.1.2. *Change of discretization*

In truncated singular value decomposition, it is the imaging device that more or less imposes the discretization through singular functions of the corresponding operator. However, we can also avoid the difficulties raised by the poor conditioning of matrix \mathbf{A} , as a consequence of the object's being discretized on a Cartesian grid for example, by choosing a *parsimonious parameterization* of the object better suited to its prior properties. This is what wavelet-based decomposition methods [STA 02, KAL 03] do, for example. The principle remains the same: *thresholding* is applied to the coefficients of the decomposition so as to eliminate the subspace dominated by the noise components.

This mode of discretization solves the problem of stability or poor conditioning analyzed above but, even so, does not always provide a satisfactory solution. Everything depends on the decomposition that is chosen.

2.1.1.3. *Iterative methods*

A very popular family of methods is made up of *iterative methods* of the form:

$$x^{(n+1)} = x^{(n)} + \alpha (y - Ax^{(n)}), \quad n = 0, 1, \dots \quad (2.3)$$

where $0 < \alpha < 2/\|A\|$ (Bialy's method [BIA 59]). If A is a non-negative, bounded, linear operator (i.e., $\langle Ax, x \rangle \geq 0, \forall x \in \mathcal{X}$) and if $y = Ax$ has at least one solution, then the series of $x^{(n)}$ converges and:

$$\lim_{n \rightarrow \infty} x^{(n)} = P x^{(0)} + \hat{x}^{\text{GI}},$$

where P is the orthogonal projection operator on $\text{Ker } A$ and \hat{x}^{GI} the generalized inverse solution. In fact, this method looks for the fixed point of the operator $G : Gx = \alpha y + (I - \alpha A)x$, but if A is compact and \mathcal{X} is of infinite dimension, then $I - \alpha A$ is not a contraction and the method diverges. Moreover, we have also seen that, even in finite dimensions, the generalized inverse solution is often dominated by the noise.

The non-negativity condition excludes a lot of operators but the method can be applied to solve the normal equation $A^*y = A^*Ax$ since A^*A is a non-negative operator. We thus obtain Landweber's method [LAN 51]:

$$x^{(n+1)} = x^{(n)} + \alpha A^* (y - Ax^{(n)}), \quad n = 0, 1, \dots \quad (2.4)$$

with $0 < \alpha < 2/\|A^*A\|$. The well known Gerchberg-Saxton-Papoulis-VanCittert [BUR 31] method for extrapolating a limited-spectrum signal is a special case of the

Bialy-Landweber method. It is in this same category of iterative methods that we can place Lucy's method [LUC 74], which is very popular in astronomy.

All these methods can provide an acceptable solution only on the condition that the number of iterations is limited (which plays the role of the inverse of a regularization coefficient) [DIA 70]. This is often done empirically, as the initial framework does not take observation noise into account and a theory regulating the number of iterations so as to limit noise amplification cannot but be heteronomous [LUC 94]. This explains why the rest of this book will focus on regularization methods of the second family, which operate by minimization under constraint and are, from this point of view, more autonomous.

2.1.2. *Minimization of a composite criterion*

The principal characteristic of the regularization methods of this second large family is to require the solution to be a trade-off between fidelity to the measured data and fidelity to the prior information [TIT 85]. This trade-off is reached using a single optimality criterion. The approach can be interpreted as follows.

The least squares solutions to equation (1.17) minimize the energy of the discrepancy between the model Ax and the data y . In this sense, they achieve the greatest fidelity to the data. However, when the observation noise is broadband, relation (1.15) shows that the high spatial frequency components of the restored or reconstructed object have large amplitudes because of noise amplification. The least squares solutions thus prove unacceptable because we expect the real object to have markedly smoother spatial variations. We therefore need to introduce a little infidelity to the data to obtain a solution that is smoother than the least squares solution and closer to the idea that we have *a priori*. A widely accepted means of doing this is by the minimization of a composite criterion [NAS 81, TIK 63, TIK 77]. The basic idea is to give up any hope of reaching the exact solution from imperfect data, to consider as *admissible* any solution for which Ax is not far from y , and to look among the admissible solutions to find the one that can be considered as the physically most reasonable, i.e., compatible with certain prior information. This is usually done by finding a solution x_α that minimizes a criterion of the form:

$$\mathcal{J}(x) = \mathcal{G}(y - Ax) + \alpha \mathcal{F}(x), \quad 0 < \alpha < +\infty, \quad (2.5)$$

specifically designed so that:

- the solution is faithful to the data up to a certain point (first term of the criterion);
- certain desirable properties that sum up our prior knowledge about the solution are reinforced (second term).

The choice of the functionals \mathcal{F} and \mathcal{G} is qualitative and determines how the regularization is carried out. Conversely, the choice of α , which is the coefficient

of regularization here, is quantitative and allows the compromise between the two sources of information to be adjusted. Perfect fidelity to the data is obtained with $\alpha = 0$, while perfect fidelity to the prior information is obtained if $\alpha = \infty$.

One of the most widely studied regularization methods is obtained by minimizing the functional:

$$\mathcal{J}(x) = \|y - Ax\|_Y^2 + \alpha \|Cx\|_{\mathcal{X}}^2, \quad (2.6)$$

where C is a constraint operator [NAS 76]. The existence of a solution is ensured when C is bounded with $\text{Im } C$, for example, but that excludes the very interesting case of a differential operator, as in Tikhonov's seminal article [TIK 63]:

$$\|Cx\|_{\mathcal{X}}^2 = \sum_{p=0}^P \int c_p(r) |x^{(p)}(r)|^2 dr,$$

where the weighting functions $c_p(r)$ are strictly positive and $x^{(p)}$ designates the p th order derivative of x . The corresponding regularizer can be written:

$$R_\alpha = (A^*A + \alpha C^*C)^{-1}A^*. \quad (2.7)$$

In this case, $x_\alpha = R_\alpha y$ exists and is unique when the domain of C is dense in \mathcal{X} and the equations $Ax = 0$ and $Cx = 0$ only have in common the trivial solution $x = 0$. This solution takes a very simple form when A is compact and $C = I$, the identity operator in \mathcal{X} . By using the singular value decomposition of A of section 1.3, we obtain:

$$x_\alpha = \sum_{n \in E} \frac{\sigma_n}{\sigma_n + \alpha} \frac{1}{\sigma_n} \langle y, v_n \rangle u_n. \quad (2.8)$$

It is thus essentially a “*filtered*” version of the non-regularized solution (1.14), or generalized inverse, of equation (1.10). We will often find this idea of *linear filtering* later, associated with the oldest regularization methods, but, for the moment, we will concern ourselves mainly with discrete problems in finite dimensions.

In the discrete case, the literature on the subject is dominated by a few functionals. Below are the ones most frequently found [TIT 85].

2.1.2.1. Euclidian distances

The squared Euclidian distance between two objects \mathbf{x}_1 and \mathbf{x}_2 is defined by:

$$\|\mathbf{x}_1 - \mathbf{x}_2\|_{\mathbf{P}}^2 = (\mathbf{x}_1 - \mathbf{x}_2)^T \mathbf{P} (\mathbf{x}_1 - \mathbf{x}_2),$$

where \mathbf{P} is a symmetric positive semi-definite matrix, chosen to express certain desirable characteristics of the proximity measurement. Such a squared distance is the

habitual choice for \mathcal{G} in the case where the noise \mathbf{b} is assumed to be zero-mean, Gaussian, independent of \mathbf{x} , and of probability density:

$$p(\mathbf{b}) \propto \exp \left\{ -\frac{1}{2} \mathbf{b}^T \mathbf{P} \mathbf{b} \right\}, \quad (2.9)$$

i.e., of covariance matrix \mathbf{P}^{-1} . Such a distance is also very often used for \mathcal{F} in order to penalize objects \mathbf{x} of large amplitude.

Applied to the interferometry example of Chapter 1 (which, we recall, is continuous-discrete), this mode of regularization leads us to look for:

$$\widehat{\mathbf{x}}^{(\alpha)} = \arg \min_{\widehat{\mathbf{x}} \in L^2_{\mathbb{C}}[0,1]} \left(\|\mathbf{x}_N - \mathbf{y}\|^2 + \alpha \int_0^1 |\widehat{\mathbf{x}}(\nu)|^2 d\nu \right),$$

where $\mathbf{x}_N = [x_1, \dots, x_N]^T$ and $x_k = \int_0^1 \widehat{\mathbf{x}}(\nu) \exp \{2j\pi k\nu\} d\nu$. The solution reads:

$$\widehat{\mathbf{x}}^{(\alpha)}(\nu) = \frac{1}{\alpha + 1} \widehat{\mathbf{x}}^{\text{gi}}(\nu).$$

The spectrum thus regularized is therefore proportional to the periodogram (1.7) that is obtained as a limit case ($\alpha \rightarrow 0$). Hence, this type of regularization is not suitable in this example.

However, we will see that the *linear-quadratic* framework above (that combines the *linear* nature of the direct model (1.3) and the *quadratic* nature of the functionals \mathcal{F} and \mathcal{G}) turns out to be very handy in practice. The minimization of a criterion such as:

$$\mathcal{J}(\mathbf{x}) = \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_{\mathbf{P}}^2 + \alpha \|\mathbf{x} - \overline{\mathbf{x}}\|_{\mathbf{Q}}^2, \quad (2.10)$$

where $\overline{\mathbf{x}}$ is a *default* solution (it is the solution obtained when $\alpha \rightarrow \infty$, i.e., when the weight given to the data tends towards zero), provides an explicit expression of the minimizer:

$$\widehat{\mathbf{x}} = (\mathbf{A}^T \mathbf{P} \mathbf{A} + \alpha \mathbf{Q})^{-1} (\mathbf{A}^T \mathbf{P} \mathbf{y} - \mathbf{Q} \overline{\mathbf{x}}) \quad (2.11)$$

which, thanks to the matrix inversion lemma [SCH 17, SCH 18], can be written:

$$\widehat{\mathbf{x}} = \overline{\mathbf{x}} + \mathbf{Q}^{-1} \mathbf{A}^T (\mathbf{A} \mathbf{Q}^{-1} \mathbf{A}^T + \alpha^{-1} \mathbf{P}^{-1})^{-1} (\mathbf{y} - \mathbf{A} \overline{\mathbf{x}}). \quad (2.12)$$

In this expression, the matrix to be inverted has, in general, different dimensions from that of (2.11).

2.1.2.2. Roughness measures

A very simple way of measuring the roughness of an image is to apply an appropriate difference operator and then calculate the Euclidian norm of the result. As the differentiation operation is linear with respect to the original image, the resulting measure of roughness is quadratic:

$$\mathcal{F}(\mathbf{x}) = \|\nabla^k(\mathbf{x})\|^2 = \|\mathbf{D}_k \mathbf{x}\|^2. \quad (2.13)$$

The order k of the difference operator ∇^k is habitually 1 or 2. Measure (2.13) is minimum when \mathbf{x} is constant ($k = 1$), affine ($k = 2$), etc.

2.1.2.3. *Non-quadratic penalization*

Another way of preserving the discontinuities in an object, better than the regularization methods using quadratic criteria, is to use *non-quadratic* penalty functions [IDI 99]. This is precisely what was done to process the interferometry data of Chapter 1, section 1.2, by finally choosing criterion (1.8). The principle is to use a function that increases more slowly than a parabola so as to apply smaller penalties to large variations. These functions are of two main types:

- L_2L_1 functions, i.e., continuously differentiable, convex functions that behave quadratically at the origin and are asymptotically linear. A typical example is the branch of a hyperbola;
- L_2L_0 functions, which differ from the previous ones by being asymptotically constant and thus non-convex.

This time, it is no longer possible to obtain an explicit solution but the first functions have the advantage of being convex, so the standard minimization techniques are sure to converge to the global minimum and give some *robustness* to the solution [BOU 93]. The others enable the discontinuities to be effectively detected but at the expense of some instability and high computing costs [GEM 92].

2.1.2.4. *Kullback pseudo-distance*

In many image processing problems, it is essential to preserve the positivity of the pixel intensities. One way of doing this is to consider that the positive object can be identified, after normalization, with a probability distribution, and then to use the distance measures between probability laws. In particular, the Kullback pseudo-distance (or divergence, or information) of a probability π with respect to a probability π_0 (such that π is absolutely continuous with respect to π_0) can be written:

$$\mathcal{K}(\pi_0, \pi) = \int \left(-\log \frac{d\pi}{d\pi_0} \right) d\pi_0.$$

For a *reference object* whose components m_j are positive,

$$\mathcal{F}(\mathbf{x}) = \mathcal{K}(\mathbf{x}, \mathbf{m}) = \sum_{j=1}^M x_j \log \frac{x_j}{m_j} \quad (2.14)$$

is often used. Here again, it is not possible to obtain an explicit expression for the solution; it has to be calculated iteratively [LEB 99].

Criterion (2.5) sums up a view of regularization that can be called *deterministic*, since the only probability law used is, at least implicitly though the choice of the functional \mathcal{G} , the one for noise. It has led to important theoretical developments, essentially in mathematical physics. However, the questions of choice of the regularizing functional $\mathcal{F}(\mathbf{x})$ [CUL 79] and the adjustment of the regularization coefficient α [THO 91]

are still very open. Section 2.3 presents the principal methods for adjusting the regularization coefficient that exist in this framework. However, whether we want to set this *hyperparameter* by such a *supervised* method or not, we need to be capable of minimizing the regularized criterion (2.5) in practice afterwards, as a function of \mathbf{x} . This important aspect of inversion is dealt with below.

2.2. Criterion descent methods

Implicitly or explicitly, most inversion methods are based on the minimization of a criterion. According to the properties of the latter, the computing cost of the solution can vary enormously, typically by a factor of a thousand between the minimization of a quadratic criterion by inversion of a linear system and that of a multimodal criterion by a relaxation technique such as *simulated annealing*, everything else being equal. Finally, the choice of the “right” inversion method depends on the computing facilities available. And we still need to know which algorithm to use for a given optimization problem. For instance, in the comparison above, it would be possible, but completely inefficient, to use simulated annealing to minimize a quadratic criterion. This section gives a non-exhaustive overview of optimization problems in the context of inverse problems in signal and image processing, with the associated algorithms. It obviously cannot replace the literature devoted to optimization as a whole, such as [NOC 99] or [BER 95].

2.2.1. Criterion minimization for inversion

By *criterion minimization*, we understand: finding the $\hat{\mathbf{x}}$ that minimizes $\mathcal{J}(\mathbf{x})$ among the elements of \mathcal{X} . In the rest of this section, we consider the case of real vectors¹: $\mathcal{X} \subset \mathbb{R}^M$. The criterion \mathcal{J} and the set \mathcal{X} may depend on the data, and structural properties (additional terms in the expression for \mathcal{J} expressing “soft” constraints, whereas the specification of \mathcal{X} is likely to impose “hard” constraints), hyperparameters managing the compromise between fidelity to data and regularity.

Thus, in the case of the generalized inverse of section 1.4, $\mathcal{J}(\mathbf{x}) = \|\mathbf{x}\|$ and $\mathcal{X} = \{\mathbf{x}, \mathbf{A}^T \mathbf{A} \mathbf{x} = \mathbf{A}^T \mathbf{y}\}$ is the set of minimizers of $\|\mathbf{y} - \mathbf{A} \mathbf{x}\|$. In the case of the specification of composite criteria dealt with in section 2.1.2,

$$\mathcal{J}(\mathbf{x}) = \mathcal{G}(\mathbf{y} - \mathbf{A} \mathbf{x}) + \alpha \mathcal{F}(\mathbf{x}), \quad (2.15a)$$

$$\text{with } \mathcal{X} = \mathbb{R}^M \text{ (non-constrained case)} \quad (2.15b)$$

$$\text{or } \mathcal{X} = \mathbb{R}_+^M \text{ (positivity constraint)} \quad (2.15c)$$

1. The case where \mathbf{x} is a function (more precisely, the case of a space \mathcal{X} of infinite dimension) poses mathematical difficulties that come under functional analysis.

Defining $\hat{\mathbf{x}}$ formally as the minimizer of a criterion hides three main levels of difficulty in terms of implementation. In order of increasing complexity we have:

- ① \mathcal{J} is quadratic: $\mathcal{J} = \mathbf{x}^T \mathbf{M} \mathbf{x} - 2 \mathbf{v}^T \mathbf{x} + \text{const.}$ and $\mathcal{X} = \mathbb{R}^M$, or else \mathcal{X} is affine: $\mathcal{X} = \{\mathbf{x}_0 + \mathbf{B}\mathbf{u}, \mathbf{u} \in \mathbb{R}^P, P < M\}$;
- ② \mathcal{J} is a differentiable convex function and $\mathcal{X} = \mathbb{R}^M$ or a convex (closed) subset of \mathbb{R}^M ;
- ③ \mathcal{J} has no known properties.

2.2.2. The quadratic case

In situation ①, with $\mathcal{X} = \mathbb{R}^M$ assuming \mathbf{M} is symmetric and invertible, $\hat{\mathbf{x}}$ is the solution of the *linear system* $\mathbf{M} \hat{\mathbf{x}} = \mathbf{v}$ of dimensions $M \times M$, which expresses the fact that the gradient becomes zero, $\nabla \mathcal{J}(\hat{\mathbf{x}}) = \mathbf{0}$. We have already encountered a similar expression in (2.11) and we will meet it again in the Gaussian linear probabilistic framework of Chapter 3.

In the variant constrained to a space \mathcal{X} that is affine, we need only to replace \mathbf{x} by its expression in \mathbf{u} to get back to the unconstrained minimization of a quadratic criterion, in \mathbb{R}^P .

2.2.2.1. Non-iterative techniques

A finite number of operations is sufficient to invert any linear system: of the order of M^3 operations (and M^2 memory locations) for an $M \times M$ system. If the *normal matrix* $\mathbf{M} = \{m_{ij}\}$ has a particular structure, the system inversion cost may decrease.

In signal processing, the stationary nature of a signal is expressed by the Toeplitz character of the normal matrix (i.e., $m_{ij} = \mu_{j-i}$). In image processing using a stationary hypothesis, the normal matrix is Toeplitz-block-Toeplitz (i.e., Toeplitz by blocks, the blocks of sub-matrices themselves being Toeplitz). In both these cases, we find inversion algorithms costing of the order of M^2 operations and M memory locations (Levinson algorithm) and even fast algorithms using a fast Fourier transform costing only of the order of $M \log M$ operations. The spectral expression for the “Wiener filter” of Chapter 4 is a special case where “fast” implementation is possible for the case of a *circulant* normal matrix (i.e., $m_{ij} = \mu_{j-i \bmod M}$).

The *sparse* nature of the normal matrix can also be used to good advantage: if only ML coefficients of \mathbf{M} are non-zero, we can hope to decrease the inversion cost in terms of the number of operations and variables to be stored. For example, if \mathbf{M} is a *band* matrix ($m_{ij} = 0$ if $|j - i| \geq \ell < M$: a band matrix is sparse and L is of the same order as ℓ), the inversion cost does not exceed $M\ell^2$ operations and $M\ell$ memory locations. In particular, a normal matrix $\mathbf{M} = \mathbf{A}^T \mathbf{A}$ is band if \mathbf{A} corresponds to filtering by a small finite impulse response.

2.2.2.2. Iterative techniques

If the number of unknowns M is very large (e.g., pixels in image restoration, or voxels for 3D objects), the memory cost of non-iterative techniques often becomes prohibitive. It is then preferable to use a *fixed point* method, iteratively engendering a series $\hat{\mathbf{x}}^{(i)}$ with a limit $\hat{\mathbf{x}} = \mathbf{M}^{-1}\mathbf{v}$. All the conventional variants verify $\mathcal{J}(\hat{\mathbf{x}}^{(i+1)}) \leq \mathcal{J}(\hat{\mathbf{x}}^{(i)})$. Three families can be distinguished.

2.2.2.2.1. “Column-action” algorithms

A single component differs between $\hat{\mathbf{x}}^{(i)}$ and $\hat{\mathbf{x}}^{(i+1)}$. The M components are scanned cyclicly during the iterations. This is the principle of the Gauss-Seidel method, or *coordinate descent* [BER 95, p. 143], also called ICM (*iterative conditional modes*) or ICD (*iterative coordinate descent*) in image restoration [BES 86, BOU 93]. It can be generalized for blocks of components and is all the more interesting and partially parallelizable if \mathbf{A} is sparse.

2.2.2.2.2. “Row-action” algorithms

A single component of \mathbf{y} is taken into account to calculate $\hat{\mathbf{x}}^{(i+1)}$ from $\hat{\mathbf{x}}^{(i)}$. The N data are scanned cyclicly during the iterations, which makes this approach inevitable if the data are too numerous to be processed simultaneously. The *algebraic reconstruction techniques* (ART), long-standing references in medical imaging by X-ray tomography, follow this principle to minimize the least squares criterion $\|\mathbf{y} - \mathbf{Ax}\|^2$ [GIL 72]. They can be generalized to the penalized criterion $\|\mathbf{y} - \mathbf{Ax}\|^2 + \alpha \|\mathbf{x}\|^2$ [HER 79], can process blocks of data and are all the more interesting and partially parallelizable if \mathbf{A} is sparse.

2.2.2.2.3. “Global” techniques

At each iteration, all the unknowns are updated according to all the data. The *gradient* algorithms are prototypes of the global approach:

$$\hat{\mathbf{x}}^{(i+1)} = \hat{\mathbf{x}}^{(i)} - \lambda(\hat{\mathbf{x}}^{(i)}) \nabla \mathcal{J}(\hat{\mathbf{x}}^{(i)}),$$

with $\nabla \mathcal{J}(\hat{\mathbf{x}}^{(i)}) = 2 \mathbf{M} \hat{\mathbf{x}}^{(i)} - 2 \mathbf{v}$. Note that the Landweber method defined by (2.4) is in fact a gradient technique minimizing the non-regularized criterion $\|\mathbf{y} - \mathbf{Ax}\|^2$. The *conjugate gradient* or *pseudo-conjugate gradient* algorithms are variants that converge more rapidly, in which the successive descent directions combine the previously calculated gradients to avoid the zigzag trajectory of the simple gradient [PRE 86, p. 303]. These variants are of first order, thus occupying little memory; they use only the M first derivatives $\partial \mathcal{J} / \partial x_m$. The preconditioning technique can further increase the efficiency of CG algorithms, as explained in Chapter 4, section 4.4.4, in the context of deconvolution.

We will end with second order techniques. In situation ①, with $\mathcal{X} = \mathbb{R}^M$, and taking \mathbf{M} to be symmetric and invertible, each iteration of the standard Newton’s

method can be written:

$$\hat{\mathbf{x}}^{(i+1)} = \hat{\mathbf{x}}^{(i)} - (\nabla^2 \mathcal{J}(\hat{\mathbf{x}}^{(i)}))^{-1} \nabla \mathcal{J}(\hat{\mathbf{x}}^{(i)}) = \mathbf{M}^{-1} \mathbf{v},$$

considering that $\nabla \mathcal{J}(\mathbf{x}) = 2\mathbf{M}\mathbf{x} - 2\mathbf{v}$ and $\nabla^2 \mathcal{J}(\mathbf{x}) = 2\mathbf{M}$. In other words, a single iteration of this algorithm is equivalent to solving the problem itself. Unless \mathbf{M} has a specific structure, the computation cost is prohibitive for most realistic inversion problems. Some *quasi-Newton* variants become iterative again by approaching \mathbf{M}^{-1} by a series of matrices $\mathbf{P}^{(i)}$. The most popular among them is the BFGS (Broyden-Fletcher-Goldfarb-Shanno) method [NOC 99, Chapter 8]. For large-sized problems, the computation burden of such quasi-Newton methods is still too high. A better choice is to resort to *limited-memory* BFGS, which can be seen as an extension of the CG method, in-between first and second order techniques [NOC 99, Chapter 9].

2.2.3. The convex case

The quadratic criteria are part of a larger family of functions \mathcal{J} that are convex, i.e., such that $\forall \mathbf{x}_1, \mathbf{x}_2 \in \Omega, \theta \in (0, 1)$,

$$\mathcal{J}(\theta \mathbf{x}_1 + (1 - \theta) \mathbf{x}_2) \leq \theta \mathcal{J}(\mathbf{x}_1) + (1 - \theta) \mathcal{J}(\mathbf{x}_2)$$

with $\mathcal{X} = \mathbb{R}^M$. Similarly, \mathcal{X} is a convex set if $\forall \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}, \theta \in (0, 1)$, we have $\theta \mathbf{x}_1 + (1 - \theta) \mathbf{x}_2 \in \mathcal{X}$.

The specification that the criteria be convex but not necessarily quadratic gives a wider choice as far as modeling is concerned. The Kullback pseudo-distance (2.14) is convex over \mathbb{R}_+^M ; the Markov penalty functions $\mathcal{F}(\mathbf{x}) = \sum_j \varphi(x_j - x_{j+1})$ are convex over \mathbb{R}^M if φ is a convex scalar function such as $\varphi(x) = \sqrt{\tau^2 + x^2}$, which was used in the spectrometry example of Chapter 1, section 1.2.

The minimization of non-quadratic convex criteria, although more difficult and more costly than the minimization of quadratic criteria, remains altogether compatible with modern computing resources, which explains the increasingly frequent use of convex penalty functions in signal and image restoration [IDI 99]. Let us start by recalling a few fundamental properties of convex criteria [BER 95, App. B]:

- a convex continuous criterion \mathcal{J} is *unimodal*: any local minimum is global and the set of its minimizers is convex;
- if $\mathcal{J}_1, \mathcal{J}_2$ are convex and $\alpha_1, \alpha_2 \geq 0$, then $\alpha_1 \mathcal{J}_1 + \alpha_2 \mathcal{J}_2$ is convex²;

2. This property “explains” why we are interested in convexity rather than unimodality: for example, the penalized criterion (2.15a) is convex (so unimodal) if \mathcal{G} and \mathcal{F} are convex, whereas the unimodality of \mathcal{G} and \mathcal{F} would not be enough to guarantee the unimodality of the criterion.

– if \mathcal{J} is *strictly* convex, there exists one and only one minimizer \hat{x} in any convex \mathcal{X} that is *closed* (i.e., the boundary of \mathcal{X} belongs to \mathcal{X}).

On the other hand, if the criterion is non-quadratic, the minimizer \hat{x} is a function of the data that is generally *neither linear, nor explicit*. Owing to this, the non-iterative inversion techniques for linear systems of section 2.2.2 are no longer valid. In contrast, the three families of iterative techniques based on the successive reduction of the criterion give algorithms that converge towards \hat{x} if \mathcal{J} is convex and differentiable and if $\mathcal{X} = \mathbb{R}^M$. The case of a criterion that is convex but not differentiable is slightly trickier; modern techniques, known as *interior point techniques*, approach the solution by minimizing a succession of differentiable convex approximations [BER 95, p. 312].

There are also other possible families of convergent techniques: *reweighted least squares*, also called *semi-quadratic* algorithms (see Chapter 6), or the approaches based on maximizing a *dual* criterion [BER 95, HEI 00, LUE 69].

If \mathcal{X} is a closed convex subset of \mathbb{R}^M , some adaptation is necessary: *projected gradient* or *conditional gradient* versions in the family of “global techniques” [BER 95, Chapter 2], and techniques of *projection on convex sets* [SEZ 82, YOU 82] in the family of “row-action” techniques. As for “column-action” techniques, they remain particularly simple if the constraints are *separable*, i.e., if \mathcal{X} is a Cartesian product, e.g. the positivity corresponds to $\mathcal{X} = \mathbb{R}_+ \times \cdots \times \mathbb{R}_+$. Finally, certain constrained problems are equivalent to a non-constrained problem in the dual domain, which justifies the use of dual methods.

2.2.4. General case

In the case of non-convex criteria, the possible existence of local minima makes the use of descent techniques risky, in the sense that any local minimizer is a possible fixed point for most of these techniques. Whether we have convergence towards \hat{x} rather than towards a local solution then depends on the initialization. Several strategies can be envisaged for avoiding these local solutions. Apart from exceptional cases, they are notoriously more costly than the descent methods and yet still do not guarantee convergence towards the global minimizer. Without guaranteeing convergence mathematically, some techniques are nevertheless sufficiently robust to avoid aberrant solutions. They then give results that could not have been obtained by minimization of a convex criterion, for applications such as automatic image segmentation or object detection.

Two types of approach can be distinguished. On the one hand we have deterministic methods that, in the absence of mathematical convergence properties, favor robustness. For instance, the principle of *gradual non-convexity* (GNC) [BLA 87, NIK 98]

consists of gradually minimizing a series of criteria using a conventional descent technique, starting with a convex criterion and finishing with the non-convex criterion. The robustness of this technique comes from the quality of the initial solution. Its implementation cost and complexity are relatively low. On the other hand, we have the pseudo-random methods (simulated annealing [GEM 84] and adaptive random search [PRO 84]), which make use of the generation of a large number of random samples to avoid the traps. Simulated annealing has (probabilistic) convergence properties but the high computing cost of such techniques explains why their use is still limited in the signal and image restoration field.

2.3. Choice of regularization coefficient

There are few methods for determining the hyperparameters in the framework of this chapter [THO 91]. The most frequently used are the following.

2.3.1. Residual error energy control

One of the most intuitive and oldest ideas for setting the value of α that comes into the regularized, or penalized, criterion (2.5) is to consider α as a Lagrange multiplier in the equivalent problem:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \mathcal{F}(\mathbf{x}) \text{ s. t. } \mathcal{G}(\mathbf{y} - \mathbf{A}\mathbf{x}) = c. \quad (2.16)$$

The degree of regularization is fixed by the value of c , which can be considered as a statistic for which the probability distribution can be deduced from $p(\mathbf{y} | \mathbf{x})$. When $\mathcal{G} = \|\cdot\|^2$ and \mathbf{x}_0 is the *true* solution, the vector of the *residuals* $\mathbf{y} - \mathbf{A}\mathbf{x}_0$ follows the law of the noise, which is implicitly taken to be homogeneous, zero-mean, white and Gaussian. It results from this that c/σ^2 is a variable of χ^2 with N degrees of freedom if σ^2 is the variance of the noise. It is then recommended to set c to its expectation value, i.e., $N\sigma^2$. However, such a choice often leads to overregularization of the solution. One explanation is that the regularized solution $\hat{\mathbf{x}}$ inevitably differs from the true solution and that the residual errors $\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}$ that are effectively calculable to obtain the value of \mathcal{G} do not follow any known distribution. Moreover, in many problems, the graph of the function $\mathcal{G}(\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}) = \mathcal{G}(\alpha)$ is practically horizontal over a large range of values of α : any error in the estimation of σ^2 thus leads to large variations in the value of α that satisfies constraint (2.16).

2.3.2. “L-curve” method

It is also possible to use an alternative method that has proved its worth in linear inverse problems of form (2.5) and in the case where the regularization functional $\mathcal{F}(\mathbf{x})$

is quadratic. This is the “L-curve” method [HAN 92]. It consists of using a log-log scale to plot the regularization functional $\mathcal{F}(\hat{\mathbf{x}}(\alpha))$ against the least squares criterion $\|\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}(\alpha)\|^2$ by varying the regularization coefficient α . This curve generally has a characteristic L shape (whence its name) and the value of α corresponding to the corner of the L provides a good compromise between the contradictory requirements of fidelity to the data and fidelity to the prior information.

To understand why this is so, we know that, if \mathbf{x}_0 is the exact solution, then the error $\hat{\mathbf{x}}(\alpha) - \mathbf{x}_0$ can be divided into two parts: a *perturbation* error due to the presence of the measuring error \mathbf{b} and a *regularization* error due to the use of a regularizing operator instead of an inverse operator (see (2.2)). The vertical part of the L-curve, described for low values of α , corresponds to solutions for which $\mathcal{F}(\hat{\mathbf{x}}(\alpha))$ is very sensitive to variations in α , as the measurement error \mathbf{b} dominates $\hat{\mathbf{x}}(\alpha)$ and does not satisfy the discrete Picard condition [HAN 92]. The horizontal part of the curve, described for high values of α , corresponds to solutions for which it is the sum of the squares of the residuals $\|\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}(\alpha)\|^2$ that is the most sensitive to variations of α , since $\hat{\mathbf{x}}(\alpha)$ is dominated by the regularization error as long as $\mathbf{y} - \mathbf{b}$ satisfies the discrete Picard condition.

2.3.3. Cross-validation

In the case where the hyperparameters of problem (2.5) are limited simply to the regularization coefficient and where \mathcal{F} and \mathcal{G} are quadratic, *cross-validation* methods also provide acceptable solutions [GOL 79, WAH 77].

The aim is to find a value of the regularization coefficient α such that the regularized solution:

$$\hat{\mathbf{x}}(\alpha, \mathbf{y}) = \arg \min_{\mathbf{x}} (\mathcal{G}(\mathbf{y} - \mathbf{A}\mathbf{x}) + \alpha \mathcal{F}(\mathbf{x})) \quad (2.17)$$

is as close as possible to the actual object \mathbf{x} . Let Δ_x be a measure of the distance between $\hat{\mathbf{x}}(\alpha, \mathbf{y})$ and \mathbf{x} . With the choice of quadratic distances for \mathcal{F} and \mathcal{G} , it is natural to also choose a quadratic distance for Δ_x :

$$\Delta_x(\alpha, \mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \hat{\mathbf{x}}(\alpha, \mathbf{y})\|^2. \quad (2.18)$$

Δ_x can be interpreted as a loss function measuring the *risk* involved in using $\hat{\mathbf{x}}(\alpha, \mathbf{y})$ instead of \mathbf{x} . A reasonable method for choosing α would be to choose the value that minimizes this risk on the average, i.e., the mean square error (MSE):

$$MSE(\alpha, \mathbf{x}) = \int \Delta_x(\alpha, \mathbf{x}, \mathbf{y}) p(\mathbf{y} | \mathbf{x}) d\mathbf{y} \quad (2.19)$$

which is an expectation with respect to the noise probability distribution (2.9). Unfortunately, the solution to this problem:

$$\alpha^{\text{MSE}}(\mathbf{y}, \mathbf{x}) = \arg \min_{\alpha} MSE(\alpha, \mathbf{x}) \quad (2.20)$$

depends on the real object which, obviously, is unknown. As the regularized solution $\hat{\mathbf{x}}(\alpha, \mathbf{y})$ can also be seen as a predictor of the observations through $\hat{\mathbf{y}}(\alpha, \mathbf{y}) = \mathbf{A}\hat{\mathbf{x}}(\alpha, \mathbf{y})$, it is possible to measure the difference between the real and predicted observations with the following loss function:

$$\Delta_y(\alpha, \mathbf{x}, \mathbf{y}) = \|\mathbf{A}\mathbf{x} - \mathbf{A}\hat{\mathbf{x}}(\alpha, \mathbf{y})\|^2. \quad (2.21)$$

The value of α could be obtained by minimizing the corresponding mean risk, which, in this case, is the MSE on the prediction:

$$MSEP(\alpha, \mathbf{x}) = \int \Delta_y(\alpha, \mathbf{x}, \mathbf{y}) p(\mathbf{y} | \mathbf{x}) d\mathbf{y} \quad (2.22)$$

but, there again, the solution depends on the real object. The difficulty can, however, be overcome because the criterion $MSEP(\alpha, \mathbf{x})$ can be estimated by *generalized cross-validation* (GCV). Its basic principle is the following. Let $\hat{\mathbf{x}}(\alpha, \mathbf{y}^{[-k]})$ be the minimizer of the criterion:

$$\mathcal{J}^{[-k]}(\mathbf{x}) = \sum_{n \neq k} |y_n - (\mathbf{A}\mathbf{x})_n|^2 + \alpha \|\mathbf{x}\|_{\mathbf{Q}}^2, \quad (2.23)$$

i.e., the object restored by using all the data *except* sample y_k . It is possible to use $\hat{\mathbf{x}}(\alpha, \mathbf{y}^{[-k]})$ next to predict the missing data item:

$$\hat{y}_k^{[-k]}(\alpha) = [\mathbf{A}\hat{\mathbf{x}}(\alpha, \mathbf{y}^{[-k]})]_k. \quad (2.24)$$

The method consists of looking for the value of α that minimizes a weighted energy of the prediction error $\alpha^{\text{GCV}} = \arg \min_{\alpha} V(\alpha)$, with:

$$V(\alpha) = \frac{1}{N} \sum_{k=1}^N w_k^2(\alpha) (y_k - \hat{y}_k^{[-k]}(\alpha))^2, \quad (2.25)$$

where the coefficients $w_k^2(\alpha)$ are introduced to avoid criterion (2.25) having undesirable properties, such as a lack of invariance during arbitrary rotations of the observation space, or absence of a minimum. They are given by:

$$w_k(\alpha) = \frac{1 - B_{kk}(\alpha)}{1 - \text{trace}(\mathbf{B}(\alpha)) / M},$$

where B_{kk} is the k th diagonal element of the matrix $\mathbf{B}(\alpha) = \mathbf{A}(\mathbf{A}\mathbf{A}^T + \alpha\mathbf{Q})^{-1}\mathbf{A}^T$. The calculation of the minimum relies on the “linear-quadratic” nature of the problem, which allows a simpler relation to be established:

$$V(\alpha) = \frac{N \|\mathbf{I} - \mathbf{B}(\alpha)\| \|\mathbf{y}\|^2}{(\text{trace}(\mathbf{I} - \mathbf{B}(\alpha)))^2}. \quad (2.26)$$

This clearly shows that the GCV function $V(\alpha)$ is, in fact, the sum of the squares of the residual errors *weighted by a coefficient that depends on α* . This method has interesting asymptotic statistical properties. For example [LI 86], $\hat{\mathbf{x}}^{(\alpha^{\text{GCV}}, \mathbf{y})}$ gives almost surely the minimum of $\|\mathbf{A}\mathbf{x} - \mathbf{A}\hat{\mathbf{x}}(\alpha, \mathbf{y})\|^2$ when $N \rightarrow \infty$. Nevertheless, it has to be understood that such a result is of interest only in the case of parsimonious parameterization of the object sought, with a number M of parameters much smaller than the number N of data points. These asymptotic properties and numerous practical results explain why this method has so often been used in 1-D problems. Its use in image processing is more recent [FOR 93, REE 90].

These methods for choosing the regularization coefficient are only clearly justified in the framework of quadratic regularized criteria. The stochastic extension of Chapter 3 will allow us to go beyond this framework.

2.4. Bibliography

- [AND 77] ANDREWS H. C., HUNT B. R., *Digital Image Restoration*, Prentice-Hall, Englewood Cliffs, NJ, 1977.
- [BER 95] BERTSEKAS D. P., *Nonlinear Programming*, Athena Scientific, Belmont, MA, 1995.
- [BES 86] BESAG J. E., "On the statistical analysis of dirty pictures (with discussion)", *J. R. Statist. Soc. B*, vol. 48, num. 3, p. 259-302, 1986.
- [BIA 59] BIALY H., "Iterative Behandlung linearen Funktionalgleichungen", *Arch. Ration. Mech. Anal.*, vol. 4, p. 166-176, 1959.
- [BLA 87] BLAKE A., ZISSERMAN A., *Visual Reconstruction*, The MIT Press, Cambridge, MA, 1987.
- [BOU 93] BOUMAN C. A., SAUER K. D., "A generalized Gaussian image model for edge-preserving MAP estimation", *IEEE Trans. Image Processing*, vol. 2, num. 3, p. 296-310, July 1993.
- [BUR 31] BURGER H. S., VAN CITTERT P. H., "Wahre und Scheinbare Intensitätsverteilung in Spektrallinien", *Z. Phys.*, vol. 79, p. 722, 1931.
- [CUL 79] CULLUM J., "The effective choice of the smoothing norm in regularization", *Math. Comp.*, vol. 33, p. 149-170, 1979.
- [DIA 70] DIAZ J. B., METCALF F. T., "On iteration procedures for equation of the first kind, $Ax = y$, and Picard's criterion for the existence of a solution", *Math. Comp.*, vol. 24, p. 923-935, 1970.
- [FOR 93] FORTIER N., DEMOMENT G., GOUSSARD Y., "GCV and ML methods of determining parameters in image restoration by regularization: fast computation in the spatial domain and experimental comparison", *J. Visual Comm. Image Repres.*, vol. 4, num. 2, p. 157-170, June 1993.

- [GEM 84] GEMAN S., GEMAN D., "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, num. 6, p. 721-741, Nov. 1984.
- [GEM 92] GEMAN D., REYNOLDS G., "Constrained restoration and the recovery of discontinuities", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, num. 3, p. 367-383, Mar. 1992.
- [GIL 72] GILBERT P., "Iterative methods for the three-dimensional reconstruction of an object from projections", *J. Theor. Biol.*, vol. 36, p. 105-117, 1972.
- [GOL 79] GOLUB G. H., HEATH M., WAHBA G., "Generalized cross-validation as a method for choosing a good ridge parameter", *Technometrics*, vol. 21, num. 2, p. 215-223, May 1979.
- [HAN 92] HANSEN P., "Analysis of discrete ill-posed problems by means of the L-curve", *SIAM Rev.*, vol. 34, p. 561-580, 1992.
- [HEI 00] HEINRICH C., DEMOMENT G., "Minimization of strictly convex functions: an improved optimality test based on Fenchel duality", *Inverse Problems*, vol. 16, p. 795-810, 2000.
- [HER 79] HERMAN G. T., HURWITZ H., LENT A., LUNG H. P., "On the Bayesian approach to image reconstruction", *Inform. Contr.*, vol. 42, p. 60-71, 1979.
- [IDI 99] IDIER J., "Regularization tools and models for image and signal reconstruction", in *3rd Intern. Conf. Inverse Problems in Engng.*, Port Ludlow, WA, p. 23-29, June 1999.
- [KAL 03] KALIFA J., MALLAT S., ROUGÉ B., "Deconvolution by thresholding in mirror wavelet bases", *IEEE Trans. Image Processing*, vol. 12, num. 4, p. 446-457, Apr. 2003.
- [LAN 51] LANDWEBER L., "An iteration formula for Fredholm integral equations of the first kind", *Amer. J. Math.*, vol. 73, p. 615-624, 1951.
- [LEB 99] LE BESNERAIS G., BERCHER J.-F., DEMOMENT G., "A new look at entropy for solving linear inverse problems", *IEEE Trans. Inf. Theory*, vol. 45, num. 5, p. 1565-1578, July 1999.
- [LI 86] LI K. C., "Asymptotic optimality of C_L and GCV in ridge regression with application to spline smoothing", *Ann. Statist.*, vol. 14, p. 1101-1112, 1986.
- [LUC 74] LUCY L. B., "An iterative technique for the rectification of observed distributions", *Astron. J.*, vol. 79, num. 6, p. 745-754, 1974.
- [LUC 94] LUCY L. B., "Optimum strategies for inverse problems in statistical astronomy", *Astron. Astrophys.*, vol. 289, num. 3, p. 983-994, 1994.
- [LUE 69] LUENBERGER D. G., *Optimization by Vector Space Methods*, John Wiley, New York, NY, 1st edition, 1969.
- [NAS 76] NASHED M. Z., *Generalized Inverses and Applications*, Academic Press, New York, 1976.
- [NAS 81] NASHED M. Z., "Operator-theoretic and computational approaches to ill-posed problems with applications to antenna theory", *IEEE Trans. Ant. Propag.*, vol. 29, p. 220-231, 1981.

- [NIK 98] NIKOLOVA M., IDIER J., MOHAMMAD-DJAFARI A., "Inversion of large-support ill-posed linear operators using a piecewise Gaussian MRF", *IEEE Trans. Image Processing*, vol. 7, num. 4, p. 571-585, Apr. 1998.
- [NOC 99] NOCEDAL J., WRIGHT S. J., *Numerical Optimization*, Springer Texts in Operations Research, Springer-Verlag, New York, NY, 1999.
- [PRE 86] PRESS W. H., FLANNERY B. P., TEUKOLSKY S. A., VETTERLING W. T., *Numerical Recipes, the Art of Scientific Computing*, Cambridge University Press, Cambridge, MA, 1986.
- [PRO 84] PRONZATO L., WALTER E., VENOT A., LEBRUCHEC J.-F., "A general-purpose global optimizer: implementation and applications", *Mathematics and Computers in Simulation*, vol. 26, p. 412-422, 1984.
- [REE 90] REEVES S. J., MERSEREAU R. M., "Optimal estimation of the regularization parameter and stabilizing functional for regularized image restoration", *Opt. Engng.*, vol. 29, p. 446-454, 1990.
- [SCH 17] SCHUR I., "Über Potenzreihen, die im Innern des Einheitskreises beschränkt sind", *J. Reine Angew. Math.*, vol. 147, p. 205-232, 1917.
- [SCH 18] SCHUR I., "Über Potenzreihen, die im Innern des Einheitskreises beschränkt sind", *J. Reine Angew. Math.*, vol. 148, p. 122-145, 1918.
- [SEZ 82] SEZAN M. I., STARK H., "Image restoration by the method of convex projections: Part 2 – Applications and numerical results", *IEEE Trans. Medical Imaging*, vol. MI-1, num. 2, p. 95-101, Oct. 1982.
- [STA 02] STARK J.-L., PANTIN E., MURTAGH F., "Deconvolution in astronomy: a review", *Publ. Astr. Soc. Pac.*, vol. 114, p. 1051-1069, 2002.
- [THO 91] THOMPSON A., BROWN J. C., KAY J. W., TITTERINGTON D. M., "A study of methods of choosing the smoothing parameter in image restoration by regularization", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-13, num. 4, p. 326-339, Apr. 1991.
- [TIK 63] TIKHONOV A., "Regularization of incorrectly posed problems", *Soviet. Math. Dokl.*, vol. 4, p. 1624-1627, 1963.
- [TIK 77] TIKHONOV A., ARSENIN V., *Solutions of Ill-Posed Problems*, Winston, Washington, DC, 1977.
- [TIT 85] TITTERINGTON D. M., "Common structure of smoothing techniques in statistics", *Int. Statist. Rev.*, vol. 53, num. 2, p. 141-170, 1985.
- [WAH 77] WAHBA G., "Practical approximate solutions to linear operator equations when the data are noisy", *SIAM J. Num. Anal.*, vol. 14, p. 651-667, 1977.
- [YOU 82] YOULA D. C., WEBB H., "Image restoration by the method of convex projection: part 1 – Theory", *IEEE Trans. Medical Imaging*, vol. MI-1, num. 2, p. 81-94, Oct. 1982.

Chapter 3

Inversion within the Probabilistic Framework

There are at least two reasons that encourage us to consider solving inverse problems in a Bayesian framework [DEM 89]. It was in this framework that local energy functions and Markov modeling, which have had a lasting influence on low-level image processing, were introduced. It is also this same framework that provides the most consistent and complete answers to problems left in abeyance in other approaches, such as the choice of hyperparameters or the optimization of a multimodal criterion.

3.1. Inversion and inference

To make the link between inversion and statistical inference more explicit, it is useful at this stage to sum up the analysis carried out in Chapter 1. After discretization, the direct problem takes the general form $A(\mathbf{x}, \mathbf{y}) = 0$, where A is an operator linking the unknown object $\mathbf{x} \in \mathbb{R}^M$ to the experimental data $\mathbf{y} \in \mathbb{R}^N$. Often, it even takes the explicit form $\mathbf{y} = A(\mathbf{x})$ or the linear form $\mathbf{y} = \mathbf{A}\mathbf{x}$, \mathbf{A} being a matrix. Inversion, i.e., the calculation of \mathbf{x} when \mathbf{A} and \mathbf{y} are known, is very often an ill-posed problem in two senses.

Firstly, the operator \mathbf{A} is often singular, in the sense that there is a class \mathcal{K} of solutions $\mathbf{x} \in \mathcal{K}$ such that $\mathbf{A}\mathbf{x} = 0$ (the kernel $\text{Ker } \mathbf{A} = \mathcal{K}$ is thus not empty). Any element of \mathcal{K} can be added to any solution to give another solution and we cannot, therefore, invert the direct relation to determine \mathbf{x} uniquely from \mathbf{y} . This lack of uniqueness makes the discrete inverse problem ill-posed in Hadamard's sense. This situation occurs whenever the instrument response destroys part of the information

necessary for the reconstruction of the object. Let us not forget, however, that this ambiguity can be removed by using a more or less empirical rule for choosing among all the solutions, such as taking the minimum norm solution, for example.

Secondly, and more critically, no experimental device is completely free of uncertainty, the simplest source being the finite accuracy of the measurements. It is thus more realistic to consider that the object sought and the measurements taken are connected by an equation of the form $\mathbf{y} = A(\mathbf{x}) \diamond \mathbf{b}$, in which A is an operator describing the essential part of the experiment and $\diamond \mathbf{b}$ accounts for the deterioration of this ideal representation by various sources of error (of discretization, measurement, etc.) grouped together in the *noise* term. When the observation mechanism can be approximated by a linear distortion and the addition of noise, this equation reduces to (1.3): $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$. The presence of this noise has the effect of “spreading” the set \mathcal{K} , since any element \mathbf{x} such that $\mathbf{A}\mathbf{x} = \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon}$ is “small” relative to the assumed level of noise, can be added to any possible solution to obtain another acceptable solution. However, above all, if the ambiguity is removed by taking a rule for choosing an acceptable solution, it is observed in practice that the latter behaves in an unstable way; small changes in the data entail large variations in the calculated solution. This can easily happen even when the solution is unique and depends continuously on the data, i.e., when the problem is well-posed in Hadamard’s sense. In fact, the instability comes from the fact that \mathbf{A} is *ill-conditioned* (see section 1.5).

So we see that, in ill-posed problems, obtaining a solution is not so much a problem of mathematical deduction as a problem of *inference*, i.e., of information processing, which can be summed up in the following question: “how can we draw the best possible conclusions from the incomplete information at our disposal?”

To be acceptable, any scientific inference method should: 1) take *all the available pertinent information* into account; 2) carefully *avoid* assuming information is available when it is not. Probabilistic modeling is a handy, consistent way of describing a situation of incomplete information. We will now see how it leads to a Bayesian statistical approach.

3.2. Statistical inference

It should be made clear from the start that any problem dealt with through a Bayesian approach has to be *well-posed* in the sense that enough information must be provided to allow the probability distributions needed for the calculation to be attributed without ambiguity. This means, at least, that an exhaustive set of possibilities must be specified at the start of each problem. We will call this the *data space* (or *proof space*) if it concerns possible results of the experiment, or the *hypothesis space* if it specifies the hypotheses that we wish to verify. It is also useful to distinguish between two classes of problems, called *estimation* and *choice of model*. The first

studies the consequences of choosing a particular model that is assumed “true”, while the aim of the choice of model is to select one model by comparison with one or more other possible candidates.

In an estimation problem, we assume that the model is true for *one* (unknown) value \mathbf{x}_0 of its parameters and we explore the constraints imposed on the parameters by the data. The hypothesis space is thus the set of all possible values of the parameters $\mathcal{H} = \{\mathbf{x}_i\}$. The data consist of one or more samples. For the problem to be well-posed, the space of all the possible samples, $\mathcal{S} = \{\mathbf{z}_i\}$, must also be stated. The spaces \mathcal{H} and \mathcal{S} can both be discrete or continuous.

Before making the estimation, it is necessary to state a *logical environment* I which defines our working framework (hypothesis space, data space, relationships between parameters and data, any additional information). Typically, I is defined as a logical proposition stating:

- that the true value of the parameter is in \mathcal{H} ;
- that the observed data consist of N samples of the space \mathcal{S}^N ;
- how the parameters are connected with the data (this is the role of the direct model A);
- any additional information.

Of course, the physical nature of the parameters and data is implicitly specified in \mathcal{H} , \mathcal{S} and A . Implicitly, all the developments that follow will be within the framework defined by I , which signifies that any probability distribution will be conditioned by I . This conditioning will not be indicated explicitly in order to lighten the notation.

We can now get started on the estimation problem by calculating the probability that each of the possible values of the parameter is the actual value. Let D designate the proposition affirming the values of the experimental data actually observed and H the proposition $\mathbf{x}_0 = \mathbf{x}$ affirming that one of the possible values of the parameter \mathbf{x} is the actual value \mathbf{x}_0 .

3.2.1. Noise law and direct distribution for data

In any statistical inference method intended to solve a problem such as (1.3), it is necessary to start by choosing a probability law $q(\mathbf{b})$ describing our information – or our uncertainty – on the errors \mathbf{b} . This is an essential step as it allows the *direct*, or *sampling*, distribution to be found:

$$p(\mathbf{y} | \mathbf{x}) = q(\mathbf{y} - A(\mathbf{x})) . \quad (3.1)$$

In the vast majority of cases, a centered Gaussian distribution, independent of \mathbf{x} , is chosen for the errors, which gives:

$$p(\mathbf{y} | \mathbf{x}) = (2\pi)^{-N/2} |\mathbf{R}|^{-1/2} \exp \left\{ -\frac{1}{2} \|\mathbf{y} - A(\mathbf{x})\|_{\mathbf{R}^{-1}}^2 \right\},$$

where \mathbf{R} designates the covariance matrix of the distribution $q(\mathbf{b})$. It is often diagonal, or even proportional to the identity. A question arises immediately: What sense is to be given to such a choice and in what situations is such a model appropriate?

With a *frequentist's* interpretation of a probability, the distribution for the noise should be that of the frequencies of its values in a very large number of repeated measurements. It is then justified by reference to the central limit theorem which says, under fairly broad conditions, that if the noise in a sample of data is the result of a large number of accumulated elementary effects that are “random” and independent, the Gaussian distribution is a good approximation of the real frequency distribution. However, except for fluctuations of electronic origin in a measurement system, the noise is not generally the result of independent effects (think, for example, of the discretization errors that depend on the solution \mathbf{x}_0). Moreover, to be able to make an inference with this interpretation, it would be necessary for us to have numerous results of other measurements so as to be able to determine these frequencies, which is an extremely rare experimental situation.

This Gaussian “hypothesis” is thus not a hypothesis on the “random” nature of the noise. We are not at all claiming that whatever gives rise to the noise is really random and follows a Gaussian distribution. It is not even a hypothesis in the true sense of the word; it is rather the least compromising – or the most conservative – choice that we can make for the noise distribution in a situation of uncertainty. We are assuming two things here: 1) that the noise can take any real value but that its average value is zero; in other words, there is no systematic measurement error (or if there is, we have been able to detect and correct it), and 2) that we expect there to be a “typical scale” of noise; in other words, large contributions to the noise are not as probable as small ones. To put it another way, we think that the distribution for the noise should have a mean value of zero and a finite standard deviation, even if we have no precise idea of the value of the latter. On the other hand, we have no idea as to the existence or otherwise of cumulants of order greater than two. In these conditions, the least compromising choice with respect to the characteristics that we do not know – which can be justified by information principles [JAY 82] – is that of a Gaussian distribution. In addition, if we suspect that the noise components affecting the N samples have different scales and are correlated, the covariance matrix of the distribution is there to express this hypothesis. It is not necessary to specify its value but if it is unknown, its elements, grouped together in a vector of *hyperparameters* $\boldsymbol{\theta}$, will in general only complicate the problem. They are called *nuisance* parameters for this reason.

This choice is appropriate whenever this information is all we know about the noise. As this is a frequent situation, the choice is often made. If we have additional information about the noise, which leads us to choose a non-Gaussian distribution, we can include it in the same way but the result will be significantly better only if the distribution is very different from a Gaussian one. There are situations – such as imaging with a low particle count – where the data are integers and have low values. Choosing a binomial or Poisson distribution can then improve the results.

3.2.2. Maximum likelihood estimation

With simply this direct distribution $p(\mathbf{y} | \mathbf{x}, \boldsymbol{\theta})$, we could define the solution of the inverse problem as being that of *maximum likelihood* (ML), the likelihood being the direct distribution in which \mathbf{y} takes its observed value and parameter \mathbf{x} becomes the variable:

$$\hat{\mathbf{x}}^{\text{ML}} = \arg \max_{\mathbf{x} \in \mathcal{H}} p(\mathbf{y} | \mathbf{x}, \boldsymbol{\theta}).$$

In general, the justification for this choice comes from the “good” statistical characteristics (more often than not asymptotic) of this estimator. The *least squares* solution is the special case of the maximum likelihood solution when the direct distribution is Gaussian:

$$\hat{\mathbf{x}}^{\text{LS}} = \arg \min_{\mathbf{x} \in \mathcal{H}} (\mathbf{y} - A(\mathbf{x}))^T \mathbf{R}^{-1} (\mathbf{y} - A(\mathbf{x})).$$

Introduced in this way, it is still a *weighted* least squares method (weighted by the matrix \mathbf{R}^{-1}) that possesses the indispensable property of invariance under changes of units in \mathcal{H} and \mathcal{S} . In many simple situations, this inference method provides all the information we are looking for. However, in inverse problems where the parameterization is not parsimonious, the direct distribution does not contain all the information needed to make the problem well-posed and it does not provide all the technical apparatus necessary for the calculation:

1) In the special case of an indeterminate linear problem $\mathbf{y} = \mathbf{A}\mathbf{x}$, where \mathbf{A} is singular (a problem known as *generalized inversion*), there is no “noise” and so no direct distribution, except in the rudimentary sense where $p(\mathbf{y} | \mathbf{x})$ is constant if \mathbf{x} is in the class \mathcal{C} of possible antecedents of \mathbf{y} , and zero otherwise. As the likelihood is constant in class \mathcal{C} , maximizing it is of no help for the choice within this class. The essence of the problem does not lie in the presence of “random” noise perturbing the data, but rather in the fact that our information is incomplete, although essentially noise free.

2) In the linear case (1.3), matrix \mathbf{A} of the direct problem is often *ill-conditioned*. The solving operator $\mathbf{A}^\dagger = (\mathbf{A}^T \mathbf{R}^{-1} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{R}^{-1}$ is unstable and the solution $\hat{\mathbf{x}}^{\text{ML}} = \mathbf{A}^\dagger \mathbf{y}$ is unacceptable: the amplification of the noise is excessive.

3) The problem can have nuisance parameters that are of no interest to us, and they may be numerous. When matrix \mathbf{R} is full, $N(N-1)/2$ hyperparameters are added

which, when they are unknown, generally have to be estimated by ML as parameters of interest x , and the global maximum may then no longer be a point but a whole region.

4) We may have highly pertinent information on the solution we are looking for. For example, we may know that it has to be positive, or satisfy certain constraints (as in astronomical imaging where the integral of the object may already be known), or that it is made up of homogeneous regions separated by clear boundaries. Such information is not contained in the direct distribution but it would be most unreasonable to ignore it.

5) In many problems, it is necessary to obtain not only a solution but also an indication of the confidence we can have in it. If we simply have the direct distribution (3.1), the *confidence intervals* given by the frequency approach only give us information on the *long term* behavior of the solution, i.e., its average behavior over a very large number of repeats of the experiment. However, we only possess the results of a single experiment, which often cannot be reproduced.

6) Finally, the estimation of the parameters of a model that is assumed to be valid is often just one step and we may need to judge the relative merits of various models.

It is therefore necessary to go beyond inference by ML. All the extensions mentioned above are “automatically” provided by the Bayesian approach.

3.3. Bayesian approach to inversion

Bayesian inference is so named because it makes great use of Bayes’ rule, which itself is a consequence of a fundamental rule in probability calculation, the *product rule* [COX 61]. Let H be a hypothesis whose truth we want to evaluate and D a set of data connected with this hypothesis. The product rule stipulates that:

$$\Pr(H, D) = \Pr(H | D) \Pr(D) = \Pr(D | H) \Pr(H)$$

where, for example, $\Pr(H | D)$ usually designates the probability that H is true knowing D . From this we draw Bayes’ rule:

$$\Pr(H | D) = \Pr(H) \Pr(D | H) / \Pr(D)$$

which is none other than a *learning* rule. It tells us how we should adjust the probability attributed to the truth of a hypothesis when our state of knowledge changes with the acquisition of data. The probability *a posteriori* for H , $\Pr(H | D)$, is obtained by multiplying its probability *a priori*, $\Pr(H)$, by the probability of having observed the data D assuming the hypothesis is true, $\Pr(D | H)$, and dividing the whole by the probability of having observed the data independently of whether the hypothesis is true or not, $\Pr(D)$. This last term, sometimes called the *global likelihood*, plays the role of a normalization constant.

A large part of statistical inference is based on the use of prior information on the quantities to be estimated, which adds to the information given by the data. Thus, it is not surprising, if we think about the deep nature of the regularization principle set out in Chapter 2, that it shows a close link with Bayesian inference.

In the case of an inverse problem such as (1.3) and assuming that the probability distributions concerned admit a density, the prior information on object \mathbf{x} is expressed, in a Bayesian context, in the form of an *a priori* probability density function (pdf) $p(\mathbf{x} | \boldsymbol{\theta})$. Bayes' rule allows us to combine this with the information contained in the data to obtain the *a posteriori* law:

$$p(\mathbf{x} | \mathbf{y}, A, \boldsymbol{\theta}) = \frac{p(\mathbf{x} | \boldsymbol{\theta}) p(\mathbf{y} | \mathbf{x}, A, \boldsymbol{\theta})}{p(\mathbf{y} | A, \boldsymbol{\theta})} = \frac{p(\mathbf{x}, \mathbf{y} | A, \boldsymbol{\theta})}{p(\mathbf{y} | A, \boldsymbol{\theta})}. \quad (3.2)$$

In this equation, $\boldsymbol{\theta}$ is a vector of *hyperparameters* composed of the parameters of the *a priori* distributions of the errors and the object, and $p(\mathbf{y} | \mathbf{x}, A, \boldsymbol{\theta})$ designates the data law conditioned by the true solution \mathbf{x} . It is completely determined by the knowledge of the direct model (1.3) and the noise probability law. The last term ensures the normalization of the *a posteriori* law:

$$p(\mathbf{y} | A, \boldsymbol{\theta}) = \int p(\mathbf{y} | \mathbf{x}, A, \boldsymbol{\theta}) p(\mathbf{x} | \boldsymbol{\theta}) d\mathbf{x}. \quad (3.3)$$

In the Bayesian approach, the knowledge (or uncertainty) about object \mathbf{x} after observation of *data* \mathbf{y} only is wholly described by the probability distribution (3.2). This probability is equal, with just a multiplying factor, to the product of the likelihood introduced in section 3.2 by the *a priori* probability $p(\mathbf{x} | \boldsymbol{\theta})$. If we assume that, in the case of section 3.2, the knowledge of \mathbf{x} (which then comes purely from observations and from the structure of the problem) is represented by the likelihood, we observe that, in the Bayesian approach, taking prior information into consideration by means of $p(\mathbf{x} | \boldsymbol{\theta})$ modifies our knowledge and, in general, has the effect of reducing the uncertainty on the parameter \mathbf{x} . But above all, because of the framework adopted, the Bayesian approach enables a wider range of answers to the question “*given a probability distribution for a continuous or discrete parameter \mathbf{x} , what best estimate can be made and with what accuracy?*”. There is not a single answer to this question; the problem concerns the theory of the decision that answers the question “*what should we do?*”. This implies value judgements and consequently goes beyond the principles of inference, which only answers the question “*what do we know?*”. Thus, we can equally well deduce a point estimator or a region of uncertainty from (3.3) [MAR 87, TAR 87]. The *maximum a posteriori* is a frequent choice for the estimator. It consists of giving \mathbf{x} the value that maximizes the distribution *a posteriori*:

$$\hat{\mathbf{x}}^{\text{MAP}} = \arg \max p(\mathbf{x} | \mathbf{y}, A, \boldsymbol{\theta}). \quad (3.4)$$

However, this is only one of the possible solutions. This MAP estimation corresponds to the minimization of a mean decision cost with an all-or-nothing cost function, the

limit (when $\varepsilon \rightarrow 0$) of the mean cost $\Pr(\|\hat{\mathbf{x}} - \mathbf{x}_0\| > \varepsilon)$. Other cost functions have been proposed in the framework of image modeling by Markov fields. They lead to the maximization of the marginal probabilities [BES 86, MAR 87].

3.4. Links with deterministic methods

In the case that interests us here, i.e., an inverse problem in a finite dimension, it is clear that regularizing according to the general principle indicated in Chapter 2, and thus minimizing a criterion such as (2.5), is equivalent to choosing the solution that maximizes the following *a posteriori* law:

$$p(\mathbf{x} | \mathbf{y}, A, \boldsymbol{\theta}) \propto \exp \left\{ -\frac{1}{2\sigma^2} (\mathcal{G}(\mathbf{y} - A(\mathbf{x})) + \alpha \mathcal{F}(\mathbf{x})) \right\}. \quad (3.5)$$

where σ^2 is the variance of the noise. The above probability law is only one of the possible choices since any strictly monotonic function other than an exponential would do. However, this choice is particularly suitable here because, with the linear model (1.3), taking the usual hypotheses that the noise is Gaussian and independent, as \mathcal{G} is a Euclidian norm, the conditional law $p(\mathbf{y} | \mathbf{x}, A, \boldsymbol{\theta})$ is really:

$$p(\mathbf{y} | \mathbf{x}, A, \boldsymbol{\theta}) \propto \exp \left\{ -\frac{1}{2\sigma^2} \mathcal{G}(\mathbf{y} - A(\mathbf{x})) \right\}. \quad (3.6)$$

For the analogy to be complete, the *a priori* law must take the following form:

$$p(\mathbf{x} | \boldsymbol{\theta}) \propto \exp \left\{ -\frac{\alpha}{2\sigma^2} \mathcal{F}(\mathbf{x}) \right\}, \quad (3.7)$$

and, for it to be rigorous, the *a posteriori* law (3.5) must be proper, a sufficient condition being that (3.6) and (3.7) are also proper:

$$\int_{\mathbb{R}^N} \exp \left\{ -\frac{1}{2\sigma^2} \mathcal{G}(\mathbf{y} - A(\mathbf{x})) \right\} d\mathbf{y} < +\infty, \quad \int_{\mathbb{R}^M} \exp \left\{ -\frac{\alpha}{2\sigma^2} \mathcal{F}(\mathbf{x}) \right\} d\mathbf{x} < +\infty.$$

Many local energy functions used in image processing were introduced in a Bayesian framework. They define \mathbf{x} as a Markov field (see Chapter 7). Although the *energy* point of view is also held by some members of the image processing community, criteria of form (2.5) can generally be reinterpreted in a Bayesian framework, even if it means making minor changes in \mathcal{F} to ensure the normalization of equation (3.7).

In consequence, the maximum *a posteriori* estimator, which is the Bayesian estimator the most used in inversion, becomes the same as the minimizer of the penalized criterion (2.5):

$$\begin{aligned} \hat{\mathbf{x}}^{\text{MAP}} &= \arg \max_{\mathbf{x}} p(\mathbf{x} | \mathbf{y}, A, \boldsymbol{\theta}) = \arg \max_{\mathbf{x}} p(\mathbf{x}, \mathbf{y} | A, \boldsymbol{\theta}) \\ &= \arg \min_{\mathbf{x}} \mathcal{G}(\mathbf{y} - A(\mathbf{x})) + \alpha \mathcal{F}(\mathbf{x}) \end{aligned}$$

under the technical conditions that allow this development (principally, that the problem brings in a finite number of variables). It is thus obvious that the Bayesian framework gives a statistical sense to the minimization of penalized criteria. The question is not, however, whether the Bayesian approach is a justification of the other approaches. We could also, and conversely, say that the same result gives a deterministic interpretation of the probabilistic estimator of the maximum *a posteriori* and that an estimator, once defined, depends no more on the formal framework that engendered it than on the digital means used to calculate it. The question is rather one of seeing that the Bayesian approach provides an answer to the problems raised in section 3.2. In addition to its great consistency, it makes original tools available:

- *marginalization* (everything that does not interest us is simply integrated out of the problem);
- *regression* (the conditional expectation does not have an equivalent in the energy framework);
- *stochastic sampling* (Monte Carlo methods, simulated annealing algorithms, genetic algorithms), not conceivable without the Bayesian approach (on this point, see Chapter 7, section 7.4.2).

3.5. Choice of hyperparameters

The Bayesian framework appreciably extends the range of methods available for determining the hyperparameters. To be applied effectively, all the methods described in Chapter 2 require us to choose the value of the regularization coefficient α and, more generally, all hyperparameters θ defining the \mathcal{F} and \mathcal{G} distance measures: the variance of the noise, the object correlation parameters and the parameters of the local energy functions. The determination of θ is the most delicate step in image restoration and reconstruction methods. Although the problem is still open, the Bayesian approach provides consistent tools for tackling it.

Hyperparameters θ constitute a second level in the description of the problem, which is indispensable to “rigidify” the first level composed of the parameters themselves – i.e., the object x . In an ill-posed problem, the value of the parameters is important for obtaining an acceptable solution but has no intrinsic interest. In a Bayesian approach, two levels of inference can be distinguished. The first is inference on x , for a given value of θ , through the *a posteriori* distribution of equation (3.2). The second is inference on θ through the analog relationship:

$$p(\theta | y, A) = p(\theta | A) p(y | \theta, A) / p(y | A). \quad (3.8)$$

Here again, we find a characteristic of the use of Bayes’ rule: the marginal likelihood $p(y | \theta, A)$ attached to the data in the second level is the coefficient of normalization in the first.

If, as is often the case, this term is sufficiently peaked, i.e., if the data \mathbf{y} contain enough information, the influence of the *a priori* distribution $p(\boldsymbol{\theta} | A)$ is negligible and the second level of inference can be solved by maximizing the likelihood. But to do this, we have to solve the marginalization problem corresponding to the calculation of the integral in (3.3). Such integrals rarely lead to an explicit result. One notable exception is the joint Gaussian distribution $p(\mathbf{x}, \mathbf{y} | \boldsymbol{\theta}, A)$, as we will see in section 3.8.

To get around the problem posed by the explicit calculation of a marginal likelihood, we can introduce “hidden variables” \mathbf{q} which complete the observations \mathbf{y} in such a way that the new likelihood $p(\mathbf{y}, \mathbf{q} | \boldsymbol{\theta}, A)$ is simpler to calculate. We are then led to maximize the conditional expectations by iterative, deterministic or stochastic techniques (EM and SEM algorithms) [DEM 77], the algorithm converging towards the solution of ML. The need for such stochastic approaches appeared when it was found to be impossible to implement convergent likelihood maximization methods by conventional optimization techniques, as the likelihood was not calculable.

Furthermore, the joint distribution or *generalized likelihood*:

$$p(\mathbf{y}, \mathbf{x} | \boldsymbol{\theta}, A) = p(\mathbf{x} | \mathbf{y}, \boldsymbol{\theta}, A) p(\mathbf{y} | \boldsymbol{\theta}, A) = p(\mathbf{y} | \mathbf{x}, \boldsymbol{\theta}, A) p(\mathbf{x} | \boldsymbol{\theta}) \quad (3.9)$$

sums up all the information at the first level of inference. Its maximization with respect to \mathbf{x} and $\boldsymbol{\theta}$ can be envisaged. Thus, the integration problem raised by (3.3) is obviously removed. At fixed $\boldsymbol{\theta}$, the *generalized maximum likelihood* (GML) coincides with the MAP; at fixed \mathbf{x} , it corresponds to the usual ML for $\boldsymbol{\theta}$, \mathbf{x} and \mathbf{y} being known. Nevertheless, repeated alternation of these two steps is hazardous: the characteristics of the corresponding estimator are not those of the usual ML [LIT 83]. It can even happen sometimes that the GML is not defined because the likelihood may have no maximum, even local [GAS 92]. This technique thus has a marked empirical character.

Thus, the Bayesian approach leads fairly naturally to the use of estimators based on likelihood for the estimation of the hyperparameters. Despite definite difficulties of implementation, interesting results have been obtained in a one-dimensional framework. In a two- or three-dimensional framework, we have to be more cautious. Although it is possible to estimate the hyperparameters in several cases, the values obtained using this approach do not necessarily lead to good results for the estimation of the parameter of interest \mathbf{x} , particularly when the latter comes from “natural” data. The cause could lie in there being too great a difference between these natural data and the behavior of the *a priori* model. The question of hyperparameter estimation thus remains wide open.

3.6. *A priori* model

A reproach that is often levelled against Bayesian estimation is that it depends on the knowledge of a hypothetical, uncertain “true model” that engendered the object

to be reconstructed. To formulate this reproach, we have to implicitly accept that reality can be “enclosed” in a mathematical model. This opens up a huge philosophical debate... In the case of the probabilistic approach to inversion as we see it, the frequentist’s interpretation of the probabilities maintains an annoying confusion. It does, however, seem important to recall that our probabilistic hypotheses are not hypotheses on the “random” character of the object but choices of a way of representing incomplete prior information – or uncertain knowledge – compatible with the chosen inference tool. This situation is far from unusual, as it is rare for the prior information available in a real problem to come in a form directly suited to the theoretical framework chosen for its processing.

Let us remember that the advantages of the Bayesian approach stem not so much from the additional information introduced by the prior – the energy and deterministic interpretations of the functional of regularization $\mathcal{F}(x)$ of section 3.4 show that this information is not proper to the Bayesian approach, and the information on nuisance parameters is *diffuse* most of the time – as from the access it provides to a layer of tools that does not exist in the other approaches, such as marginalization, regression and pseudo-random algorithms.

Having said this, the conversion of prior information into probabilities is a tricky problem that is still far from being solved. To describe object x , the prior is often chosen *pragmatically*, as we will see later. There are, however, some *formal rules* that lead to reasonable choices [BER 94, KAS 94, ROB 97] and are used in particular for the hyperparameters. They often lead to an *improper law*, which does not cause any special difficulty if it is handled correctly [JEF 39]. Here are a few examples.

Some methods rely on *transformation group* theory to determine the “natural” reference measure for the problem and to satisfy certain invariance principles. In practice though, this approach has done little more than justify the use of Lebesgue’s method for the *localization* parameters (thus providing an extension to the continuous case of the uniform distribution resulting from the application of Bernouilli’s “indifference principle” in the discrete case) and the Jeffreys measure in the case of *scale* parameters [JEF 39, POL 92].

Other methods are based on information principles. These are mainly *maximum entropy* methods (MEM), in which we look for the distribution that is closest to the reference distribution (in the Kullback divergence sense) whilst verifying incomplete prior information [JAY 82]. There again, this approach has mainly just helped to justify certain choices after the event. In addition, it is only really workable when the prior information is made up of linear constraints on the distribution we are looking for (moments). We are thus working in the family of *exponential distributions*.

Another formal principle consists of using a *conjugate prior*, i.e., a prior belonging to the same family as the direct distribution of the problem, to obtain an *a posteriori*

distribution in the family [ROB 97]. This is only of interest if the family in question is as small as possible and parametrized. In this case, the step from *a priori* to *a posteriori*, by application of Bayes' rule, comes down to updating the parameters. The interest of this method is essentially technical, as the *a posteriori* is always calculable, at least up to a certain point. A partial justification can also be found by invariance reasoning: if the data \mathbf{y} change $p(\mathbf{x})$ into $p(\mathbf{x} | \mathbf{y})$, the information that \mathbf{y} contributes about \mathbf{x} is clearly limited; it should not lead to a change of the whole structure of $p(\mathbf{x})$, but only of its parameters. It is obvious though that the main motivation for using the method is its convenience. However, only certain families of direct distributions, such as *exponential families* [BRO 86], guarantee the existence of conjugate priors and it is often necessary to limit use of the method to this class of distributions. In addition, the “automatic” nature of this way of making choices is rather deceptive because additional *hyperparameters* – the values of which have to be specified – inevitably appear.

A last, very important class is composed of “tailor made” constructions, in other words, constructions that are not based on general principles like the previous ones but make pragmatic use of probabilistic methods that express the properties expected of the solutions as well as possible. It is into this category that we must put the Gibbs-Markov fields, which have undergone spectacular development in imaging since 1984 [GEM 84] and which allow essential local properties that an object must possess to be incorporated into an *a priori* distribution. The construction of these models requires considerable know-how but is a very powerful way of incorporating elaborate prior information. The price to be paid for this is high complexity, both in the handling of the models and in the implementation of the resulting estimators. Chapter 7 is entirely devoted to Gibbs-Markov models.

3.7. Choice of criteria

The Bayesian approach brings inversion down to the determination of an *a posteriori* law. Since we cannot envisage calculating such laws completely, we content ourselves with looking for a point estimator, which is often the maximum *a posteriori* one. There are alternatives (*marginal maximum a posteriori*, *mean a posteriori*, etc.) but it is important to assess the consequences of such a choice carefully and, if necessary, think about alternatives.

It is reasonable to raise the question of the necessity for the solution to be continuous with respect to the data and, consequently, the need for convexity of the regularization criteria. While quadratic and entropy approaches are well known for making inverse problems well-posed, the minimization of a non-convex functional cannot guarantee that the solution will be continuous: a small variation in the data can induce a “jump” from one valley to another and thus a loss of continuity. However, in many

problems, these transitions are not only desirable but necessary to restore discontinuities, edges, interfaces, bright spots, etc. without limits in terms of spatial resolution. We can shed a different light on this problem by noting that certain non-convex criteria introduced in imaging have an equivalent expression implying *hidden variables*. In this case, the problem leaves convex analysis and incorporates a measure of combinatory analysis or hypothesis testing, which comes more under decision theory than estimation. Bayesian analysis remains pertinent in this combined detection-estimation context. Much recent work has followed this direction, combining several levels of variables, mixing low- and high-level descriptions, or data acquired by different experimental means. It is in this sense that the conventional concepts of regularization, such as continuity with respect to the data, are not completely appropriate and an effort should be made to extend them.

3.8. The linear, Gaussian case

The Gaussian laws associated with linear direct models provide a linear estimation structure and thus a very convenient algorithmic framework. However, they only allow us to incorporate crude information, basically limited to second order characteristics. Thus, in standard regularization theory [TIT 85], the choice of a quadratic term for fidelity to the data: $\mathcal{G}(\mathbf{y} - \mathbf{A}\mathbf{x}) = \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_{\mathbf{P}}^2$ is equivalent to choosing a Gaussian distribution for the noise: $q(\mathbf{b} | \mathbf{R}_b) \sim \mathcal{N}(0, \mathbf{R}_b)$, with $\mathbf{R}_b \propto \mathbf{P}^{-1}$. Similarly, choosing quadratic penalization: $\mathcal{F}(\mathbf{x}) = \|\mathbf{D}_k \mathbf{x}\|^2$ is also equivalent to choosing a Gaussian prior distribution for the object: $p(\mathbf{x} | \mathbf{R}_x) \sim \mathcal{N}(0, \mathbf{R}_x)$, with $\mathbf{R}_x \propto (\mathbf{D}_k^T \mathbf{D}_k)^{-1}$, assuming, of course, that the matrix $\mathbf{D}_k^T \mathbf{D}_k$ is defined as positive. Deterministic “linear-quadratic” regularization is thus rigorously equivalent to Gaussian linear estimation and the solution, which is explicit, is given by equations (2.11) and (2.12):

$$\hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{R}_b^{-1} \mathbf{A} + \mathbf{R}_x^{-1})^{-1} \mathbf{A}^T \mathbf{R}_b^{-1} \mathbf{y}, \quad (3.10)$$

$$= \mathbf{R}_x \mathbf{A}^T (\mathbf{A} \mathbf{R}_x \mathbf{A}^T + \mathbf{R}_b)^{-1} \mathbf{y}, \quad (3.11)$$

and has the remarkable characteristic of being a linear function of data \mathbf{y} . This “linear-quadratic” or linear Gaussian inversion holds a dominant position in inversion problems and it is a common reaction to say “*inverse problems aren’t complicated; you just need to smooth the data before doing the inversion*”. This way of seeing things is not wrong and is, in fact, sufficient for many problems but it is limiting; it stops us from going further and induces a cascading scheme – linear filtering of a generalized inverse solution – that is only justified in the “linear-quadratic” framework.

3.8.1. Statistical properties of the solution

Solution (3.10) is, in the Gaussian case, the mode, the mean and the median of the *a posteriori* probability distribution (3.5) all at once. It minimizes several very

commonly used cost criteria, in particular the *mean quadratic error*. Obviously, in this case, we are talking about a *mean with respect to the a posteriori distribution*, but many physicists and engineers only know the mean square error (MSE) defined as a *mean with respect to the direct distribution* (3.6). It is therefore useful to study the MSE, which is the sum of the bias energy and the trace of the covariance matrix: $MSE(\hat{\mathbf{x}}) = \|\mathbf{E}(\hat{\mathbf{x}}) - \mathbf{x}_0\|^2 + \text{trace Cov}(\hat{\mathbf{x}})$, designating the “true” solution by \mathbf{x}_0 . For the sake of simplicity, we will assume that the noise is stationary and white: $\mathbf{R}_b = \sigma_b^2 \mathbf{I}$ and that we can write $\mathbf{R}_x = \sigma_x^2 (\mathbf{D}^T \mathbf{D})^{-1}$. We thus have $\alpha = \sigma_b^2 / \sigma_x^2$.

The expectation of regularized solution (2.11), for direct distribution (3.6), can be written:

$$\begin{aligned} \mathbf{E}(\hat{\mathbf{x}}) &= \mathbf{E}((\mathbf{A}^T \mathbf{A} + \alpha \mathbf{D}^T \mathbf{D})^{-1} \mathbf{A}^T (\mathbf{A} \mathbf{x}_0 + \mathbf{b})) \\ &= (\mathbf{A}^T \mathbf{A} + \alpha \mathbf{D}^T \mathbf{D})^{-1} \mathbf{A}^T \mathbf{A} \mathbf{x}_0. \end{aligned}$$

Thus, for the bias to be zero ($\mathbf{E}(\hat{\mathbf{x}}) - \mathbf{x}_0 = 0$), we would need $\alpha = 0$, i.e., we must not regularize! The bias energy is:

$$\|\mathbf{E}(\hat{\mathbf{x}}) - \mathbf{x}_0\|^2 = \|((\mathbf{A}^T \mathbf{A} + \alpha \mathbf{D}^T \mathbf{D})^{-1} \mathbf{A}^T \mathbf{A} - \mathbf{I}) \mathbf{x}_0\|^2,$$

an increasing function of α , that equals zero and has a zero derivative at $\alpha = 0$ and that tends towards $\|\mathbf{x}_0\|^2$ when $\alpha \rightarrow \infty$.

The covariance matrix of the solution can be written:

$$\begin{aligned} \text{Cov}(\hat{\mathbf{x}}) &= \mathbf{E}((\hat{\mathbf{x}} - \mathbf{E}(\hat{\mathbf{x}})) (\hat{\mathbf{x}} - \mathbf{E}(\hat{\mathbf{x}}))^T) \\ &= \sigma_b^2 (\mathbf{A}^T \mathbf{A} + \alpha \mathbf{D}^T \mathbf{D})^{-1} \mathbf{A}^T \mathbf{A} (\mathbf{A}^T \mathbf{A} + \alpha \mathbf{D}^T \mathbf{D})^{-1}. \end{aligned}$$

To calculate its trace, we assume that matrices \mathbf{A} and \mathbf{D} have the same singular vectors¹, so that we have the factorizations:

$$\mathbf{A}^T \mathbf{A} = \mathbf{U} \mathbf{\Lambda}_a^2 \mathbf{U}^T \quad \text{and} \quad \mathbf{D}^T \mathbf{D} = \mathbf{U} \mathbf{\Lambda}_d^2 \mathbf{U}^T,$$

where $\mathbf{\Lambda}_a$ and $\mathbf{\Lambda}_d$ are diagonal matrices composed respectively of the singular values $\lambda_a(k)$ of \mathbf{A} and $\lambda_d(k)$ of \mathbf{D} , $k = 1, 2, \dots, M$. We thus obtain:

$$\text{trace Cov}(\hat{\mathbf{x}}) = \sigma_b^2 \sum_{k=1}^M \frac{\lambda_a^2(k)}{(\lambda_a^2(k) + \alpha \lambda_d^2(k))^2},$$

a strictly decreasing function of α , tending towards zero when $\alpha \rightarrow \infty$.

1. This is the case, for example, if \mathbf{D} is the identity matrix, or if \mathbf{A} and \mathbf{D} are two circulant matrices, such as those we will encounter in Chapter 4.

Thus, there is an “optimum”, strictly positive, value of α , that makes the MSE minimum. It is worth noting, however, that it depends on the true solution \mathbf{x}_0 through the bias energy and that efforts to find it would therefore be in vain. Also note that an approach, frequent in statistics, consisting of looking for estimators without bias and, if a degree of freedom remains, with minimum variance², leads to the generalized inverse solution, the MSE of which is:

$$MSE(\hat{\mathbf{x}}^{\text{GI}}) = \text{trace Cov}(\hat{\mathbf{x}}^{\text{GI}}) = \sigma_b^2 \sum_{k=1}^M \frac{1}{\lambda_a^2(k)}.$$

This can be considerable when some singular values $\lambda_a(k)$ are small, which is precisely the case in discretized and ill-conditioned problems. It can thus be said that, in terms of MSE, regularization consists of voluntarily introducing a bias in order to considerably reduce the variance of the solution.

3.8.2. Calculation of marginal likelihood

The linear, Gaussian case is one of the few that allow an explicit calculation of the marginal likelihood of equation (3.3), used to adjust the values of hyperparameters θ . When these are limited to the variances σ_b^2 and σ_x^2 for example (or to the pair σ_b^2 and $\alpha = \sigma_b^2/\sigma_x^2$), we have:

$$p(\mathbf{x}, \mathbf{y} | \sigma_x^2, \sigma_b^2) = (2\pi \sigma_b^2)^{-N/2} (2\pi \sigma_x^2)^{-M/2} |\mathbf{D}^T \mathbf{D}|^{1/2} e^{-\mathcal{Q}/2\sigma_b^2},$$

$$\text{where } \mathcal{Q} = (\mathbf{y} - \mathbf{A}\mathbf{x})^T (\mathbf{y} - \mathbf{A}\mathbf{x}) + \alpha \mathbf{x}^T \mathbf{D}^T \mathbf{D} \mathbf{x}.$$

To calculate the ordinary, or marginal, likelihood of α and σ_b^2 , we have to “integrate \mathbf{x} out of the problem”. To prepare this integration, a perfect square is conventionally made to appear in \mathcal{Q} :

$$\mathcal{Q} = (\mathbf{x} - \hat{\mathbf{x}})^T (\mathbf{A}^T \mathbf{A} + \alpha \mathbf{D}^T \mathbf{D}) (\mathbf{x} - \hat{\mathbf{x}}) + \mathcal{S}(\alpha),$$

with $\mathcal{S}(\alpha) = \mathbf{y}^T (\mathbf{y} - \mathbf{A}\hat{\mathbf{x}})$, which leads to a Gaussian integral:

$$\begin{aligned} p_{\mathbf{y}} | \alpha, \sigma_b^2 &= \int p_{\mathbf{x}, \mathbf{y}} | \sigma_x^2, \sigma_b^2 d\mathbf{x} \\ &= (2\pi \sigma_b^2)^{-N/2} \alpha^{M/2} |\mathbf{D}^T \mathbf{D}|^{1/2} |\mathbf{A}^T \mathbf{A} + \alpha \mathbf{D}^T \mathbf{D}|^{-1/2} e^{-\mathcal{S}(\alpha)/2\sigma_b^2}. \end{aligned}$$

2. This strategy has no serious basis. Good asymptotic properties (when $N \rightarrow \infty$) are often mentioned for these estimators *without bias and with minimum variance*, but an estimator such as (3.10) also converges towards \mathbf{x}_0 in the same conditions, and faster, since for any finite N , its MSE is smaller.

By switching to logarithms, we obtain the log-marginal likelihood:

$$L(\alpha, \sigma_b^2) = \frac{M}{2} \log \alpha - \frac{N}{2} \log(2\pi\sigma_b^2) + \frac{1}{2} \log |\mathbf{D}^T \mathbf{D}| - \frac{1}{2} \log |\mathbf{A}^T \mathbf{A} + \alpha \mathbf{D}^T \mathbf{D}| - \frac{\mathcal{S}(\alpha)}{2\sigma_b^2}.$$

If this likelihood is sufficiently peaked, we can then satisfy ourselves with finding the $(\hat{\alpha}, \hat{\sigma}_b^2)$ pair that maximizes $L(\alpha, \sigma_b^2)$. We have:

$$\frac{\partial L}{\partial \sigma_b^2} = -\frac{N}{2\sigma_b^2} + \frac{\mathcal{S}(\alpha)}{2\sigma_b^4} = 0 \quad \implies \quad \hat{\sigma}_b^2 = \frac{\mathcal{S}(\alpha)}{N},$$

the “usual” estimator for variance. It is, however, difficult to maximize L as a function of α . We will thus content ourselves with finding $\hat{\alpha}$ by exploring a discrete grid, since the result $\hat{x}(\alpha)$ is, in general, sensitive only to variations of the order of magnitude of α [FOR 93, THO 91].

3.8.3. Wiener filtering

The “linear-quadratic” framework is the only one that allows a statistical interpretation to be given in the *infinite dimension* problem [FRA 70]:

$$y = Ax + b, \quad x \in \mathcal{X}, \quad y \in \mathcal{Y}. \quad (3.12)$$

For this, we assume that the functions x , y and b appearing in equation (3.12) are particular *trajectories* or *realizations*, of *stochastic processes* X , Y and B , linked by an analog relation³:

$$Y = AX + B. \quad (3.13)$$

If the zero-mean process X depends on a variable r , its *covariance function* is defined as $\Gamma_X(r, r') = E(X(r)X(r'))$, and we assume that the functions x , trajectories of the process X , belong to a Hilbert space \mathcal{X} and the functions y and b , the respective trajectories of Y and B , belong to the same Hilbert space \mathcal{Y} (which may be distinct from \mathcal{X}). The covariance (function) of X can thus be considered as the kernel of an operator R_X defined on the space \mathcal{X} :

$$(R_X \phi)(r) = \int \Gamma_X(r, r') \phi(r') dr', \quad \phi \in \mathcal{X}.$$

The inverse problem is to estimate a realization x of X , given the observation data of the realization y of Y and probabilistic prior knowledge on the processes X and B .

3. Here, for the sake of simplicity, we also assume that processes X , Y and B have zero mean. This hypothesis is not restrictive as, if they do not, the processes can always be centered and, thanks to the linearity of A , relation (3.13) remains true for the centered processes.

In the special case where X is a Gaussian process (or any linear transformation – such as the derivative – of a Gaussian process), the *a priori* probability law for X can be written symbolically⁴:

$$p_X(x) \propto \exp \left\{ -\frac{1}{2} \langle x, R_X^{-1} x \rangle_{\mathcal{X}} \right\}.$$

If we take the hypothesis that the noise process B is additive, white and Gaussian of variance σ^2 , the *a posteriori* law can be written:

$$p_X(x | Y = y) \propto \exp \left\{ -\frac{1}{2\sigma^2} \left(\|y - Ax\|_{\mathcal{Y}}^2 + \sigma^2 \langle x, R_X^{-1} x \rangle_{\mathcal{X}} \right) \right\}.$$

The best estimator of x , given the observation data y , depends on the choice of the optimality criterion but, in this case, if we choose the maximum of the *a posteriori* law or the MSE and if we factorize the covariance operator according to:

$$R_X = (C^* C)^{-1}, \quad (3.14)$$

the solution minimizes the criterion $\|y - Ax\|_{\mathcal{Y}}^2 + \sigma^2 \|Cx\|_{\mathcal{X}}^2$. It follows that $\hat{x} = Gy$, where G is given by (2.7) with $\alpha = \sigma^2$. Moreover, if we define the operator $R_B = \sigma^2 Id$, where Id is the identity operator in \mathcal{Y} (R_B is the covariance operator of white noise), then G can also be written in the form:

$$G = R_X A^* (A R_X A^* + R_B)^{-1}, \quad (3.15)$$

which is the form of the *Wiener filter*. Put differently, the Tikhonov regularizer (2.7) is analogous to a Wiener filter in the case of white noise, provided that the constraint operator C that appears in it is linked to the covariance operator R_X by relation (3.14). Note, however, that second order ergodic processes have trajectories of finite power but infinite energy: \mathcal{X} is not a summable square function space.

Equation (3.15) differs from the usual expression for a Wiener filter expressed in the Fourier domain. In fact, the expression above is more general. We will find the usual formulation again in Chapter 4, by taking advantage of additional hypotheses such as the convolutional structure of operator A and the weak stationarity (second order) of random processes X and B .

In contrast, in the case of non-quadratic functionals \mathcal{G} or \mathcal{F} , the minimization of criterion (2.5) does not have a systematic statistical interpretation. In substance, the difficulty comes from the fact that the mathematical quantity characterizing the probability of a random process indexed on a space of finite dimension is, in this case, a set of functions having no direct relation with (2.5) and not allowing the likelihood function to be defined naturally.

4. In fact, the law of a process is given by the joint law of the n random variables $X(r_1), X(r_2), \dots, X(r_n), \forall n \in \mathbb{N}, \forall (r_1, r_2, \dots, r_n) \in \mathbb{R}^n$.

3.9. Bibliography

- [BER 94] BERNARDO J. M., SMITH A. F. M., *Bayesian Theory*, Wiley, Chichester, UK, 1994.
- [BES 86] BESAG J. E., “On the statistical analysis of dirty pictures (with discussion)”, *J. R. Statist. Soc. B*, vol. 48, num. 3, p. 259-302, 1986.
- [BRO 86] BROWN L. D., *Foundations of Exponential Families*, vol. 9, Hayward, CA, IMS Lecture Notes, Monograph Series edition, 1986.
- [COX 61] COX R., *The Algebra of Probable Inference*, Johns Hopkins University Press, Baltimore, MD, 1961.
- [DEM 77] DEMPSTER A. P., LAIRD N. M., RUBIN D. B., “Maximum likelihood from incomplete data via the EM algorithm”, *J. R. Statist. Soc. B*, vol. 39, p. 1-38, 1977.
- [DEM 89] DEMOMENT G., “Image reconstruction and restoration: overview of common estimation structure and problems”, *IEEE Trans. Acoust. Speech, Signal Processing*, vol. ASSP-37, num. 12, p. 2024-2036, Dec. 1989.
- [FOR 93] FORTIER N., DEMOMENT G., GOUSSARD Y., “GCV and ML methods of determining parameters in image restoration by regularization: fast computation in the spatial domain and experimental comparison”, *J. Visual Comm. Image Repres.*, vol. 4, num. 2, p. 157-170, June 1993.
- [FRA 70] FRANKLIN J. N., “Well-posed stochastic extensions of ill-posed linear problems”, *J. Math. Anal. Appl.*, vol. 31, p. 682-716, 1970.
- [GAS 92] GASSIAT E., MONFRONT F., GOUSSARD Y., “On simultaneous signal estimation and parameter identification using a generalized likelihood approach”, *IEEE Trans. Inf. Theory*, vol. 38, p. 157-162, Jan. 1992.
- [GEM 84] GEMAN S., GEMAN D., “Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images”, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, num. 6, p. 721-741, Nov. 1984.
- [JAY 82] JAYNES E. T., “On the rationale of maximum-entropy methods”, *Proc. IEEE*, vol. 70, num. 9, p. 939-952, Sep. 1982.
- [JEF 39] JEFFREYS, *Theory of Probability*, Oxford Clarendon Press, Oxford, UK, 1939.
- [KAS 94] KASS R. E., WASSERMAN L., Formal Rules for Selecting Prior Distributions: A Review and Annotated Bibliography, Technical report no. 583, Department of Statistics, Carnegie Mellon University, 1994.
- [LIT 83] LITTLE R. J. A., RUBIN D. B., “On jointly estimating parameters and missing data by maximizing the complete-data likelihood”, *Amer. Statist.*, vol. 37, p. 218-220, Aug. 1983.
- [MAR 87] MARROQUIN J. L., MITTER S. K., POGGIO T. A., “Probabilistic solution of ill-posed problems in computational vision”, *J. Amer. Stat. Assoc.*, vol. 82, p. 76-89, 1987.
- [POL 92] POLSON N. G., “On the expected amount of information from a non-linear model”, *J. R. Statist. Soc.*, vol. 54, num. B, p. 889-895, 1992.

- [ROB 97] ROBERT C. P., *The Bayesian Choice. A Decision-Theoretic Motivation*, Springer Texts in Statistics, Springer Verlag, New York, NY, 1997.
- [TAR 87] TARANTOLA A., *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*, Elsevier Science Publishers, Amsterdam, The Netherlands, 1987.
- [THO 91] THOMPSON A., BROWN J. C., KAY J. W., TITTERINGTON D. M., "A study of methods of choosing the smoothing parameter in image restoration by regularization", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-13, num. 4, p. 326-339, Apr. 1991.
- [TIT 85] TITTERINGTON D. M., "Common structure of smoothing techniques in statistics", *Int. Statist. Rev.*, vol. 53, num. 2, p. 141-170, 1985.

PART II

Deconvolution

Chapter 4

Inverse Filtering and Other Linear Methods

4.1. Introduction

Many physical systems can, as a good approximation, be considered as linear with respect to their inputs and invariant under translation of the variables of the input signal (e.g. one time variable or two space variables). They can then be modeled using convolution. For this reason, *deconvolution* is a generic inverse problem that comes into many applications: non-destructive testing, geophysics, medical imaging, astronomy, communications, etc. Several of these applications have chapters devoted to them in Part II of this book.

In this chapter, we present the specificities of convolution, and then the techniques for *linear* deconvolution. This is equivalent to *spectral equalization*; and therefore cannot restore frequency components that have been cut out by the observation system; it can simply compensate for some of the attenuations introduced. The resolution of linear solutions is fundamentally limited by the spectral content of the data. To go beyond this limit, we need to use nonlinear techniques, which require additional information or hypotheses on the object to be restored and will be seen in later chapters.

The main interest of linear solutions lies in their speed and ease of use – although implementation techniques have evolved greatly in the past 20 years – and in their remarkable robustness with respect to hypotheses on the object, precisely because of

their simplicity¹. These solutions thus serve as references to which all more elaborate solutions are compared. They can also be intermediate means in calculations, as in the semi-quadratic methods presented in Chapter 6. So we will pay particular attention to these questions of implementation here.

Finally, tuning linear solutions is relatively simple: it consists of choosing the second-order characteristics of the signals under consideration. It is possible to use the numerous results on the second-order modeling of signals and specify their correlation function, or their spectral density, or a difference equations model. As a last resort, a single scalar parameter, the coefficient of regularization, can be adjusted. In this case, it is essentially a ratio between the input powers and the noise, and thus a *signal-to-noise ratio* (SNR), one of the most familiar notions we could wish for. Furthermore, there are simple, proven techniques for choosing the regularization parameter, which are presented in section 3.8 and section 2.3.3.

4.2. Continuous-time deconvolution

In a continuous-time setting, the convolution equation can be written:

$$y(t) = (Ax)(t) = (h \star x)(t) = \int_{\mathbb{R}} x(t - t') h(t') dt', \quad t \in \mathbb{R}. \quad (4.1)$$

In this equation, x is the input to be restored, h the kernel of the convolution operator A or the impulse response (IR) of the observation system², and y represents the observed data. This is a special case of equation (1.9) of Chapter 1.

4.2.1. Inverse filtering

If we restrict ourselves to the case where x , y and h are generalized functions or tempered distributions possessing Fourier transforms (FT), one solving method seems to impose itself. By Fourier transformation, equation (4.1) becomes:

$$\widehat{y}(\nu) = \widehat{x}(\nu) \widehat{h}(\nu), \quad \nu \in \mathbb{R}, \quad (4.2)$$

whence the solution by *inverse filtering*:

$$\widehat{x}(\nu) = \widehat{y}(\nu) / \widehat{h}(\nu) \quad \text{and} \quad x(t) = \mathcal{F}^{-1}\{\widehat{x}(\nu)\}. \quad (4.3)$$

1. It is always wise to remember the old law of cybernetics, somewhat neglected nowadays, that says “[...] the quantity of information required is a measure of the machine’s tendency to go wrong” [WIE 71].

2. In this chapter, it is assumed to be known. See also Chapter 8, which deals with situations where this is not the case.

As the transform $\widehat{x}(\nu)$ determines $x(t)$ completely, it is sufficient, in principle, to know $\widehat{y}(\nu)$ and $\widehat{h}(\nu)$. However, things are not that simple [ARS 66].

If we take the situation of a convolution algebra \aleph , the necessary and sufficient condition for $\forall y \in \aleph, y = h \star x$ to have a solution in \aleph is that h possesses a convolution inverse h^{-1} such that $h \star h^{-1} = \delta$. We then obtain the solution immediately: $y \star h^{-1} = x$. However, h^{-1} does not always exist if h is, for example, a distribution. Even worse, h^{-1} may exist only for certain functions y . Moreover, if \aleph admits zero divisors, there is an infinite number of solutions.

We should not lose sight of the purely formal aspect of equation (4.3). For it to have meaning, $1/\widehat{h}(\nu)$ has to exist and be a tempered distribution. Thus $\widehat{h}(\nu)$ has to be a function that is never zero for any value of ν and does not tend towards zero at infinity faster than any power of $1/\nu$. These are very strict conditions that are never satisfied in practice. What is more, for most real observation systems, $|\widehat{h}(\nu)| \rightarrow 0$ when $\nu \rightarrow \infty$ and the solution given by (4.3) is inevitably *unstable*. The experimental data always contain errors:

$$y(t) = \int_{\mathbb{R}} x(t-t') h(t') dt' + b(t), \quad (4.4)$$

and there is no reason why the FT of the error, or the measurement noise, $b(t)$, should decrease at infinity like $\widehat{h}(\nu)$. Finally, introducing boundedness hypotheses on the support of $\widehat{x}(\nu)$ or $x(t)$ does not solve anything:

- if the class of inputs is restricted to *limited bandwidth* functions, it is necessary, if a solution is to be found, for the support of the input spectrum to be included in the frequency band where $\widehat{h}(\nu)$ is known, which is the classic Rayleigh criterion used in instrumentation. If we consider the instability due to small values of h , it could seem useful to assume a bounded support for FT x if such a support was much narrower than the support of FT h and restricted to regions with high SNR. However, this would imply having an observation system of much higher resolution than that needed for signal recovery, which is a rare situation of little interest;

- if, conversely, the class of inputs is limited to functions with *bounded support*, we know that $\widehat{x}(\nu)$ is an *analytical function* that can, since Weierstrass, be extended starting from our knowledge of it on an interval. However, these analytical extension methods require the function and all its derivatives to be known exactly, and their implementation is too sensitive to measurement errors [KHU 77].

The example of a moving average (MA) filter:

$$y(t) = \frac{1}{T} \int_{t-T}^t x(t') dt', \quad (4.5)$$

that will be used throughout this chapter and whose frequency response is shown in Figure 4.1 accumulates several difficulties. Its frequency response is a cardinal sine

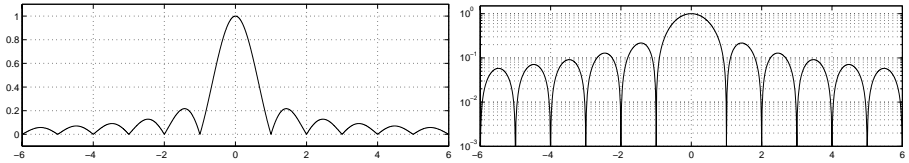


Figure 4.1. *Modulus of transfer function of the MA filter defined by (4.5) ($T = 1$). Scale is linear on the left, logarithmic on the right*

that attenuates the high frequencies – they are inevitably drowned in the noise – and, what is more, is canceled out for frequencies that are multiples of $1/T$. Inverse filtering cannot provide a solution.

4.2.2. Wiener filtering

The simplicity of form of equation (4.3) is deceptive and we find ourselves in the heart of the difficulties described in Chapter 1. If x , y and h are assumed to be functions with summable squares – a hypothesis that is very often acceptable from a physical point of view – the associated operator A is compact and bounded but its image $\text{Im } A$ is not closed. The deconvolution problem is thus ill posed and regularization methods are needed. In the framework of this chapter, we will limit ourselves to penalty methods using quadratic regularization functionals, and will choose the Tikhonov regularizer (2.7) or the Wiener filter (3.15).

The regularized solutions to the continuous-time deconvolution problem date back to Wiener's work on optimal filtering in the 1950s, which was applied immediately to solve equation (4.4) (e.g. in geophysics [ROB 54]). In Wiener's approach, presented in Chapter 3, the input signal and the noise are modeled by second-order random processes, characterized by their covariance function. In the stationary case, the covariance functions are invariant under translation and are thus functions of a single variable, the correlation functions $r^x(t)$ and $r^b(t)$ which can also be specified by their FT, the power spectral densities $S_x(\nu)$ and $S_b(\nu)$.

With these hypotheses of stationarity and operator A being a convolution, the general form of Wiener filter (3.15) takes a particularly simple expression by passing into the Fourier domain as its frequency transfer function can be written:

$$\widehat{g}(\nu) = \frac{\widehat{h}^*(\nu) S_x(\nu)}{|\widehat{h}(\nu)|^2 S_x(\nu) + S_b(\nu)} = \frac{1}{\widehat{h}(\nu)} \frac{|\widehat{h}(\nu)|^2 S_x(\nu)}{|\widehat{h}(\nu)|^2 S_x(\nu) + S_b(\nu)}. \quad (4.6)$$

Thus, we see that the Wiener filter is, in fact, a cascade of an inverse filter – of transfer function $1/\widehat{h}(\nu)$ – and a stabilizing filter. The latter has a practically unit transfer in

frequency bands with a high SNR, i.e., when $|\widehat{h}(\nu)|^2 S_x(\nu) \gg S_b(\nu)$ and the whole then behaves like an inverse filter. On the other hand, when the SNR becomes zero in other frequency bands, the transfer becomes zero, as does that of the stabilizing filter, thus controlling the divergence of the inverse filter. Thus we see clearly why Wiener's solution, or "linear quadratic" regularization can only manage to equalize the spectral content of the data in the band of frequencies where the SNR is sufficient.

Filter (4.6) is generally not realizable (being non-causal), which is obviously a nuisance for implementation using an analog electronic device. Many authors have thus taken an interest in obtaining solutions under constraints, e.g. of causality or finite duration. The first causal solutions used spectral factorization of the signals under consideration, i.e., they represented a signal as the output from a filter with white noise as the input [VAN 68].

In the 1960s, such representations were replaced by state representations (see section 4.5), which not only enabled the implementation problems for constrained versions of the Wiener filter to be solved but, above all, allowed the results to be easily extended to the non-stationary case. Kalman-Bucy filtering [KAL 60] was born, and was rapidly applied to deconvolution, e.g. in geophysics, in the 1970s. The reference [BAY 70] gives a good introduction to continuous-time Kalman filtering together with examples of its use in geophysics, illustrated by the processing of analog data.

Nowadays, with the development of personal computers, the signals are most often discretized as soon as they are acquired. So, before presenting the corresponding methods, we are going to study the consequences of this discretization on the nature of the problem.

4.3. Discretization of the problem

As the signal to be deconvolved y is discretized, the input to be restored x is also discretized most of the time, which leads to the processing of a discrete-discrete inverse problem (see Chapter 1). The discretization of the output often comprises an integration (as, for example, with a CCD camera). The developments that follow, concerning the case of simple sampling, can be mostly used as they stand, the integration being included in the system response h .

4.3.1. Choice of a quadrature method

To discretize the problem, the input to be restored $x \in \mathcal{X}$ (see section 1.3) is decomposed on a family of functions $\{g_m\}$:

$$x(t) = \sum_{m=1}^M x_m g_m(t) + x^*(t), \quad (4.7)$$

where the x_m are the decomposition coefficients and x^* is a residual truncation error term.

If the observed signal y is *sampled* regularly at a finite number, N , of instants $t_n = n\Delta t$, we can write:

$$y_n = y(n\Delta t) = \sum_{m=1}^M h_{n,m} x_m + b_n, \quad n = 1 \dots, N, \quad (4.8)$$

where $h_{n,m} = \int g_m(n\Delta t - t) h(t) dt$, and $b_n = b(t_n)$ represents the sum of the measurement noise and the filtered truncation error $\int x^*(n\Delta t - t) h(t) dt$.

It is common practice to decompose x on a Δt -shifted kernel basis $g_m(t) = g(t - m\Delta t)$, which allows the problem to be written in the form of a discrete-time convolution:

$$y_n = \sum_{m=1}^M \tilde{h}_{n-m} x_m + b_n, \quad n = 1, 2, \dots, N,$$

but it should be noted that the discrete IR is then sampled from a filtered version of h :

$$\tilde{h}_k = (h \star g)(k\Delta t). \quad (4.9)$$

Hence, in general, *the cut-off frequency of the discrete-time system is different from that of the continuous-time system.*

When the basis kernel g is an interval indicator of width Δt , the coefficients x_m and \tilde{h}_k are the mean values of the input and the IR, respectively, over intervals of width Δt . When the basis kernel is a cardinal sine of pseudo-period Δt , convolution (4.9) truncates \tilde{h} to the band $(-\Delta t/2, \Delta t/2]$. If h is of limited bandwidth included in this interval, Shannon's condition is verified and $\tilde{h}_k = h(k\Delta t)$. Otherwise, the cut-off frequency of the discrete system is lower than that of the continuous problem, which in fact corresponds to a regularization by dimension control.

Another element to be taken into consideration is the time-domain truncation of the IR \tilde{h}_k , which is indispensable to obtain a finite-dimension problem. From this point of view, choosing a cardinal sine, which has a very slow decay, as the basis kernel leads to greater truncation effects than the use of indicators. Intermediate choices exist, of course, e.g. spline functions or prolate spheroidal functions.

By concatenating the N equations (4.8) describing the relations among the values of the observed signal y_n , those of the signal to be restored x_n and the coefficients of the IR h_n (direct problem), we obtain a linear $N \times M$ system of the form (1.3)

$$y = \mathbf{H} x + b. \quad (4.10)$$

We know (see Chapter 2) that the difficulties of inverting such a system are connected with the conditioning of matrix \mathbf{H} , which we will study after having clarified its structure.

4.3.2. Structure of observation matrix \mathbf{H}

In what follows, the notation h_k will be used to mean the IR of the discrete problem and we will take it that its *effective* support – i.e., the domain for which the coefficients h_k have significant values – is smaller than the time interval $N\Delta t$ over which y is observed. As this IR is not necessarily causal, we write: $\mathbf{h} = [h_{-Q}, \dots, h_0, \dots, h_{+P}]^T$. These $P + Q + 1$ significant coefficients are thus such that: $P + Q + 1 < N$. The discrete convolution equation is then:

$$y_n = \sum_{p=-Q}^P h_p x_{n-p} + b_n, \quad n = 1, 2, \dots, N, \quad (4.11)$$

The deconvolution problem is often considered to consist of estimating a vector $\mathbf{x} = [x_1, \dots, x_N]^T$ from an observed vector $\mathbf{y} = [y_1, \dots, y_N]^T$ having the same support. Unfortunately, equation (4.11) shows that it is then impossible to establish a relationship like (4.10) between these two vectors, because of *boundary effects*. Actually, if we write out the components of matrix equation (4.10), we obtain:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} h_P & \dots & h_0 & \dots & h_{-Q} & 0 & \dots & 0 \\ 0 & h_P & \dots & h_0 & \dots & h_{-Q} & 0 & \dots \\ & \ddots & \ddots & & \ddots & & \ddots & \ddots \\ & & \dots & 0 & h_P & \dots & h_0 & \dots & h_{-Q} & 0 \\ 0 & & \dots & 0 & h_P & \dots & h_0 & \dots & h_{-Q} \end{bmatrix} \begin{bmatrix} x_{-P+1} \\ \vdots \\ x_0 \\ x_1 \\ \vdots \\ x_N \\ x_{N+1} \\ \vdots \\ x_{N+Q} \end{bmatrix} \quad (4.12)$$

and the problem is undetermined because $M = \dim \mathbf{x} = N + P + Q > \dim \mathbf{y} = N$.

Matrix \mathbf{H} thus has a *Toeplitz form* with a band structure, as $P + Q + 1 < N$. In the case of an image (2D problem), discrete convolution (4.11) involves summing over two indices and it is usually brought down to the same matrix expression (4.10) by constructing vectors \mathbf{x} and \mathbf{y} by a lexicographical scan – e.g. line by line from left to right and top to bottom – of the tables containing the values of the pixels of the object to be restored and of the blurred observed image [HUN 73, JAI 89]. Matrix \mathbf{H} deduced from this has a block-Toeplitz structure and each block is itself Toeplitz. To

simplify the notation, we will call them Toeplitz-block-Toeplitz matrices from now on.

It should first of all be noted that the structure of \mathbf{H} allows the product $\mathbf{H}\mathbf{x}$ to be calculated quickly. It is always possible to complete this matrix with additional rows so as to obtain a *circulant* square matrix \mathbf{C}_h of order $M = N + P + Q$. A circulant matrix is a Toeplitz matrix entirely defined by its first row: a row is obtained from the previous one by circular permutation. Circulant square matrices can be factorized in the Fourier basis [HUN 71]:

$$\mathbf{C}_h = \mathbf{W}^* \mathbf{\Lambda}_h \mathbf{W}, \quad (4.13)$$

where $\mathbf{W}_{k,\ell} = e^{-2j\pi(k-1)(\ell-1)/L} / \sqrt{M}$ is the unitary Fourier matrix of order M and $\mathbf{\Lambda}_h$ is a diagonal matrix in which the diagonal elements are obtained by discrete Fourier transform (DFT calculable by *fast Fourier transform* — FFT) of the first column of \mathbf{C}_h . The product $\mathbf{H}\mathbf{x}$ can then be obtained by the following operations:

- 1) DFT of \mathbf{x} and of the first column of \mathbf{C}_h ;
- 2) component by component product of the two DFTs obtained;
- 3) inverse DFT of result;
- 4) extraction of \mathbf{y} (of dimension N) in the resulting vector.

This technique, which uses zero-padding, is usually very advantageous in terms of computing cost. In 2D, it can be generalized by making use of the properties of circulant-block-circulant matrices (i.e., circulant by blocks and where the blocks themselves are circulant) with square blocks [HUN 73].

To return to boundary problems: if we define the vectors $\mathbf{x}_l = [x_{-P+1}, \dots, x_0]^T$ and $\mathbf{x}_r = [x_{N+1}, \dots, x_{N+Q}]^T$ to be the left and right “*borders*” of the unknown vector \mathbf{x} , linear system (4.12) can be written [NG 99]:

$$\mathbf{y} = \mathbf{H}_l \mathbf{x}_l + \mathbf{H}_c \mathbf{x} + \mathbf{H}_r \mathbf{x}_r + \mathbf{b}, \quad (4.14)$$

where from now on we note $\mathbf{x} = [x_1, \dots, x_N]^T$. \mathbf{H}_l is made up of the first P columns of \mathbf{H} , \mathbf{H}_r of the last Q columns, and \mathbf{H}_c designates the central part of \mathbf{H} .

There are then two cases:

1) Vectors \mathbf{x}_l and \mathbf{x}_r are unknown and we want to estimate them, thus trying to obtain $M = N + P + Q$ values from N equations. The resulting indetermination is not a problem in a regularized framework (see Chapter 2). On the other hand, the convolution matrix does not have a circulant structure, which leads to more costly inversion algorithms.

2) Vectors \mathbf{x}_l and \mathbf{x}_r are unknown and we do not want to estimate them. Thus we want to approximate equation (4.14) by a square $N \times N$ system. There are several approximations that correspond to boundary conditions added to the statement

of the problem. The prewindowing (respectively postwindowing) consists of taking \mathbf{x}_l (respectively \mathbf{x}_r) to be zero, which can be justified in certain situations (causal or transient phenomenon). Below, we describe two other boundary hypotheses that are very advantageous in terms of computing cost.

4.3.3. Usual boundary conditions

A much-used approximation involves taking the hypothesis that the signal to be restored is *N-periodic*: $\mathbf{x}_l = [x_{N-P+1}, \dots, x_N]^T$ and $\mathbf{x}_r = [x_1, \dots, x_Q]^T$. Equation (4.14) then becomes $\mathbf{y} = \mathbf{H}_p \mathbf{x} + \mathbf{b}$, where the new convolution matrix:

$$\mathbf{H}_p = [\mathbf{0}_{N \times (N-P)} \mid \mathbf{H}_l] + \mathbf{H}_c + [\mathbf{H}_r \mid \mathbf{0}_{N \times (N-Q)}] \quad (4.15)$$

is circulant and can thus be diagonalized in a Fourier basis: $\mathbf{H}_p = \mathbf{W}^* \mathbf{\Lambda}_h \mathbf{W}$. Its eigenvalues are obtained by DFT of the first column \mathbf{h}_p of \mathbf{H}_p , which is an *N*-periodic zero-padded RI. In the case of the non-causal RI \mathbf{h} introduced in section 4.3.2, it writes $\mathbf{h}_p = [h_0, \dots, h_P, 0, 0, \dots, 0, h_{-Q}, \dots, h_{-1}]^T$.

Another, less frequent, approximation imposes a “*mirror condition*” (also known as a Neumann boundary condition [NG 99]) on the boundaries or, in other words, assumes that $\mathbf{x}_l = [x_Q, x_{Q-1}, \dots, x_1]^T$ and $\mathbf{x}_r = [x_N, x_{N-1}, \dots, x_{N-P+1}]^T$. The new convolution matrix coming from equations (4.10) and (4.14) takes the form:

$$\mathbf{H}_m = [\mathbf{0}_{N \times (N-P)} \mid \mathbf{H}_l] \mathbf{J} + \mathbf{H}_c + [\mathbf{H}_r \mid \mathbf{0}_{N \times (N-Q)}] \mathbf{J} \quad (4.16)$$

where \mathbf{J} is the unit Hankel matrix, or reversal matrix, of dimensions $N \times N$. In this case, matrix \mathbf{H}_m is neither Toeplitz nor circulant, but “Toeplitz-plus-Hankel”. When the IR of the system is even ($h_k = h_{-k}$ and $P = Q$), these matrices can be factorized like circulant matrices but by discrete cosine transform (DCT).

These approximations can be immediately generalized to the 2D case [NG 99]. It is easy to see the interest here: as circulant or Toeplitz-plus-Hankel matrices can be diagonalized in a complex exponential or cosine basis, their factorization, which is a preparation for the subsequent inversion, has a reduced computing cost thanks to the DFT or DCT. It remains to be seen what consequences these approximations have on signals that satisfy neither a periodic nor a mirror boundary hypothesis. We will look at this later using an example.

4.3.4. Problem conditioning

The calculation of the conditioning of the problem is based on the study of the non-zero eigenvalues of matrix $\mathbf{H}^T \mathbf{H}$ (the problem of zero eigenvalues being settled by resorting to a generalized inverse). According to the boundary hypotheses, either

$\mathbf{H}^T \mathbf{H}$ is circulant or one of the two matrices $\mathbf{H} \mathbf{H}^T$ or $\mathbf{H}^T \mathbf{H}$ is Toeplitz (or Toeplitz-block-Toeplitz in 2D). We have seen in equation (4.13) that eigenvalues of circulant matrices are easy to calculate via DFT. For non-circulant Toeplitz matrices, however, there are only asymptotic results. In both cases, the results bring in the transfer function associated with the discrete IR:

$$\widehat{H}(\nu) = \sum_{p=-Q}^P h_p e^{-2j\pi p\nu}, \quad \nu \in [0, 1).$$

4.3.4.1. Case of the circulant matrix

When the signal to be restored is N -periodic, we have seen that the resulting convolution matrix \mathbf{H}_p is circulant of order N , and equation (4.13) shows that the eigenvalues of the normal matrix of the problem $\mathbf{H}^T \mathbf{H}$ form a regular sampling of the square of the modulus of \widehat{H} : $\lambda_h^2(k) = |\widehat{H}(k/N)|^2$, $0 \leq k \leq N-1$.

The conditioning of the matrix is thus directly linked to a framing of the values of the function $|\widehat{H}(\nu)|^2$: an example is given in Figure 4.2 for the case of the MA filter of equation (4.5). The problem was discretized with a sampling period $\Delta t = T/20$. Although none of the eigenvalues is identically zero in this example, the condition number c of problem (1.22) is very high: $c > 10^{10}$ and we can expect a few difficulties during generalized inversion, as we will see in section 4.3.5.

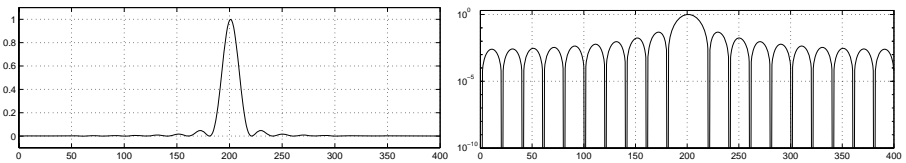


Figure 4.2. Spectrum of eigenvalues of the normal matrix $\mathbf{H}^T \mathbf{H}$ for the case of a signal deteriorated by the MA filter (4.5) (periodic boundary condition)

4.3.4.2. Case of the Toeplitz matrix

In the general case of equation (4.12), matrix $\mathbf{H} \mathbf{H}^T$ (which is of order N) is symmetric band-Toeplitz. It is always possible to include it in a circulant matrix of order $N + P + Q$ where the first row is a permutation of the autocorrelation of the IR completed by zeros and has eigenvalues $|\widehat{H}(k/(N + P + Q))|^2$. For large N , matrix $\mathbf{H} \mathbf{H}^T$ appears as a perturbation of the circulant matrix. Szegő's theorem [GRE 58] describes the convergence of the spectra of these two matrices towards each other, which is another way of saying that the eigenvalues of $\mathbf{H} \mathbf{H}^T$ tend to be evenly spread over

$|\widehat{H}(\nu)|^2$ when N tends towards infinity. It can thus be shown that the conditioning of the problem tends towards the ratio:

$$\frac{\max_{\nu \in [0,1)} |\widehat{H}(\nu)|^2}{\min_{\nu \in [0,1)} |\widehat{H}(\nu)|^2}.$$

4.3.4.3. Opposition between resolution and conditioning

We inevitably find ourselves faced with the following dilemma:

- either the IR has a limited spectrum and we choose a sampling period Δt small enough to respect Shannon's condition. There is no change in the cut-off frequency due to the sampling (see section 4.3) but the problem is very ill-conditioned as the function $|\widehat{H}(\nu)|^2$ takes very small values;

- or the sampling period is chosen big enough for there to be a change in the cut-off frequency of the system: the conditioning is better but we give up the possibility of reaching a high-resolution solution during deconvolution.

This situation is unusual in numerical analysis: choosing an increasingly fine discretization of equation (4.1) in order to reduce the quadrature error makes solving equation (4.10) more and more delicate. This is precisely the distinctive feature of ill-posed problems.

4.3.5. Generalized inversion

To illustrate the influence of poor conditioning on deconvolution, Figure 4.3 presents an example of a least squares solution (1.17) in the case of a signal filtered by the MA filter of equation (4.5). If the normal matrix has zero eigenvalues, the generalized inverse (1.19) behaves in a similar way. We observe, as predicted in Chapter 1 (section 1.3), a strong amplification of the noise components at the frequencies attenuated by the filter, even though the SNR is very favorable (30 dB). In particular, we

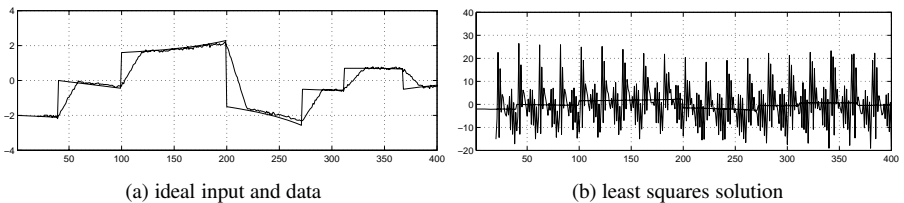


Figure 4.3. Example of 1D deconvolution (MA filter of (4.5) with a SNR of 30 dB): the solution (b) is unacceptable

find again the influence of the first zero transmission in the neighborhood of which the eigenvalues of $\mathbf{H}^T \mathbf{H}$ are very small (see Figure 4.2), which engenders an oscillation of period T , the duration of the integration window of the filter.

4.4. Batch deconvolution

Processing the data *as a batch* supposes having computing and memory capacity suited to the value of N . If these conditions are satisfied, the implementation essentially comes down to the inversion of a matrix, which is often large but structured. This is what we are going to see in this section. When the above conditions are not satisfied, we can turn to recursive solutions as presented in section 4.5.

4.4.1. Preliminary choices

The methodology explained in this book consists of choosing and minimizing a composite regularized criterion such as (2.5):

$$\mathcal{J}(\mathbf{x}) = \mathcal{G}(\mathbf{y} - \mathbf{H}\mathbf{x}) + \alpha \mathcal{F}(\mathbf{x}),$$

and this chapter is devoted to the quadratic terms (we denote $\|\mathbf{x}\|_{\mathbf{A}} = \mathbf{x}^T \mathbf{A} \mathbf{x}$):

$$\mathcal{G}(\mathbf{y} - \mathbf{H}\mathbf{x}) = \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_{\mathbf{R}_b}^2 \quad \text{and} \quad \mathcal{F}(\mathbf{x}) = \|\mathbf{x} - \mathbf{m}_x\|_{\mathbf{Q}}^2.$$

According to the Bayesian interpretation of Chapter 3, this criterion is equivalent to maximum *a posteriori* estimation with a Gaussian distribution for the noise: $\mathbf{b} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_b)$ and a Gaussian prior for the object: $\mathbf{x} \sim \mathcal{N}(\mathbf{m}_x, \mathbf{R}_x)$, with $\mathbf{R}_x \propto \mathbf{Q}^{-1}$, assuming of course that \mathbf{Q} is defined positive.

A dominant characteristic of the deconvolution methods presented in this chapter is that they are based on two strong properties: 1) the *invariance under translation* of the observation system, characteristic of a convolution, which makes \mathbf{H} a Toeplitz matrix, and 2) the *stationarity* of the phenomena being considered, which leads to Toeplitz covariance matrices \mathbf{R}_b and \mathbf{R}_x .

With generalization in mind, we could choose a non-zero mean \mathbf{m}_b in the distribution for the noise so as to take account of a *systematic error* during data acquisition. If this additional degree of freedom is to be really useful, it is necessary for this error to be accessible, otherwise it would be impossible, in this “linear-quadratic” framework, to separate \mathbf{m}_b and $\mathbf{H}\mathbf{x}$. If this error is accessible, we just have to subtract it from the data before performing the inversion. In the case of a stationary white noise, i.e., $\mathbf{R}_b = \sigma_b^2 \mathbf{I}$, the regularization parameter appears equivalent to an inverse SNR $\alpha = \sigma_b^2 / \sigma_x^2$.

The *a priori* mean for the object \mathbf{m}_x , or *default solution*, is not obliged to be stationary, i.e., constant, and the use of an appropriately chosen, non-constant mean has been suggested in order to “*Gaussianize*” the histogram of the experimental values \mathbf{y} and thus come closer to the normality hypothesis implicit in these methods [HUN 76]. However, not too much should be expected of this technique. The quality of the solution is fundamentally limited by its linearity with respect to the data and the use of a non-zero mean \mathbf{m}_x does not greatly change the performance of the estimator.

As far as matrix \mathbf{Q} is concerned, a simple and usual choice for the quadratic regularization term is

$$\alpha \mathcal{F}(\mathbf{x}) = \alpha_0 \sum x_n^2 + \alpha_1 \sum (x_{n+1} - x_n)^2. \quad (4.17)$$

This allows penalties to be applied to the norm of the solution and to that of the first differences (penalization of non-smooth solutions). The two terms of (4.17) are often called the zeroth-order and first-order penalties. Note that the choice $\alpha_1 > 0$ and $\alpha_0 = 0$ leads to a singular matrix \mathbf{Q} (and an improper *a priori* distribution, see Chapter 7, section 7.3.1.3).

This type of regularization is generalized by introducing a discretized derivation operator \mathbf{D} (differences of order p in 1D, Laplacian in 2D) and defining:

$$\mathbf{Q} = \mathbf{D}^T \mathbf{D}. \quad (4.18)$$

4.4.2. Matrix form of the estimate

With the above choices, the criterion writes

$$\mathcal{J}(\mathbf{x}) = \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_{\mathbf{R}_b^{-1}}^2 + \alpha \|\mathbf{D}\mathbf{x}\|^2 \quad (4.19)$$

and the solution is

$$\hat{\mathbf{x}} = (\mathbf{H}^T \mathbf{R}_b^{-1} \mathbf{H} + \alpha \mathbf{D}^T \mathbf{D})^{-1} \mathbf{H}^T \mathbf{R}_b^{-1} \mathbf{y}. \quad (4.20)$$

An example of such a solution is shown in the right-hand part of Figure 4.4 in the example of the MA filter of equation (4.5), and using a first-order regularization ($\alpha_0 = 0$ in (4.17)). It suffers from the defects belonging to the quadratic framework. Relation (2.8) shows that it is impossible to correctly restore all the components of the solution (which are the coefficients of its Fourier series expansion under the periodic hypothesis). Those in the neighborhood of the eigenvalues $\lambda_h^2(k)$ that are too small or zero are forced to zero and the discontinuities of the input signal, which require a very large number of Fourier coefficients to be correctly restored, are thus filtered. This results in parasite oscillations in their neighborhood, known as *ringing* (or the Gibbs phenomenon).

Calculating solution (4.20) has *a priori* a cost proportional to $O(M^3)$ elementary arithmetic operations (scalar multiplications and additions). With specific boundary hypotheses, such as the *periodic hypothesis*, the structure of the matrix to be inverted makes it possible to use algorithms with a cost proportional to $O(M \log M)$, but the *boundary effects* can be large. For instance in the example in Figure 4.4, the periodic hypothesis enables a rapid calculation (see section 4.4.3) but estimates a signal \hat{x} that has a very marked boundary effect expressed by large-amplitude oscillations (see left-hand part of the figure). It is worth noting however that the central part of the solution with periodic hypothesis is practically the same as using an exact calculation.

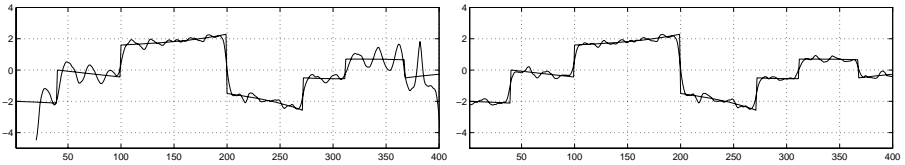


Figure 4.4. 1D deconvolution: “periodic” solution (left) and “exact” solution (right). The actual input and the data are shown in Figure 4.3a. The choice of $\alpha_1 = 0.05$ is obtained by minimizing the quadratic error between the exact solution and the true input, which is calculable in this simulated example

To avoid boundary effects, border signals x_1 and x_r must be estimated as well as the central part. It is thus necessary to invert the matrices that are *only* Toeplitz or close to Toeplitz. The inversion of such matrices has been the subject of many works. The algorithms given by Levinson [LEV 47] enable inversion to be carried out with a computing cost proportional to $O(N^2)$ [GOL 96]. In the 1980s, algorithms in $O(N(\log N)^2)$ were proposed [MOR 80]. We will not go into the detail of these works as we prefer to present fast solutions based on the DFT [HUN 73] or the conjugate gradient (CG) algorithm [COM 84, CHA 88, CHA 93, CHA 96, NAG 96], which keep us in an algebraic framework close to that used in the other chapters of this book.

4.4.3. Hunt’s method (periodic boundary hypothesis)

Hunt’s method [HUN 73] is based on the periodic approximation introduced in section 4.3.3 and on a circulant approximation \mathbf{D}_p of the finite difference operator \mathbf{D} . Both circulant operators are diagonalized in the Fourier basis

$$\mathbf{H}_p = \mathbf{W}^* \Lambda_h \mathbf{W}, \quad \text{and} \quad \mathbf{D}_p = \mathbf{W}^* \Lambda_d \mathbf{W}.$$

Diagonal components of Λ_h are obtained by DFT of the first column \mathbf{h}_p of \mathbf{H}_p : $\widehat{\mathbf{h}} = \mathbf{W}\mathbf{h}_p$. We have seen in section 4.3.3 that \mathbf{h}_p is a N -periodic zero-padded RI and it is

easy to show that its DFT forms a regular sampling of \widehat{H} :

$$[\mathbf{\Lambda}_h]_{\ell,\ell} = \widehat{h}_\ell = \widehat{H}((\ell - 1)/N), \quad 1 \leq \ell \leq N.$$

The diagonal of $\mathbf{\Lambda}_d$ is the vector $\widehat{\mathbf{d}} = \mathbf{W}\mathbf{d}_p$, where \mathbf{d}_p denotes the first column of the circulant derivation matrix \mathbf{D}_p and forms a regular sampling of the corresponding transfer function \widehat{D} . For instance, in the case of a first-order penalty (see section 4.4.1), the derivation kernel is $\mathbf{d} = [-1, 1]^T$, $\mathbf{d}_p = [-1, 1, 0, \dots, 0]^T$, $\widehat{D}(\nu) = \exp(-2j\pi\nu) - 1$ and $\widehat{d}_\ell = \widehat{D}((\ell - 1)/N)$.

In the sequel we focus on the case of a stationary white observation noise $\mathbf{R}_b = \mathbf{I}$. Using diagonalized operators in (4.19-4.20), we obtain a separable criterion in the Fourier domain, in terms of the DFT $\widehat{\mathbf{x}} = \mathbf{W}\mathbf{x}$ and $\widehat{\mathbf{y}} = \mathbf{W}\mathbf{y}$:

$$\mathcal{J}(\mathbf{x}) = \sum_{\ell=1}^N \left\{ |\widehat{y}_\ell - \widehat{h}_\ell \widehat{x}_\ell|^2 + \alpha |\widehat{d}_\ell \widehat{x}_\ell|^2 \right\}, \quad (4.21)$$

and the solution is

$$\widehat{x}_\ell = \frac{\widehat{h}_\ell^* \widehat{y}_\ell}{|\widehat{h}_\ell|^2 + \alpha |\widehat{d}_\ell|^2}, \quad 1 \leq \ell \leq N. \quad (4.22)$$

Hunt's algorithm can then be summarized by the following operations:

- 1) calculate the DFT of the data $\widehat{\mathbf{y}} = \mathbf{W}\mathbf{y}$;
- 2) calculate the DFT of the N -periodic zero-padded RI $\widehat{\mathbf{h}} = \mathbf{W}\mathbf{h}_p$;
- 3) calculate the DFT of the N -periodic zero-padded derivation kernel $\widehat{\mathbf{d}} = \mathbf{W}\mathbf{d}_p$;
- 4) calculate the solution in the Fourier domain using (4.22) and the time-domain solution $\widehat{\mathbf{x}}$ by inverse DFT.

As each Fourier transform is performed using the FFT, the resulting algorithm is very fast. Note that 2D periodic deconvolution with stationary covariances, which benefits from the circulant-block-circulant structure of the matrices, leads to exactly the same operations but using 2D DFT, the 2D RI of the observation system and a 2D derivation mask (for instance a Laplacian mask).

Solution (4.22) can be considered as a time-limited implementation of the Wiener deconvolution filter (4.6) in the case of white noise ($S_b(\nu) = \text{const.}$) the signal spectrum S_x being replaced by $1/|\widehat{\mathbf{d}}|^2$. This is not only an analogy, but matches the Bayesian interpretation of matrix $\mathbf{Q} = \mathbf{D}^T \mathbf{D}$ as the inverse of the correlation matrix of the signal to be restored (except for a multiplicative coefficient, which is accounted for in the regularization parameter α). For a first order regularization, we get $S_x(\nu) \propto 1/(1 - \cos 2\pi\nu)$ and, by expanding the cosine in $\nu < 1$, we can observe that

this regularization amounts to choosing a signal power spectra decreasing as $1/\nu^2$. A generalization is to choose a power decrease model for the signal spectrum such as

$$S_x(\nu_\ell) = \frac{1}{|\widehat{d}_\ell|^2} = \frac{1}{1 + (\nu_\ell/\nu_c)^p}, \quad \nu_\ell = \frac{\ell-1}{N}, \quad 1 \leq \ell \leq N, \quad (4.23)$$

with $2 \leq p \leq 4$, where ν_ℓ is the normalized discrete frequency and ν_c is a cut-off frequency, or the inverse of a mean correlation length.

Finally, it should be noted that the above calculations (diagonalization in a transformed space) are similar to those of [NG 99] for a mirror condition when the IR is even. The diagonalization is then carried out by DCT.

4.4.4. *Exact inversion methods in the stationary case*

In the stationary case, it may be necessary to make boundary hypotheses that do not lead to a fast solution. In fact, we often satisfy ourselves with a fast solution for one of the following reasons:

- 1) the object is confined to the center of the reconstructed support and has identical values at each of its boundaries;
- 2) the object is spread but we are interested in its values in the central zone only (so the boundaries will be eliminated after restoration);
- 3) the noise level is such that the regularization used “erases” the boundary effects;
- 4) the IR is symmetric and the use of the mirror condition is sufficient.

Thus, it is relatively rare to deal with the exact stationary problem, i.e., without an approximate boundary hypothesis³.

If the case arises, it is possible to use a conjugate gradient (CG) algorithm preferably with a preconditioner (PCG). Circulant preconditioner CG (CPCG) algorithms are particularly well suited to deconvolution problems.

We know that conjugate gradient algorithms enable quadratic problems of dimension N to be solved in N iterations. The cost of each iteration is dominated by that of multiplying a matrix by a vector, the dimensions of which are those of the data. In the case of Toeplitz matrices, this cost falls to $O(N \log N)$, thanks to the use of the FFT.

The use of preconditioners allows us to obtain algorithms that converge towards the solution linearly (and sometimes even a little faster; see [CHA 96]). This reduces

3. For instance, the 2D example which is dealt with in an exact stationary setting in the references [CHA 93, CHA 96, NAG 96] comes into category 1 and could be suitably treated by periodic approximation.

the cost of a solution, which is not exact but acceptable in the sense of a given stop criterion, to $O(N \log N)$ operations. In fact, in many cases, the CPCG algorithms can be stopped after less than about ten iterations, which explains the saving with respect to the standard CG. Using CPCG for deconvolution was proposed in 1984 by Commenges [COM 84]. Rediscovered somewhat later [CHA 88], deconvolution by CPCG has become a reference method today, as it is as fast as the Toeplitz matrix inversion algorithms and easier to use [CHA 96, NAG 96, NG 99].

To simplify the notation, in this section we will consider that the noise is white and use the regularization based on a discrete operator \mathbf{D} as in (4.18) so that the criterion to be minimized can be written:

$$\begin{aligned}\mathcal{J}(\mathbf{x}) &= \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 + \alpha \|\mathbf{D}\mathbf{x}\|^2 \\ &= \|\mathbf{v} - \mathbf{S}\mathbf{x}\|^2, \\ \mathbf{v} &= \begin{bmatrix} \mathbf{y} \\ \mathbf{0} \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} \mathbf{H} \\ \sqrt{\alpha}\mathbf{D} \end{bmatrix}.\end{aligned}$$

The preconditioning consists of introducing a new vector $\mathbf{u} = \mathbf{\Pi}\mathbf{x}$, where $\mathbf{\Pi}$ is a matrix (the preconditioner) that is close to \mathbf{S} while being quick to compose and to invert. We then use the CG algorithm to minimize:

$$\mathcal{J}(\mathbf{u}) = \|\mathbf{v} - \mathbf{S}\mathbf{\Pi}^{-1}\mathbf{u}\|^2,$$

which leads to a much smaller number of iterations than for the minimization of \mathbf{x} , but implies an added cost per iteration corresponding to the products of $\mathbf{\Pi}^{-1}$ and $(\mathbf{\Pi}^{-1})^T$ by the vectors. By using circulant preconditioners [CHA 88], we can carry out these products by FFT, which allows the cost of an iteration to be kept to $O(N \log N)$.

For example, taking our inspiration from deconvolution in the periodic hypothesis, we can use a preconditioner deduced from the circulant matrix:

$$\mathbf{C} = \mathbf{W}^* \text{diag} \left\{ \sqrt{|\widehat{h}_\ell|^2 + \alpha |\widehat{d}_\ell|^2} \right\} \mathbf{W}, \quad \ell = 1, 2, \dots, M,$$

where \widehat{h}_ℓ and \widehat{d}_ℓ are the coordinates of the respective DFT of \mathbf{h} and \mathbf{d} , the first columns of matrices \mathbf{H} and \mathbf{D} [COM 84, NAG 96]. This method can be extended to the 2D case by using a circulant-block-circulant matrix constructed from the 2D transfer function of the instrument and the DFT of the derivation kernel; see [NAG 96]. By initializing the CPCG algorithm to zero, the first iteration of the algorithm coincides with Hunt's solution that we saw in the previous section. So, after a first circulant inversion step, we may very well decide to continue with the CPCG algorithm of [NAG 96] if solution (4.22) has boundary effects that are too pronounced.

4.4.5. Case of non-stationary signals

In the non-stationary case, the Toeplitz or Toeplitz-block-Toeplitz characteristic of the matrices is lost and the computing cost of a matrix inversion is, *a priori*, proportional to $O(N^3)$. It is always possible to implement a PCG algorithm to calculate the solution, but choosing the preconditioner is trickier. On this subject, see the recent works by Fessler *et al.* [FES 99]. Also note the possibility of using Kalman smoothing, which will be presented in section 4.5.

4.4.6. Results and discussion on examples

4.4.6.1. Compromise between bias and variance in 1D deconvolution

The expressions for the bias and the covariance matrix of the linear operators were recalled in section 3.8 for the general case. In the specific case of deconvolution with the periodic, stationary hypothesis, all the matrices involved are square, circulant and can be diagonalized in the Fourier basis, and it is interesting to work in the frequency domain.

In the example of the MA filter of equation (4.5) regularized according to (4.17), relation (4.22) shows that the regularizer \mathbf{R}_α of equation (2.7), defined by $\hat{\mathbf{x}} = \mathbf{R}_\alpha \mathbf{y}$, has a frequency transfer function:

$$\widehat{g}_\ell = \frac{\widehat{h}_\ell^*}{|\widehat{h}_\ell|^2 + \alpha_0 + \alpha_1 |\widehat{d}_\ell|^2}, \quad \ell = 1, \dots, N.$$

The DFT of the bias is thus easy to calculate (\mathbf{x}° is the original object):

$$\mathbb{E}(\widehat{\mathbf{x}}) - \mathbf{x}^\circ \xrightarrow{\text{DFT}} \left\{ (\widehat{g}_\ell \widehat{h}_\ell - 1) \widehat{x}_\ell^\circ \right\}_{\ell=1}^N, \quad (4.24)$$

The squares of the moduli of each of the coordinates of this DFT – the sum of which gives the bias energy, by isometry – are plotted on the left-hand part of Figure 4.5, as a function of the reduced discrete frequency $\nu_\ell = (\ell - 1)/N$. We observe that, for frequencies where the SNR is high ($|\widehat{h}_\ell|^2 \gg \alpha_0 + \alpha_1 |\widehat{d}_\ell|^2$ in the example considered), the frequency components of the bias energy of the regularized solution have a negligible modulus. On the other hand, at frequencies that are very much degraded by the system (high frequencies and the neighborhood of the transmission zeros), these components take notable values, which are nevertheless lower than the values reached by the true solution (section 3.8.1).

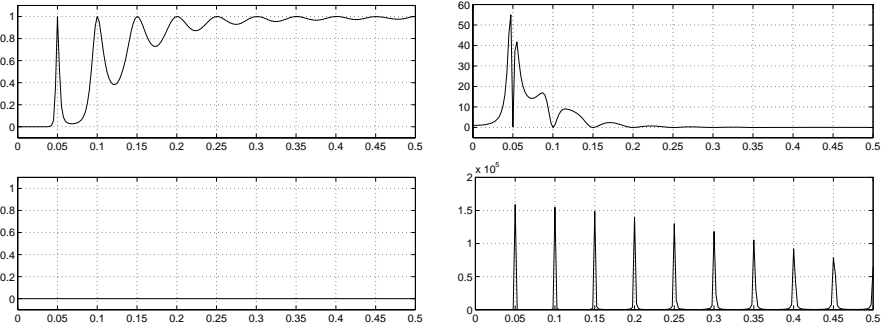


Figure 4.5. Components of the bias energy (normalized by \hat{x}_0), left, and components of the trace of the covariance matrix of the estimator (normalized by σ_b^2), right, versus the reduced discrete frequency for the example of the MA filter of section 4.2. Solution regularized with the prior (4.17) at top ($\alpha_0 = 0$, $\alpha_1 = 0.05$), least squares solution at bottom (note the scale of the vertical axes)

Similarly, the trace of the covariance matrix \mathbf{R}_α of the solution can be expressed as a sum of frequency components:

$$\text{trace Cov}(\hat{\mathbf{x}}) = \sigma_b^2 \sum_{\ell=1}^N |\hat{g}_\ell|^2,$$

which are plotted on the right-hand side of Figure 4.5, as a function of the reduced discrete frequency. This time, we observe that, for frequencies that are very much degraded by the system, these frequency components have an amplitude that is clearly reduced with respect to that obtained with a non-regularized solution.

Overall, the mean quadratic error, the sum of the bias energy and the trace of the covariance matrix of the solution, decreases as soon as $\alpha_1 > 0$, i.e., as soon as regularization is performed. This is what Figure 4.6 clearly shows. This error has a minimum for α_1 close to $\alpha_1 = 0.02$, which approximately corresponds to the value $\alpha_1 = 0.05$ determined by minimizing the simple quadratic error, defined as $\|\hat{\mathbf{x}} - \mathbf{x}^\circ\|^2$. We also observe a general characteristic of these quadratic regularization methods: the restoration error only varies significantly for variations of about an order of magnitude in the regularization parameter. There is thus little point in trying to fine tune it. In practice, the true solution \mathbf{x}° is obviously unknown and the optimum value of the regularization coefficient cannot be set by minimizing the mean quadratic error. It can be set directly by the user (*supervised mode*), or estimated from the only data available (*non-supervised mode*), by cross validation or maximization of a marginal likelihood, the expressions for which were given in sections 2.3.3 and 3.8.2. The results obtained in

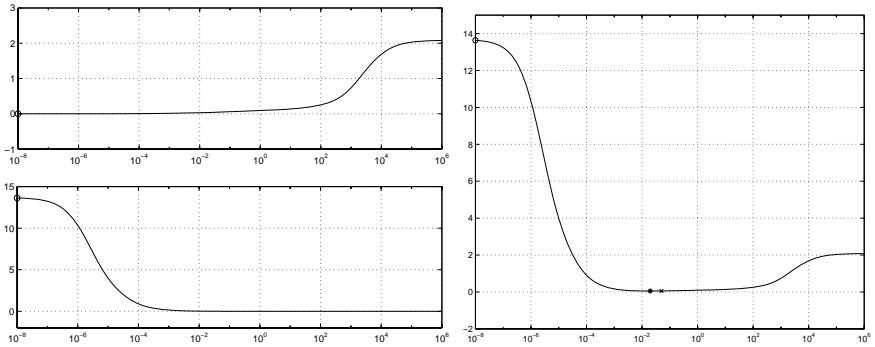


Figure 4.6. Bias energy (top left), trace of the covariance matrix of the solution (bottom left) and mean quadratic error (right) versus α_1 ($\alpha_0 = 0$), on a logarithmic scale. The minimum quadratic error is obtained for $\alpha_1 = 0.02$, and is marked by a star. The cross marks the value used previously ($\alpha_1 = 0.05$). The values corresponding to the least squares solution are marked by a circle ($\alpha_1 = 0$)

the example of the MA filter (not presented here) show that the value of the regularization coefficient obtained in this way is very close to the value that minimizes the mean quadratic error.

4.4.6.2. Results for 2D processing

In this section, we present a result of 2D deconvolution on real data provided by L.M. Mugnier and J-M. Conan (ONERA/DOTA/HRA). It concerns the observation of Ganymede (one of the moons of Jupiter) by the Haute-Provence observatory, France, using an adaptive optics system developed by ONERA. These data are presented in more detail in Chapter 10.

Ganymede, discovered by Galileo, is visible from the Earth with an ordinary telescope. Using a telescope with adaptive optics partly corrects the influence of atmospheric turbulence but the image obtained, shown on the left of Figure 4.7, still has marked defects due to residual errors of the wavefront correction. This degradation is only approximately modeled by a convolution, using a time average of the residues as the instrument response⁴. The response of the system is given in Figure 4.7. It shows a fairly fine peak on a broad circular plateau. The object is confined to the center of the field of observation, which enables the doubly periodic boundary hypothesis to be employed successfully, with Hunt's method.

4. Chapter 10 presents a treatment of this problem by *myopic deconvolution*, which is more suitable.



Figure 4.7. *Left, the observed image; center, the instrument response (IR) on a linear scale; and right, the logarithm of this response thresholded at 1/1,000th of its maximum. On the IR, a central peak can be observed, with a vast plateau spreading over most of the image support*

Figure 4.8 left shows that a truncated singular value decomposition (TSVD; see section 2.1.1) leads to a very noisy result in this example, despite the truncation. Using a quadratic regularization – which is equivalent to *damping* the singular values that are too small, see (2.8), rather than truncating them abruptly below a certain threshold – allows the noise amplification to be controlled better, as the central part of Figure 4.8 shows, in comparison with the reference image on the right which was reconstructed from images taken by probes exploring the solar system⁵. Several details visible in the restored image are confirmed by the reference image: a dark zone at upper left, and a light zone at lower center.



Figure 4.8. *Left, deconvolution using TSVD; center using Hunt's method with Laplacian regularization (4.23); right, the reference*

5. Data from the NASA/JPL base; see <http://space.jpl.nasa.gov/>.

Figure 4.9 presents a comparison between two *a priori* models of the 2D power spectral density for the object deduced from (4.23). The law for the decrease of the power spectral density as a function of the modulus of the spatial frequency is different. We note that the solutions are close and the general characteristics of linear solutions are again present: no spectral extrapolation, and ringing (Gibbs phenomenon) at intensity jumps on the edges of the object.

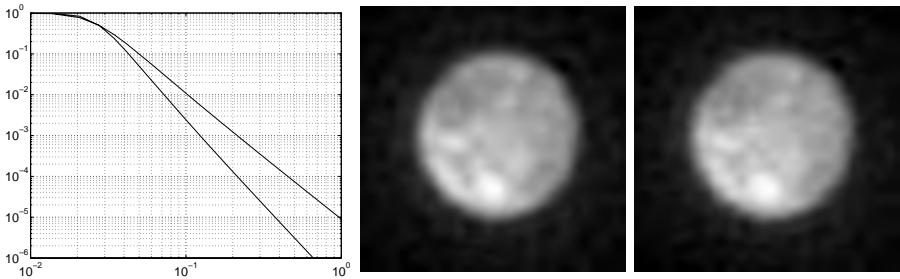


Figure 4.9. Comparison of results obtained with a prior spectrum of type (4.23) and two different laws for the power spectral density of the object (shown on left): one in $1/\nu^4$ (Laplacian regularization, result in center) and the other in $1/\nu^3$ (result shown on right)

4.5. Recursive deconvolution

It is not always possible, or even desirable, to process the data wholesale, either because the matrix to be inverted exceeds the available computer capacity, or because we want to do the processing on-line, or in real time as the data are acquired. If we give up the idea of processing the data recursively several times⁶, as in the last case, Kalman filtering naturally provides a suitable method. We will start by restricting ourselves to 1D signals but the 2D case will be rapidly reviewed in section 4.5.8.

4.5.1. Kalman filtering

For the applications that interest us here, the Kalman filter equations are based on the following state-space representation:

$$\begin{cases} \mathbf{x}_{n+1} = \mathbf{F}_n \mathbf{x}_n + \mathbf{G}_n \mathbf{u}_n, \\ y_n = \mathbf{H}_n \mathbf{x}_n + b_n, \end{cases} \quad n = 1, 2, \dots, \quad (4.25)$$

6. Otherwise, an item-by-item iterative technique can be used as in section 2.2.2.

in which the observation y_n is scalar; matched with the data for the moments of order 1 and 2:

$$\begin{aligned} \mathbb{E} \left(\begin{bmatrix} \mathbf{u}_n \\ b_n \end{bmatrix} \right) &= \mathbf{0}, \quad \mathbb{E}(\mathbf{x}_0) = \mathbf{m}_0^x, \\ \mathbb{E} \left(\begin{bmatrix} \mathbf{x}_0 - \mathbf{m}_0^x \\ \mathbf{u}_n \\ b_n \end{bmatrix} \begin{bmatrix} (\mathbf{x}_0 - \mathbf{m}_0^x)^T, \mathbf{u}_n^T, b_n \end{bmatrix} \right) &= \begin{bmatrix} \mathbf{R}_0^x & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_n^u & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & r_n^b \end{bmatrix}. \end{aligned}$$

In a Bayesian interpretation, as chosen here, the Kalman filter enables us to recursively compute the mean and the covariance matrix of the *a posteriori* probability distribution of the state vector \mathbf{x}_n , which is normal, knowing the data y_1, \dots, y_m . These are denoted $\hat{\mathbf{x}}_{n|m}$ and $\mathbf{R}_{n|m}^x$ respectively. Strictly speaking, Kalman “filtering” corresponds to the case where $m = n$. When $m > n$, we tend to speak of “Kalman smoothing” and we have two possibilities:

- the *fixed-lag* smoother which calculates $\hat{\mathbf{x}}_{n|n+p}$ for any n ;
- the *fixed interval* smoother, which calculates $\hat{\mathbf{x}}_{n|N}$ for any $n \leq N$ in a given time interval $[1, N]$; this solution corresponds rigorously to batch processing of the data.

Below, we recall the recurrence equations of the filter then the smoother. For a demonstration of these equations, the reader is invited to consult the references [AND 79, JAZ 70, VAN 68].

For the so-called “*covariance*” form chosen here, the filter recurrence repeatedly links the two following operations at each instant, in the sense of increasing time ($n = 1, 2, \dots, N$):

- *one-step prediction*:

$$\begin{aligned} \hat{\mathbf{x}}_{n|n-1} &= \mathbf{F}_{n-1} \hat{\mathbf{x}}_{n-1|n-1} \\ \mathbf{R}_{n|n-1}^x &= \mathbf{F}_{n-1} \mathbf{R}_{n-1|n-1}^x \mathbf{F}_{n-1}^T + \mathbf{G}_{n-1} \mathbf{R}_{n-1}^u \mathbf{G}_{n-1}^T \end{aligned}$$

- *correction*:

$$\begin{aligned} r_n^e &= \mathbf{H}_n \mathbf{R}_{n|n-1}^x \mathbf{H}_n^T + r_n^b \\ \mathbf{k}_n &= \mathbf{R}_{n|n-1}^x \mathbf{H}_n^T (r_n^e)^{-1} \\ \hat{\mathbf{x}}_{n|n} &= \hat{\mathbf{x}}_{n|n-1} + \mathbf{k}_n (y_n - \mathbf{H}_n \hat{\mathbf{x}}_{n|n-1}) \\ \mathbf{R}_{n|n}^x &= (\mathbf{I} - \mathbf{k}_n \mathbf{H}_n) \mathbf{R}_{n|n-1}^x \end{aligned} \tag{4.26}$$

When it is not necessary to calculate the two means and the two covariance matrices explicitly at each instant, these two sets of equations can be merged into one.

Kalman smoothing consists of calculating the estimates $\hat{\mathbf{x}}_{N-1|N}, \dots, \hat{\mathbf{x}}_{1|N}$, in a backward recursive procedure starting from the estimates $\hat{\mathbf{x}}_{n|n}$ obtained in the forward processing with the conventional Kalman filter, and initializing the procedure with $\hat{\mathbf{x}}_{N|N}$. The recurrence equations of the Kalman smoother are the following:

– *update of the mean:*

$$\begin{aligned}\hat{\mathbf{x}}_{n|N} &= \hat{\mathbf{x}}_{n|n} + \mathbf{S}_n (\hat{\mathbf{x}}_{n+1|N} - \mathbf{F}_n \hat{\mathbf{x}}_{n|n}) \\ \mathbf{S}_n &= \mathbf{R}_{n|n}^x \mathbf{F}_n^T (\mathbf{R}_{n+1|n}^x)^{-1}\end{aligned}$$

– *update of the covariance:*

$$\mathbf{R}_{n|N}^x = \mathbf{R}_{n|n}^x + \mathbf{S}_n (\mathbf{R}_{n+1|N}^x - \mathbf{R}_{n+1|n}^x) \mathbf{S}_n^T$$

A few observations can be made:

– we are indeed concerned with smoothing the estimators coming from the forward procedure, which is performed according to the state evolution model. Neither the data nor the observation model come into these equations; the smoother only requires the filtered estimates $\hat{\mathbf{x}}_{n|n}$ and the covariances $\mathbf{R}_{n|n}^x$ and $\mathbf{R}_{n+1|n}^x$;

– calculating the smoothing gain \mathbf{S}_n means inverting matrix $\mathbf{R}_{n+1|n}^x$ which is of the order of the state dimension. The additional cost of the smoothing is thus largely dependent on the *a priori* model chosen. For an autoregressive (AR) model of order 1, such as the random walk model (4.29) used below, we can use scalar equations;

– as in the filtering case, there is nothing to stop us using the smoother to treat non-stationary state models.

For an application of Kalman filtering to deconvolution, it is clear that the observation equation of state model (4.25) must be linked, at least partially, with the initial discrete convolution (4.11). However, there are several ways of working which, combined with different choices for \mathbf{x}_n , \mathbf{F}_n and \mathbf{G}_n , lead to a variety of solutions.

4.5.2. Degenerate state model and recursive least squares

A careful look at equation (4.12) shows that a first possible choice for the state evolution equation is:

$$\mathbf{x}_{n+1} = \mathbf{x}_n = \mathbf{x} = [x_{-P+1}, x_{-P+2}, \dots, x_{N+Q}]^T, \quad (4.27)$$

which entails:

$$\mathbf{F}_n = \mathbf{I}, \quad \mathbf{G}_n = \mathbf{0}, \quad \mathbf{H}_n = [0, \dots, 0, h_P, \dots, h_0, \dots, h_{-Q}, 0, \dots, 0].$$

This model is thus *degenerate*, as the state vector is static and the Kalman filter is initialized simply by choosing $\hat{\mathbf{x}}_{0|0} = \mathbf{m}_0^x$ and $\mathbf{R}_{0|0}^x = \mathbf{R}_0^x \propto \mathbf{Q}^{-1}$. It is easy to

verify that there is no reason to distinguish between one-step prediction and filtering ($\hat{\mathbf{x}}_{n+1|n} = \hat{\mathbf{x}}_{n|n}$ and $\mathbf{R}_{n+1|n}^x = \mathbf{R}_{n|n}^x$) and that the smoothing stage is not useful because the last recursion in the forward time direction gives exactly the result of batch processing: $\hat{\mathbf{x}}_{N|N}$ and $\mathbf{R}_{N|N}^x$. The recursive equations obtained:

$$\begin{aligned} \mathbf{k}_n &= \mathbf{R}_{n-1|n-1}^x \mathbf{H}_n^T (\mathbf{H}_n \mathbf{R}_{n-1|n-1}^x \mathbf{H}_n^T + r_n^b)^{-1} \\ \hat{\mathbf{x}}_{n|n} &= \hat{\mathbf{x}}_{n-1|n-1} + \mathbf{k}_n (y_n - \mathbf{H}_n \hat{\mathbf{x}}_{n-1|n-1}) \\ \mathbf{R}_{n|n}^x &= (\mathbf{I} - \mathbf{k}_n \mathbf{H}_n) \mathbf{R}_{n-1|n-1}^x \end{aligned} \quad (4.28)$$

are in fact those of a *recursive least squares* algorithm for solving the regularized least squares problem (4.26) one data-item at a time. The main characteristics of this algorithm are the following:

- no boundary hypothesis is introduced, unlike in the fast batch inversion algorithms described earlier;
- the innovation and its variance are scalar, so there is no matrix inversion;
- a smoothed estimate is obtained *on-line*, without the need for a backward pass, as mentioned above;
- the exact computing cost depends on the composition of the *a priori* covariance matrix \mathbf{R}_0^x , but is in $O(N^2)$ per recursion, i.e., $O(N^3)$ in total, as for inversion of a matrix of dimension N .

This algorithmic solution is thus of little practical interest. To reduce the computing and memory load at each recursion, there are two possible paths: 1) reduce the dimension of the state vector by other choices of matrices \mathbf{F}_n , \mathbf{G}_n and \mathbf{H}_n , as in sections 4.5.3 and 4.5.6 below, or 2) take advantage of the *shift invariance* property of matrix \mathbf{H}_n of the degenerate state model above so as to avoid having to solve a Riccati equation to calculate gain vector \mathbf{k}_n at each recursion, as in section 4.5.4 below.

4.5.3. Autoregressive state model

The signal to be deconvolved is often modeled in the form of an AR signal of order L :

$$x_n = \sum_{\ell=1}^L a_\ell x_{n-\ell} + u_n. \quad (4.29)$$

In this section, we will restrict ourselves to causal systems ($Q = 0$); extending to general FIR systems is easy. By defining the state vector as $\mathbf{x}_n = [x_n, \dots, x_{n-K+1}]^T$, of dimension $K \geq \max(P+1, L)$ and defining the vector of the regression coefficients as $\mathbf{a} = [a_1, \dots, a_L]^T$, we can write the state evolution equation in form (4.25) on

condition that we choose matrices \mathbf{F}_n and \mathbf{G}_n in the following way:

$$\mathbf{F}_n = \mathbf{F} = \begin{bmatrix} [\mathbf{a}^T, \mathbf{0}_{1 \times (K-L)}] & 0 \\ \mathbf{I}_{K \times K} & \mathbf{0}_{K \times 1} \end{bmatrix}, \quad (4.30)$$

$$\text{and} \quad \mathbf{G}_n = \mathbf{G} = [1, 0, \dots, 0]^T. \quad (4.31)$$

The state evolution noise \mathbf{u}_n is thus the non-correlated generating process u_n which is scalar. To complete the modeling, it remains to define matrix \mathbf{H} in the form $\mathbf{H} = \mathbf{h}^T = [h_0, \dots, h_P, \mathbf{0}_{1 \times (K-P-1)}]$, a line-vector of dimensions $1 \times K$. The corresponding state equation (4.25) is called the *snake chain* model, or model in *companion form* according to the field of application (communications, automatic systems, etc.)⁷.

Kalman filter (4.26) applied to this model gives the vector: $\hat{\mathbf{x}}_{n|n} = [\hat{x}(n|n), \hat{x}(n-1|n), \dots, \hat{x}(n-K+1|n)]^T$ at each recursion. Its coordinates are all estimators of the signal to be restored x_n , obtained by fixed-lag smoothing for delays up to $K-1$. If this smoothing is sufficient to ensure an acceptable mean quadratic error, a smoothed solution $\{\hat{x}(n-p|n)\}_{n=1,2,\dots}$ ($0 \leq p \leq K-1$) is obtained in a single pass by extracting the appropriate coordinates from the succession of filtered vectors $\hat{\mathbf{x}}_{n|n}$. If not, it is necessary to use a fixed-interval smoother, which requires an additional, backward pass. We will see more about this later, in an example.

4.5.3.1. Initialization

If the second-order prior information available on the signal x_n is composed of the correlation function r_n^x , a factorization algorithm such as Levinson's gives vector \mathbf{a} and the variance of the generating process r^u . It is also possible to choose the regression coefficients directly. Among the choices made by various authors, it is interesting to note:

- the white noise model ($a_\ell = 0$, $1 \leq \ell \leq L$), used, for example, in geophysics [CRU 74];

- the random walk model ($a_1 = 1$ and $a_\ell = 0$, $2 \leq \ell \leq L$), which is non-stationary but does not lead to divergence of the filter if the mean of the signal is observed. It is then a convenient choice from the regularization point of view [COM 84].

As far as implementation itself is concerned, it consists of choosing vector \mathbf{m}_0^x and initial covariance matrix \mathbf{R}_0^x . The latter can be immediately calculated from r_n^x and

7. This model is different from that used in *predictive deconvolution* [ROB 54], where it is the observed signal y_n that is taken to be autoregressive. The transfer function of the instrument is thus “all-pole” and no longer FIR. A *myopic* inversion is then performed, but the signal to be restored (the process generating the AR model) is non-correlated and the phase of the estimated transfer function is minimum, unless there are specific constraints (see also section 9.4.1).

can also be deduced from \mathbf{a} and r^u by an inverse Levinson algorithm. For \mathbf{m}_0^x , however, the choice is more delicate: obviously, the true signal is not known for $n \leq 0$. We find again, here, the questions of boundary choices encountered in batch processing. Note that, for a stationary model, inexact initialization with a zero-mean white vector with a large variance is often enough.

4.5.3.2. Criterion minimized by Kalman smoother

In practice, we often content ourselves with carrying out fixed-lag smoothing in a single pass. However, the solution $\{\hat{x}(n-p|n)\}_{n=1,2,\dots}$ thus obtained is not the minimizer of the regularized criterion (4.19) and it is important to state the underlying criterion. For this, let us examine the estimations made at instants 0 and 1 for an *invariant* state model such as equation (4.25). Let $p(\mathbf{x}_0)$ be the *a priori* pdf for \mathbf{x}_0 , a Gaussian with mean \mathbf{m}_0^x and covariance \mathbf{R}_0^x . At instant 0, we measure y_0 , for which the direct law has the density:

$$p(y_0) \propto \exp \left\{ -\frac{1}{2r^b} (y_0 - \mathbf{h}^T \mathbf{x}_0)^2 \right\}.$$

Estimator $\hat{\mathbf{x}}_{0|0}$ maximizes the conditional law $p(\mathbf{x}_0 | y_0) \propto p(y_0)p(\mathbf{x}_0)$, and thus minimizes the quadratic criterion:

$$\mathcal{J}_0(\mathbf{x}_0) = \frac{1}{r^b} (y_0 - \mathbf{h}^T \mathbf{x}_0)^2 + \|\mathbf{x}_0 - \mathbf{m}_0^x\|_{(\mathbf{R}_0^x)^{-1}}^2.$$

At instant 1, the conditional law $p(\mathbf{x}_1 | y_0, y_1)$ is the marginal of the joint law $p(\mathbf{x}_0, \mathbf{x}_1 | y_0, y_1)$ which can be factorized as follows:

$$p(\mathbf{x}_0, \mathbf{x}_1 | y_0, y_1) = p(y_0 | \mathbf{x}_0) p(y_1 | \mathbf{x}_1) p(\mathbf{x}_1 | \mathbf{x}_0) p(\mathbf{x}_0),$$

and, in the same way, corresponds to the following quadratic criterion:

$$\mathcal{J}_1(\mathbf{x}_0, \mathbf{x}_1) = \frac{1}{r^b} \sum_{k=0}^1 (y_k - \mathbf{h}^T \mathbf{x}_k)^2 + \|\mathbf{x}_1 - \mathbf{F} \mathbf{x}_0\|_{(\mathbf{G} \mathbf{R}^u \mathbf{G}^T)^{-1}}^2 + \|\mathbf{x}_0 - \mathbf{m}_0^x\|_{(\mathbf{R}_0^x)^{-1}}^2$$

The minimum of this criterion is reached for the pair of vectors $\{\hat{\mathbf{x}}_0^{(1)}, \hat{\mathbf{x}}_1^{(1)}\}$, such that $\hat{\mathbf{x}}_0^{(1)} = \hat{\mathbf{x}}_{0|1}$ and $\hat{\mathbf{x}}_1^{(1)} = \hat{\mathbf{x}}_{1|1}$. Let us generalize this procedure: at instant n , we define the joint criterion:

$$\mathcal{J}_n(\mathbf{x}_0, \dots, \mathbf{x}_n) = \frac{1}{r^b} \sum_{k=0}^n (y_k - \mathbf{h}^T \mathbf{x}_k)^2 + \sum_{k=1}^n \|\mathbf{x}_k - \mathbf{F} \mathbf{x}_{k-1}\|_{(\mathbf{G} \mathbf{R}^u \mathbf{G}^T)^{-1}}^2 + \|\mathbf{x}_0 - \mathbf{m}_0^x\|_{(\mathbf{R}_0^x)^{-1}}^2$$

which has its minimum for the $(n+1)$ -uplet of vectors $\{\hat{\mathbf{x}}_0^{(n)}, \dots, \hat{\mathbf{x}}_n^{(n)}\}$. We note that the estimator $\hat{\mathbf{x}}_{n|n}$ produced by the Kalman filter at instant n is $\hat{\mathbf{x}}_n^{(n)}$. Moreover,

the $n + 1$ estimators of the fixed-interval Kalman smoother $\hat{\mathbf{x}}_{0|n}, \dots, \hat{\mathbf{x}}_{n|n}$ are the $n + 1$ vectors that minimize \mathcal{J}_n .

The above results do not apply directly to the *a priori* AR model (4.29) above, as matrix $\mathbf{G} \mathbf{R}^u \mathbf{G}^T$ is not invertible with the choice made for \mathbf{G} in (4.31). To reach criterion \mathcal{J}_n , we replace it in the previous criterion by a diagonal matrix having the diagonal elements $\{r^u, \varepsilon^2, \dots, \varepsilon^2\}$. By making ε tend towards 0, we find that two consecutive estimated vectors $\hat{\mathbf{x}}_{m-1|n}$ and $\hat{\mathbf{x}}_{m|n}$ must have $K - 1$ equal components, in agreement with the state evolution equation. We thus define the deconvolved vector $\hat{\mathbf{x}}_n^{\text{ks}}$ as the concatenation of the first components of each vector $\hat{\mathbf{x}}_{m|n}$, and show that this vector minimizes the following criterion:

$$\mathcal{J}^{\text{ks}}(\mathbf{x}_n) = \frac{1}{r^b} \|\mathbf{y} - \mathbf{H} \mathbf{x}\|^2 + \frac{1}{r^u} \sum_{m=1}^n \left(x_m - \sum_{\ell=0}^L a_\ell x_{m-\ell} \right)^2 + \frac{1}{r^u} x_0^2, \quad (4.32)$$

where \mathbf{H} is the observation block-matrix. The initialization term comes from the choice of the initial *a priori* moments.

The above criterion is clearly connected with the quadratic regularized criteria of the beginning of the chapter. In particular, choosing a random walk model ($a_1 = 1$ and $a_\ell = 0, \ell \geq 2$) corresponds to regularizing the first differences of the signal (equation (4.17) with $\alpha_0 = 0$). Other choices correspond to more general regularization terms, see for example Chapter 11.

4.5.3.3. Example of result

Figure 4.10 illustrates the results that can be obtained with these on-line deconvolution techniques for the example of an MA filter and with the random walk model. The fixed-interval smoothing solution, which minimizes \mathcal{J}^{ks} , is comparable to the results of the batch processing presented in Figure 4.4. The solutions by simple Kalman filtering, or by fixed-lag smoothing, give a greater estimation variance than the fixed-interval smoothing, thus inducing a higher relative quadratic error. The second plot on the figure shows the variation of this error for increasing lag and demonstrates clearly that, in this example, a large lag is necessary if it is to come close to the error of the fixed-interval smoother.

4.5.4. Fast Kalman filtering

The on-line deconvolution by the two variants of Kalman filtering described above does not reduce the computing cost. They are methods in $O(N^3)$. The reason for this is that they use the standard equations (4.26), which are too general because they apply to non-stationary problems whereas, in this chapter, the observation model is invariant and the signals are stationary.

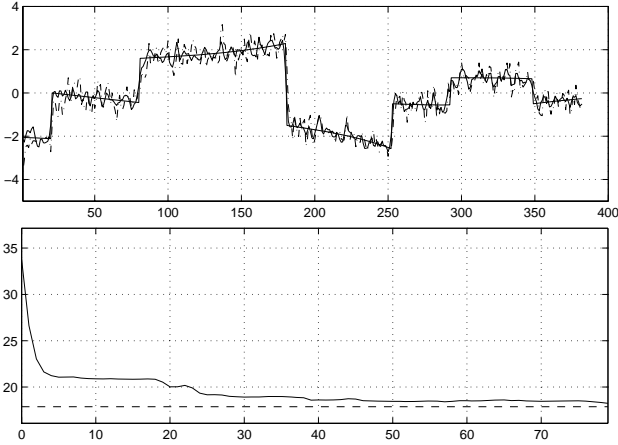


Figure 4.10. Results of on-line deconvolution of a 1D signal. Top, the actual input is represented by a dashed line, the result of Kalman filtering is shown as a dotted line, and the fixed-interval smoother result as a solid line. Bottom, variation of the relative quadratic error for increasing lags in the fixed-lag smoother is plotted as a solid line and the error of the fixed-interval smoother is shown as a dotted line. The SNR is 20 dB

When model (4.25) is stationary (i.e., when \mathbf{F} , \mathbf{G} , \mathbf{H} , \mathbf{R}^u and r^b do not depend on n), it is possible to markedly reduce the computing load by making the recursion act on *increments* of the nominal quantities of the filter $\mathbf{R}_{n|n-1}^x$ and \mathbf{k}_n rather than on the quantities themselves [DEM 89].

Thus, in the case of the AR model (4.30-4.31), we can show [COM 84] that this recursion can also be written:

$$\begin{bmatrix} (r_{n+1}^e)^{1/2} & \mathbf{0} \\ \mathbf{k}_{n+1}(r_{n+1}^e)^{-1/2} & \mathbf{v}_{n+1} \end{bmatrix} = \begin{bmatrix} (r_n^e)^{1/2} & s_n \\ \mathbf{k}_n(r_n^e)^{-1/2} & \mathbf{F}\mathbf{v}_n \end{bmatrix} \Theta_n, \quad (4.33)$$

where s_n and \mathbf{v}_n are auxiliary quantities defined at initialization and Θ_n is a J -orthogonal transformation matrix⁸. With a random walk model, this algorithm only requires $5(P + Q + 1)$ scalar multiplications per recursion and thus has a total cost in $O(N(P + Q + 1))$. However, the nature of transformation Θ_n shows that

8. A J -orthogonal transformation is such that $\Theta_n \mathbf{J} \Theta_n^T = \mathbf{J}$, where \mathbf{J} is a signature matrix, i.e., a diagonal matrix in which the diagonal elements have the values $+1$ or -1 . It does not preserve the Euclidian norm, unlike orthogonal transformations, and we lose a means of controlling the propagation of numerical errors.

this gain itself has a price: this type of algorithm has a potential numerical instability and it is not enough to use a *square root* form (4.33) as is usually recommended [AND 79, VER 86]. The choice of the corresponding hyperbolic rotation is very important [LEB 93].

In the case of a degenerate model (4.27), we are no longer dealing with a strictly invariant model but a shift-invariant one (since we move from \mathbf{H}_n to \mathbf{H}_{n+1} by shifting the coordinates). The technique described above is generalized by changing the definition of the increments in the nominal quantities of the filter [DEM 85]. The corresponding algorithm this time needs $n + P + Q + 1$ scalar multiplications per recursion, giving a total cost in $O(N^2)$, which is higher than the previous one. However, it should be remembered that, in a single pass, we obtain the fixed-interval smoother result without boundary hypotheses.

These fast filtering techniques and the asymptotic filtering results below are currently used essentially for on-line, real-time processing. The implementation problems raised come under IT engineering and are beyond the scope of this book. Interested readers are encouraged to refer to [MAS 99, MOZ 99].

4.5.5. Asymptotic techniques in the stationary case

4.5.5.1. Asymptotic Kalman filtering

In the stationary case, it is well known that the covariance matrix $\mathbf{R}_{n|n}^x$ that intervenes in the filter tends, when $n \rightarrow \infty$, towards a fixed matrix \mathbf{R}_∞ that verifies a Riccati equation of the form:

$$\mathbf{R}_\infty = \mathbf{F} \mathbf{R}_\infty \mathbf{F}^T - \mathbf{F} \mathbf{R}_\infty \mathbf{H} (\mathbf{H} \mathbf{R}_\infty \mathbf{H}^T + r^b)^{-1} + \mathbf{G} \mathbf{R}^u \mathbf{G}^T. \quad (4.34)$$

The Kalman gain \mathbf{k}_n then tends towards a constant vector and, as the calculation of these two quantities does not bring in the observed data y_n , the idea arose, as soon as Kalman filtering was invented, of calculating these asymptotic quantities in advance and carrying out *invariant* recursive filtering:

$$\hat{\mathbf{x}}_{n|n} = \mathbf{F} \hat{\mathbf{x}}_{n-1|n-1} + \mathbf{k} (y_n - \mathbf{H} \mathbf{F} \hat{\mathbf{x}}_{n-1|n-1}),$$

in order to reduce the computing cost of a recursion.

The asymptotic gain \mathbf{k} nevertheless remains to be calculated and the stability of the recursive filter obtained has to be verified. Abundant literature has been devoted to these questions [AND 79]. Equation (4.34) can be solved by a wide variety of methods, iterative or not, among which Chandrasekhar factorization methods, simple or dual, are particularly suited to invariant models in the broad sense and to varied initial conditions [DU 87].

4.5.5.2. *Small kernel Wiener filter*

Asymptotic Kalman filtering provides an optimum causal solution at low cost. However, in many problems, particularly in 2D, it appears more advantageous to use a local solution taking into account the past and future of the instant under consideration. A non-causal Wiener FIR estimator, or *small kernel* Wiener filter, can thus be chosen:

$$\hat{x}_n = \sum_{j=-J}^J g_j y_{n-j}$$

the optimality of which is defined locally by considering a sub-set of the Wiener-Hopf equations (3.15) in discrete time. This leads to a matrix system $\mathbf{R}_y \mathbf{g} = \mathbf{r}_{xy}$, where \mathbf{R}_y is the covariance matrix of the process y and \mathbf{r}_{xy} the vector of the input/output covariances for lags $\{-J, \dots, 0, \dots, J\}$. These quantities are easily expressed as functions of \mathbf{R}_x , \mathbf{R}_b and h .

This approach has been proposed for handling 3D problems [PER 97] and for restoring large sized images in satellite imaging [REI 95] – although we are not dealing with a simple deconvolution problem in the latter case.

4.5.6. *ARMA model and non-standard Kalman filtering*

When the input to be restored is *a priori* non-correlated, another means of reducing the computing load of a Kalman filter is to abandon the idea of modeling the direct problem by the discrete convolution used since the beginning of this chapter and to replace it with a *minimal realization*, i.e., a minimum order ARMA model, identified from the IR of the system [MEN 83]. The coefficients of the numerator of the transfer function obtained (MA part) serve to construct matrix \mathbf{H} of the state model and those of the denominator (AR part) matrix \mathbf{F} . The signal to be restored then becomes the noise, u_n , of the state evolution, which leads to a *non-standard Kalman* filtering or smoothing problem. This method, introduced for seismic reflection problems, has the advantage of being able to deal with non-stationary deconvolution problems for a reasonable computing cost.

4.5.7. *Case of non-stationary signals*

The principle of extending Kalman filtering to the non-stationary context (variable *a priori* model, variable response h , variable noise variance) poses no problem in as far as the time laws of all the parameters are sufficiently well known for the state model to be completely defined (4.25). In practice, however, non-stationarity is often associated with a poor knowledge of these laws and we are faced with a *myopic* problem, having a large number of parameters to be determined in addition to the samples of the input signal. This question will be tackled in Chapter 8.

4.5.8. On-line processing: 2D case

The use of Kalman filtering for 2D deconvolution was put forward in the 1970s. Once the method of scanning the image, usually in lexicographic order, has been chosen, the problem is formally a simple extension of the Kalman filtering seen above. Practically, it is difficult to reconcile this scanning mode with satisfactory probabilistic modeling of the image because of the absence of a 2D spectral factorization theorem. We are led to choose a very large state vector, which considerably increases the complexity of the algorithm. Various approximations have been used to reduce the calculations. Among them, we note the “RUM” (*Reduced Update Model*) [WOO 81] then “ROM” (*Reduced Order Model*) versions [ANG 89]. At the same time, fast versions of the Kalman filter were used for stationary problems; see, for example, [MAH 87, SAI 85, SAI 87]. Added to these difficulties is the handling of edge conditions, which can greatly complicate implementation.

It can be considered that this type of algorithm has practically fallen into disuse in image restoration. Most problems of reasonable size (up to 10^6 pixels) can be batch processed with the solutions presented in the first part of this chapter. For very large-sized problems (3D problems and 2D problems with 10^8 pixels), we can content ourselves with small kernel filtering based on stationary modeling of the problem as presented in section 4.5.5.

4.6. Conclusion

The deconvolution methods introduced in this chapter are fundamental for several reasons. First of all, many physical phenomena can be modeled, at least to a first approximation, by convolution. Secondly, this direct model can be interpreted easily in the frequency domain, which enables an extensive, intuitive analysis of the phenomena and difficulties encountered. Finally, the inversion methods proposed are based on *a priori* models that remain simple: quadratic regularization terms or Gaussian *a priori* laws, defined by their second-order properties. We thus arrive at estimators that are very simple, as they are linear, and we can thus turn our attention to questions of implementation: *block* methods (Hunt, preconditioned gradient) and *recursive* methods (Kalman filtering and smoothing).

In return, there are limitations in terms of resolution of the solutions obtained. The methods obtained perform partial *spectral equalization*, i.e., they only manage to compensate for certain attenuations in the sensor bandwidth; they do, however, avoid an explosive restitution of the frequencies that are too strongly attenuated. Thus, the resolution of the linear solutions is fundamentally limited by the spectral content of the data.

Any increase in resolution (apart from improvements to the measuring systems, which is outside the scope of this book) relies on taking into account more specific

information on the object to be reconstructed: positivity, pulse nature (see Chapter 5), or presence of contours in an image (see Chapter 6) for example. In other domains, such as digital communications, the fact that the input parameters belong to a finite alphabet can contribute to the same result.

4.7. Bibliography

- [AND 79] ANDERSON B. D. O., MOORE J. B., *Optimal Filtering*, Prentice-Hall, Englewood Cliffs, NJ, 1979.
- [ANG 89] ANGWIN D. L., KAUFMAN H., "Image restoration using reduced order models", *Signal Processing*, vol. 16, p. 21-28, 1989.
- [ARS 66] ARSAC J., *Fourier Transform and the Theory of Distributions*, Prentice-Hall, Englewood Cliffs, NJ, 1966.
- [BAY 70] BAYLESS J. W., BRIGHAM E. O., "Application of the Kalman filter to continuous signal restoration", *Geophysics*, vol. 35, num. 1, p. 2-23, 1970.
- [CHA 88] CHAN R. H., "An optimal circulant preconditionner for Toeplitz systems", *SIAM J. Sci. Comput.*, vol. 9, p. 766-771, 1988.
- [CHA 93] CHAN R. H., NAGY J. G., PLEMMONS R. J., "FFT-based preconditionners for Toeplitz-block least squares problems", *SIAM J. Num. Anal.*, vol. 30, num. 6, p. 1740-1768, Dec. 1993.
- [CHA 96] CHAN R. H., NG M. K., "Conjugate gradient methods for Toeplitz systems", *SIAM Rev.*, vol. 38, num. 3, p. 427-482, Sep. 1996.
- [COM 84] COMMENGES D., "The deconvolution problem: fast algorithms including the preconditioned conjugate-gradient to compute a MAP estimator", *IEEE Trans. Automat. Contr.*, vol. AC-29, p. 229-243, 1984.
- [CRU 74] CRUMP N. D., "A Kalman filter approach to the deconvolution of seismic signals", *Geophysics*, vol. 39, p. 1-13, 1974.
- [DEM 85] DEMOMENT G., REYNAUD R., "Fast minimum-variance deconvolution", *IEEE Trans. Acoust. Speech, Signal Processing*, vol. ASSP-33, p. 1324-1326, 1985.
- [DEM 89] DEMOMENT G., "Equations de Chandrasekhar et algorithmes rapides pour le traitement du signal et des images", *Traitement du Signal*, vol. 6, p. 103-115, 1989.
- [DU 87] DU X.-C., SAINT-FELIX D., DEMOMENT G., "Comparison between a factorization method and a partitioning method to derive invariant Kalman filters for fast image restoration", in DURRANI T. S., ABBIS J. B., HUDSON J. E., MADAN R. N., MCWHIRTER J. G., MOORE T. A. (Eds.), *Mathematics in Signal Processing*, p. 349-362, Clarendon Press, Oxford, UK, 1987.
- [FES 99] FESSLER J. A., BOOTH S. D., "Conjugate-gradient preconditionning methods for shift-variant PET image reconstruction", *IEEE Trans. Image Processing*, vol. 8, num. 5, p. 668-699, May 1999.
- [GOL 96] GOLUB G. H., VAN LOAN C. F., *Matrix Computations*, John Hopkins University Press, Baltimore, 3rd edition, 1996.

- [GRE 58] GRENANDER U., SZEGÖ G., *Toeplitz Forms and their Applications*, University of California Press, Berkeley, 1958.
- [HUN 71] HUNT B. R., “A matrix theory proof of the discrete convolution theorem”, *IEEE Trans. Automat. Contr.*, vol. AC-19, p. 285-288, 1971.
- [HUN 73] HUNT B. R., “The application of constrained least squares estimation to image restoration by digital computer”, *IEEE Trans. Computers*, vol. C-22, p. 805-812, 1973.
- [HUN 76] HUNT B. R., CANNON T. M., “Nonstationary assumptions for Gaussian models of images”, *IEEE Trans. Systems, Man, Cybern.*, p. 876-882, Dec. 1976.
- [JAI 89] JAIN A., *Fundamental of Digital Image Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1989.
- [JAZ 70] JAZWINSKI A. H., *Stochastic Process and Filtering Theory*, Academic Press, New York, NY, 1970.
- [KAL 60] KALMAN R. E., “A new approach to linear filtering and prediction problems”, *J. Basic Engng.*, vol. 82-D, p. 35-45, 1960.
- [KHU 77] KHURGIN Y. I., YAKOVLEV V. P., “Progress in the Soviet Union on the theory and applications of bandlimited functions”, *Proc. IEEE*, vol. 65, p. 1005-1029, 1977.
- [LEB 93] LE BESNERAIS G., GOUSSARD Y., “Improved square-root forms of fast linear least squares estimation algorithms”, *IEEE Trans. Signal Processing*, vol. 41, num. 3, p. 1415-1421, Mar. 1993.
- [LEV 47] LEVINSON N., “The Wiener RMS error criterion in filter design and prediction”, *J. Math. Physics*, vol. 25, p. 261-278, Jan. 1947.
- [MAH 87] MAHALANABIS A.-K., XUE K., “An efficient two-dimensionnal Chandrasekhar filter for restoration of images degraded by spatial blur and noise”, *IEEE Trans. Acoust. Speech, Signal Processing*, vol. 35, p. 1603-1610, 1987.
- [MAS 99] MASSICOTTE D., “A parallel VLSI architecture of Kalman-filter-based algorithms for signal reconstruction”, *Integration VLSI J.*, vol. 28, p. 185-196, 1999.
- [MEN 83] MENDEL J. M., *Optimal Seismic Deconvolution*, Academic Press, New York, NY, 1983.
- [MOR 80] MORF M., “Doubling algorithms for Toeplitz and related equations”, in *Proc. IEEE ICASSP*, Denver, CO, p. 954-959, 1980.
- [MOZ 99] MOZIPO A., MASSICOTTE D., QUINTON P., RISSET T., “A parallel architecture for adaptive channel equalization based on Kalman filter using MMA α ”, in *Proc. IEEE Canadian Conf. on Electrical and Computer Engng.*, Alberta, Canada, p. 554-559, May 1999.
- [NAG 96] NAGY J. G., PLEMMONS R. J., TORGENSEN T., “Iterative image restoration using approximate inverse preconditionning”, *IEEE Trans. Image Processing*, vol. 5, num. 7, p. 1151-1162, July 1996.
- [NG 99] NG M. K., CHAN R. H., TANG W.-C., “A fast algorithm for deblurring models with Neumann boundary conditions”, *SIAM J. Sci. Comput.*, vol. 21, num. 3, p. 851-866, 1999.

- [PER 97] PEREIRA S., GOUSSARD Y., "Unsupervised 3-D restoration of tomographic images by constrained Wiener filtering", in *Proc. IEEE EMB Conf.*, Chicago, IL, p. 557-560, 1997.
- [REI 95] REICHENBACH S. E., KOEHLER D. E., STRELOW D. W., "Restoration and reconstruction of AVHRR images", *IEEE Trans. Geosci. Remote Sensing*, vol. 33, num. 4, p. 997-1007, July 1995.
- [ROB 54] ROBINSON E. A., "Predictive decomposition of seismic traces", *Geophysics*, vol. 27, p. 767-778, 1954.
- [SAI 85] SAINT-FELIX D., HERMENT A., DU X.-C., "Fast deconvolution: application to acoustical imaging", in J.M. THIJSEN, V. MASSEO (Eds.), *Ultrasonic Tissue Characterization and Echographic Imaging*, Nijmegen, The Netherlands, Faculty of Medicine Printing Office, p. 161-172, 1985.
- [SAI 87] SAINT-FELIX D., DU X.-C., DEMOMENT G., "Filtres de Kalman 2D rapides à modèle d'état non causal pour la restauration d'image", *Traitement du Signal*, vol. 4, p. 399-410, 1987.
- [VAN 68] VAN TREES H. L., *Detection, Estimation and Modulation Theory*, Part 1, John Wiley, New York, NY, 1968.
- [VER 86] VERHAEGEN M., VAN DOOREN P., "Numerical aspects of different Kalman filter implementations", *IEEE Trans. Automat. Contr.*, vol. AC-31, num. 10, p. 907-917, Oct. 1986.
- [WIE 71] WIENER N., *Cybernétique et société*, Union générale d'édition, Paris, 1971.
- [WOO 81] WOODS J. W., INGLE V. K., "Kalman filtering in two dimensions: further results", *IEEE Trans. Acoust. Speech, Signal Processing*, vol. 29, num. 2, p. 568-577, Apr. 1981.

Chapter 5

Deconvolution of Spike Trains

5.1. Introduction

A point source is the idealization of a physical source that has a duration or dimension markedly smaller than the resolution of the sensor observing it. In practice, this type of source is found in signals or images observed in the form of an echo or spot whose shape is essentially a characteristic of the instrument: *impulse response* (IR). In these conditions, the only features of the observed signal that can be attributed to the source are its location and amplitude.

This type of source is encountered in astronomy, in stellar imaging, when the apparent diameter of the stars imaged is smaller than the parameter λ/D , where λ is the wavelength and D the diameter of the receiving antenna (see Chapter 10, section 10.1.2). It is also found in ultrasound echography or seismology whenever the scale of an inhomogeneity or transition is small with respect to the wavelength used (see Chapter 9).

The detection and location of a single source can be satisfactorily handled by “matched filter” techniques [VAN 68]. By extension, the matched filter will work when the “echoes” due to the sources are well separated from one another. As soon as the echoes due to several sources overlap significantly, matched filtering is no longer effective.

In such a context, spike train deconvolution aims to process an observed signal made up of overlapping echoes in order to extract the position and amplitude of the

point sources that generated the echoes. It corresponds to the observation model $y(t) = \sum_{k=1}^K r_k h(t - t_k) + b(t)$, in which K is the number of sources, r_k the amplitude of the k th source (in echography, this quantity represents the *reflectivity* of the medium, whence the traditional use of the variable r to represent this quantity), t_k is the time of arrival of the signal due to the k th source, h is the IR of the instrument, which is assumed to be known in the framework of this chapter, and b is the noise, which includes everything we do not intend to model deterministically. The noise is almost always assumed to be Gaussian and, in most cases, white, zero-mean and stationary. The developments that follow will use these hypotheses unless stated otherwise. A more realistic formulation of the problem takes the sampled and finite nature of the observations into consideration:

$$y(nT) = \sum_{k=1}^K r_k h(nT - t_k) + b(nT), \quad n = 0, \dots, N-1, \quad (5.1)$$

where T is the signal sampling period. In this form, the problem can be tackled from the standpoint of identifying a parametric model, the order, K , of which is unknown and must be estimated [WAL 97].

A noteworthy contribution adopting this approach has been made by [KWA 80], taking his inspiration from the CLEAN technique [HOG 74], which can be qualified as an “iterative matched filter”. In this respect, the impulse positioning algorithm called *multipulse*, which appeared in speech coding [ATA 82], works on a similar principle (but in a noise-free context) where h is estimated by linear prediction (see Chapter 9, section 9.4.1.3).

The same type of model has been much studied in the context of spectral ray analysis, which corresponds to a model structure close to equation (5.1) [DJU 96, DUB 97, STO 89, WON 92]. These approaches lead to a very tricky, non-convex optimization problem, one of the difficulties of which is the unknown size, K , of the space of parameters to be identified.

The approaches developed below get around this difficulty by considerably simplifying model (5.1). In practice this means substantially reducing the algorithmic complexity. The time of arrival space is discretized with a sampling step T and it is thus possible to rewrite (5.1) in the form:

$$y(nT) = \sum_m r_m h((n-m)T) + b(nT), \quad n = 0, \dots, N-1, \quad (5.2)$$

where we assume $r_m = 0$ if $mT \neq t_k, \forall k$. With the notation $y_n = y(nT)$ and $h_n = h(nT)$, (5.2) gives the invariant linear model $y = h \star r + b$ (4.11) of Chapter 4. With this model, the estimation of K , of the times of arrival and of the amplitudes is transformed into a dual problem. At each instant m we need to estimate the amplitude of a pulse that may (or may not) be located at m . Thus, estimating the number

of pulses and the positions of the times of arrival merges with the “detection” of the presence or absence ($r_m = 0$) of a pulse at each instant m . In the signal processing community, this problem is referred to as a *detection-estimation* problem. It is the detection operation that makes the corresponding optimization problem globally nonlinear.

In what follows, we will only consider a matrix version of (5.2):

$$\mathbf{y} = \mathbf{H}\mathbf{r} + \mathbf{b} \quad (5.3)$$

where \mathbf{y} , \mathbf{b} and \mathbf{r} are vectors containing, respectively, N samples of y and of b , and the corresponding M samples of r . M depends on N , on the number of samples used to represent the IR, and on the boundary assumption used (see section 4.3.3). The time-shift invariance implicit in equation (5.2) is expressed in matrix model (5.3) by the Toeplitz structure of \mathbf{H} , the $N \times M$ convolution matrix constructed from the samples of h .

The aim of spike deconvolution is to estimate \mathbf{r} . We will look at two families of approaches:

- 1) those that place the emphasis on the estimation aspect, by deferring the decisional aspect. In this context, the deconvolution operation is considered as a contrast enhancement technique. The natural framework for presenting these techniques is deterministic regularization;
- 2) those that consider the detection-estimation problem right from the start and seek to solve detection and estimation together. The traditional, and probably most appropriate, framework for setting out these techniques is the Bayesian one, with Bernoulli-Gaussian (BG) priors.

Whatever the family and the interpretation framework, these approaches all result in an optimization problem. The first family leads to techniques of nonlinear optimization on \mathbb{R}^M , while the second brings in combinatorial optimization on the discrete space $\{0, 1\}^M$.

5.2. Penalization of reflectivities, L2LP/L2Hy deconvolutions

The synthetic example of Figure 5.1 illustrates the failure of the trivial matched filter and least squares solutions on equation (5.3). The spike train signal to be estimated is the *Mendel sequence* proposed in [KOR 82] and often used since as a benchmark. It is depicted in panel (a) of Figure 5.1. The IR is represented in (c). The output (b) is corrupted by additive noise, with a signal-to-noise ratio (SNR, i.e., the mean empirical power of $h \star r$ over variance of the noise) of 10 dB. (e) shows the result of a matched filtering, i.e., filtering of observations y (continuous line) by the time-reverse of h , followed by thresholding at 0.05 (dotted lines) giving the detections marked by \times s. It

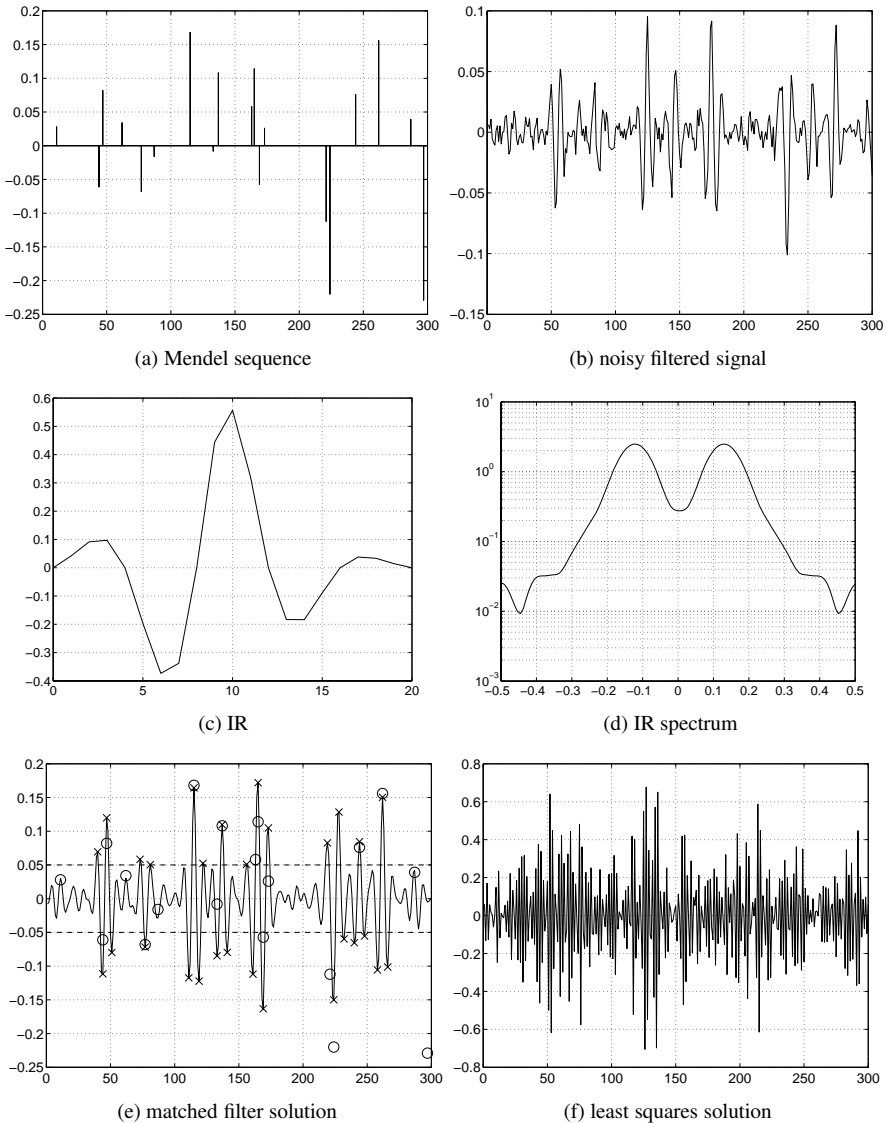


Figure 5.1. Synthetic filter based on Mendel sequence (a), synthetic spike train signal introduced in [KOR82]. This signal is filtered by the IR (c), additive Gaussian noise of variance $r_b = 5.4 \cdot 10^{-5}$ is added to give the signal (b). The crosses of (e) represent the spikes detected by a matched filter followed by thresholding, while the circles mark the spikes to be found. The least-squares solution (f) is worthless for detection purposes

turns out that no threshold gives acceptable false alarm and correct detection rates at the same time. Furthermore, the least-squares solution (f) is unusable. This is a result of the limited bandwidth of the IR (d), when we are trying to find a broadband signal.

Following the methodology presented in Chapter 2, we want to add prior information about the rarity of non-zero samples on to input r . Since the solution is assumed to be broadband, no interaction between neighboring or distant samples is introduced in our priors. In a deterministic regularization framework, this leads us to look for solutions as minimizers of a penalized least squares criterion of the form:

$$J(\mathbf{r}, \mu) = \|\mathbf{y} - \mathbf{H}\mathbf{r}\|^2 + \mu \sum_m \phi(r_m), \quad (5.4)$$

in which the regularization term can be decomposed into a sum of monovariate functions to express the *a priori* independence of samples.

The difficulty is to specify function ϕ to fully characterize the type of regularization. ϕ needs to be chosen in a way that favors values close to zero while, at the same time, exceptionally allowing values very different from zero. The definition of ϕ thus raises the same difficulties as those found in image restoration for correlated models (see Chapter 6). The evolution followed in the choice of the functions ϕ is also similar: quadratic functions, then non-convex functions and, more recently, non-quadratic convex functions.

5.2.1. Quadratic regularization

The simplest choice is to take $\phi(r) = r^2$, since the resulting estimator is linear:

$$\hat{\mathbf{r}} = (\mathbf{H}^T \mathbf{H} + \mu \mathbf{I})^{-1} \mathbf{H}^T \mathbf{y}. \quad (5.5)$$

This can be implemented using low-cost Wiener or Kalman filtering techniques [CRU 74, DEM 84, FAT 80, WOO 75]. These techniques are very widely used in deconvolution but do not, in general, yield interesting results for spike restoration unless the spectral content of the IR is sufficiently rich in high frequencies. They can only perform *spectral equalization* in the band where the SNR is large enough (see (4.6) and discussion in Chapter 4).

As an illustration, let us go back to the synthetic example of Figure 5.1. In Figure 5.2 we present the solution $\hat{\mathbf{r}}$ of (5.5) for various values of the regularization parameter. The solution obtained for $\mu = 1$ is clearly over-regularized, while the one obtained for $\mu = 10^{-3}$ is under-regularized. There is no intermediate value of μ that gives a satisfactory result as the IR has too strong a low frequency behavior (see Figure 5.1d), which makes linear processing ineffective, even with a moderate noise level.

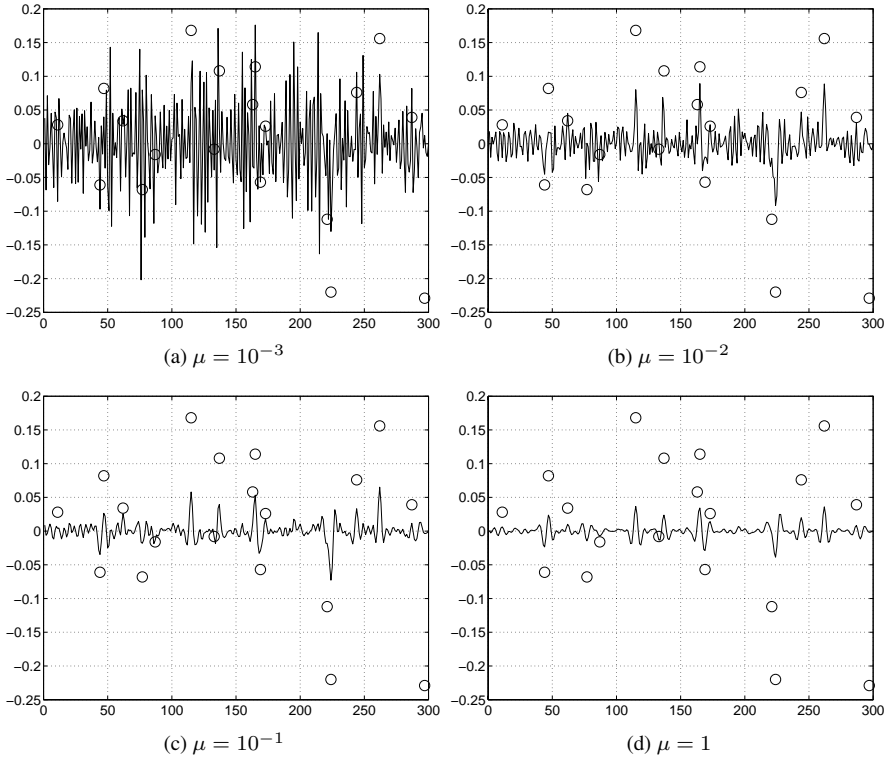


Figure 5.2. Results provided by quadratic regularization of the example of Figure 5.1. The circles mark the spikes of the Mendel sequence. The result \hat{r} is plotted for various values of the regularization parameter μ

5.2.2. Non-quadratic regularization

The quadratic function penalizes large values too much and various authors have looked into functions that increase more slowly than r^2 as r increases.

The first contribution proposing $\phi(r) = |r|$ appeared in the geophysics community in the late 1970s [TAY 79] and was followed, among others, by [OLD 86] and, more recently, [O'B 94]. The various authors try to deal with the algorithmic problems connected with the non-differentiable nature of the criterion by using simplex-type algorithms. Using a simplex means giving up the quadratic term on the data and replacing it by an absolute-value term. In practice, the method does, indeed, produce solutions with a marked spiky character. It converges in a finite time but is very costly, one of the most costly that exists according to Kaaresen's comparison [KAA 98a].

Much more efficient techniques to minimize (5.4) with $\phi(r) = |r|$ have been made available by recent advances in the field of variable selection based on the homotopy approach of [OSB 00] (see also [EFR 04]). Moreover, such techniques do not calculate $\hat{r}(\mu)$ for single values of μ . Instead, they characterize the full family of solutions $\{\hat{r}(\mu), \mu > \mu_{\min}\}$, which yields efficient ways of tuning μ either on qualitative or quantitative grounds.

On the other hand, Saito [SAI 90], taking his inspiration from the works of Leclerc [LEC 89] in computer vision, has proposed $\phi(r) = 1$ if r is non-zero, and $\phi(r) = 0$ otherwise. The regularizing part of J thus increases linearly with the number of non-zero samples. The corresponding criterion is neither convex nor differentiable (ϕ is discontinuous at $r = 0$) and Saito recommends gradual non-convexity techniques (GNC; see section 2.2.4). Saito's method produces solutions with a very marked spiky character but suffers from the optimization problems inherent in non-convex criteria.

More generally, the non-convex functions ϕ used in image restoration can be used here (see section 6.4). The corresponding models contain hidden decision processes, formally identical to the line processes in image restoration (see section 6.4.1). From this point of view, such models can be seen as close to a Bernoulli-Gaussian model as presented in section 5.3.

5.2.3. L2LP or L2Hy deconvolution

An alternative to quadratic regularization and non-convex or non-differentiable functions is to choose strictly convex and differentiable functions $\phi(r)$ that increase more slowly than r^2 when r increases. It is remarkable that the use of such functions should have been suggested only relatively recently [GAU 95]. Two classes of functions of this type have been put forward in the context of non-destructive testing (see Chapter 9): the function $|r|^p$ for $2 > p > 1$ and the hyperbolic function $\sqrt{T^2 + r^2}$, $T > 0$ with which “L2LP deconvolution” and “L2Hy deconvolution” are respectively associated. Such functions allow us to approach the behavior of function $|r|$, and thus to avoid over-penalizing large-amplitude reflectivities, while remaining differentiable and strictly convex. The strict convexity of ϕ ensures that of J and thus the existence and uniqueness of the solution and its continuity with respect to the data and parameters. The differentiability of the criterion allows standard descent techniques to be used that are very easy to adjust and very cost-competitive relative to the simplex or GNC techniques mentioned in section 5.2.2. From a practical standpoint, the uniqueness of the solution simplifies the initialization. Concretely, the continuity of the solution ensures the robustness of L2LP/L2Hy deconvolutions with respect to noise and errors in the models or the choice of parameters. Finally, the approach presented here is easy to understand, which makes it accessible to the non-expert user. The convex differentiable penalization approach is thus an excellent compromise in

terms of cost, simplicity, accuracy and robustness. Section 5.4 completes this analysis by comparative tests concerning robustness.

5.3. Bernoulli-Gaussian deconvolution

5.3.1. Compound BG model

The distinctive feature of the family of methods we are now going to look at is that they aim to explicitly materialize the presence of a spike and dissociate it from the spike amplitude r . To do this, they associate an auxiliary variable q with each sample of the signal. Each variable is binary and shows the presence ($q = 1$) or absence ($q = 0$) of a spike. The objective of the deconvolution is to simultaneously estimate the position (indicated by $q = 1$) and amplitude of each spike. The BG model corresponds to the simplest random model of this type, for which:

- Q is a Bernoulli variable of parameter $\lambda \triangleq \Pr(Q = 1) \ll 1$;
- the distribution of R , given that $Q = q$, is Gaussian and zero-mean with variance qr_x .

λ and r_x are two *hyperparameters* assumed to be known. The simplicity of the BG model in fact masks a difficulty for its use in maximum likelihood estimation: when $q = 0$, the distribution of r is Gaussian with zero mean and variance, i.e., a Dirac distribution. In general, Dirac distributions appear in the posterior likelihood function, thereby making a straightforward application of the MAP paradigm irrelevant. To get around the problem, it is possible to replace the BG model by a mixed Gaussian model with very small but non-zero variance. Another possibility is to consider the process, r_e , of the spike amplitudes: it is Gaussian, zero-mean with variance r_x , and is only defined when $q \neq 0$. This approach is developed below.

5.3.2. Various strategies for estimation

The previous set of hypotheses (observation model (5.3), white, stationary Gaussian noise, BG input model) allows the posterior likelihood of $(Q, R_e | Y = y)$ to be defined without ambiguity. However, Bayesian methodology leaves us a great deal of freedom in the choice of the type of likelihood to be optimized, even though there are limited choices in practice. Owing to the composite nature of the BG model, we can envisage performing the deconvolution either by maximizing the *joint* likelihood $p(r_e | q, y) \Pr(q | y)$, or by proceeding sequentially, estimating first q by maximizing the *marginal* likelihood $\Pr(q | y)$, then r_e by maximizing $p(r_e | \hat{q}, y)$. In simulation, the joint approach leads to poorer quality results than the marginal approach when the true values of the parameters are used. More precisely, the joint approach gives

a false alarm rate that is too high for an equivalent good detection rate. Nevertheless, it is possible to obtain comparable results for the two approaches with different hyperparameter settings.

The marginal likelihood does not give rise to difficulties of definition due to the Dirac distributions mentioned above, and the various definitions of the BG process found in the literature all lead to the same marginal likelihood. For these reasons, we describe the approach by maximum marginal likelihood below, making three points clear:

- 1) the joint and marginal criteria differ structurally only by a matrix determinant;
- 2) most of the methods used to optimize one of these criteria can also be applied to optimize the other (this is the case of the SMLR presented in section 5.3.4);
- 3) the \mathbf{q} conditional estimation of the amplitudes is the same in the joint and sequential approaches.

5.3.3. General expression for marginal likelihood

By applying Bayes' rule, the marginal likelihood can be written:

$$\Pr(\mathbf{q} | \mathbf{y}) \propto p(\mathbf{y} | \mathbf{q}) \Pr(\mathbf{q}). \quad (5.6)$$

The expression for $\Pr(\mathbf{q})$ results from Bernoulli's hypothesis of section 5.3.1

$$\Pr(\mathbf{q}) = \lambda^{M_e} (1 - \lambda)^{M - M_e}$$

where M_e is the number of non-zero components of vector \mathbf{q} . Considering the definition of BG processes given in section 5.3.1, the components of vector \mathbf{r}_e are Gaussian, zero-mean, independent and with variance r_x . As recalled in section 5.3.2, noise b is also white, Gaussian, zero-mean, stationary and with variance r_b . We thus deduce that:

$$p(\mathbf{y} | \mathbf{q}) = \mathcal{N}(\mathbf{0}, \mathbf{B}) \quad \text{with} \quad \mathbf{B} \triangleq r_x \sum_k \mathbf{h}_{t_k} \mathbf{h}_{t_k}^T + r_b \mathbf{I}, \quad (5.7)$$

where \mathbf{h}_n is the n th column of \mathbf{H} , because of the conditionally Gaussian nature of $(\mathbf{R}_e | \mathbf{Q} = \mathbf{q})$ and the linearity of input-output relationship (5.3). It should be stressed that \mathbf{B} can also be expressed:

$$\mathbf{B} = \mathbf{H} \mathbf{\Pi} \mathbf{H}^T + r_b \mathbf{I} \quad \text{with} \quad \mathbf{\Pi} \triangleq r_x \text{Diag} \{q(m)\}_{1 \leq m \leq M} \quad (5.8)$$

and it is the latter expression that we will use in what follows.

By setting aside the terms that do not depend on \mathbf{q} , we deduce from (5.6)-(5.7) that maximizing $\Pr(\mathbf{q} | \mathbf{y})$ is equivalent to maximizing:

$$L(\mathbf{q}) \triangleq -\mathbf{y}^T \mathbf{B}^{-1} \mathbf{y} - \log |\mathbf{B}| - 2M_e \log \frac{1 - \lambda}{\lambda}. \quad (5.9)$$

Before tackling the practical problems connected with this maximization, let us examine the second step of the sequential approach, i.e., the estimation of the spike amplitudes when the sequence \mathbf{q} is assumed to be known. Maximizing $p(\mathbf{r}_e | \hat{\mathbf{q}}, \mathbf{y})$ comes down to estimating the maximum *a posteriori* of the Gaussian variable \mathbf{r}_e observed through the linear system \mathbf{H} . According to the results stated previously (Chapter 3, section 3.8), the estimate takes the form:

$$\forall k, \quad (\hat{r}_e)_k = \mathbf{h}_{t_k}^T \mathbf{B}^{-1} \mathbf{y}$$

and can be calculated in the standard way except, perhaps, for very large signals.

Numerous choices – generally dictated by the operating conditions – remain to be made if we are to implement the maximization of $L(\mathbf{q})$. Particularly worth mentioning are the choices of recursive or batch processing, and the type of representation for the linear system h . A detailed examination of all these situations is beyond the scope of this chapter. For this reason, we will give a precise description of only one method, corresponding to hypotheses that will be stated, whilst nevertheless trying to bring out the *general* mechanisms of BG deconvolution and the trade-offs that always have to be made. We will then briefly mention other important techniques, highlighting the main characteristics and situating them with respect to the chosen method.

5.3.4. An iterative method for BG deconvolution

Due to the discrete nature of \mathbf{q} , the exact maximization of L is a combinatorial problem. Formally, the problem is simple as all we need to do is calculate L for the 2^M possible configurations of \mathbf{q} so as to find the maximizer of $\hat{\mathbf{q}}$. Unfortunately, this is impossible to envisage in practice, even for signals of moderate size; the calculation would be far too voluminous. To obtain a realistic method, we will confine ourselves to exploring a subset of the possible configurations of \mathbf{q} and, as far as possible, avoid calculating L by straightforward application of equation (5.9). To do this, we define the notion of *neighboring sequences* and iteratively maximize the likelihood on the neighborhoods thus defined. The efficiency of the resulting method essentially depends on three factors: the nature of the neighborhoods, the existence of simple formulae connecting the likelihood of two neighboring sequences, and the strategy for exploring the neighborhoods. In the example given here, we define the neighborhood of a sequence \mathbf{q}_0 as the set of sequences \mathbf{q}_k that differ from \mathbf{q}_0 by exactly one component. We then establish formulae connecting the likelihoods of \mathbf{q}_k and \mathbf{q}_0 . These formulae serve as the basis of a suboptimal SMLR-type [KOR 82] procedure for optimizing L , which consists of maximizing L over the whole neighborhood of a sequence \mathbf{q}_0 , then repeating the process until a local maximum is reached. We will look at some more refined variants in the next section.

In what follows, the indices 0 and k concern quantities relative to \mathbf{q}_0 and \mathbf{q}_k respectively and the k th vector of the canonical basis of \mathbb{R}^M is designated \mathbf{v}_k . To obtain

formulae for updating L that do not require many calculations, we introduce the following auxiliary quantities:

$$\begin{aligned}\mathbf{A} &\triangleq \mathbf{H}^T \mathbf{B}^{-1} \mathbf{H}, \\ \mathbf{w} &\triangleq \mathbf{H}^T \mathbf{B}^{-1} \mathbf{y}, \\ \rho_k &\triangleq \varepsilon_k r_x^{-1} + \mathbf{v}_k^T \mathbf{A}_0 \mathbf{v}_k,\end{aligned}$$

where ε_k takes the value ± 1 according to whether a spike is added to or taken away from \mathbf{q}_0 in position k . According to the expression for L established in equation (5.9), the relation between \mathbf{B}_k and \mathbf{B}_0 plays a big part in establishing the formulae we are looking for. From equation (5.8), we have:

$$\mathbf{\Pi}_k = \mathbf{\Pi}_0 + \varepsilon_k \mathbf{v}_k r_x \mathbf{v}_k^T,$$

which, by substitution in (5.7) and by application of the matrix inversion lemma, leads to:

$$\mathbf{B}_k^{-1} = \mathbf{B}_0^{-1} - \mathbf{B}_0^{-1} \mathbf{H} \mathbf{v}_k \rho_k^{-1} \mathbf{v}_k^T \mathbf{H}^T \mathbf{B}_0^{-1}. \quad (5.10)$$

From this we deduce that:

$$\mathbf{y}^T \mathbf{B}_k^{-1} \mathbf{y} = \mathbf{y}^T \mathbf{B}_0^{-1} \mathbf{y} - \mathbf{w}_0^T \mathbf{v}_k \rho_k^{-1} \mathbf{v}_k^T \mathbf{w}_0$$

and, using another traditional result [GOO 77, Appendix E], that:

$$|\mathbf{B}_k| = \varepsilon_k r_x \rho_k |\mathbf{B}_0|.$$

If we assume that all the auxiliary quantities relative to sequence \mathbf{q}_0 are known, the likelihood calculation for a sequence \mathbf{q}_k of its neighborhood can be performed using the following algorithm:

$$\mathbf{k}_k = \mathbf{A}_0 \mathbf{v}_k, \quad \rho_k = \varepsilon_k r_x^{-1} + \mathbf{v}_k^T \mathbf{k}_k, \quad (5.11)$$

$$L(\mathbf{q}_k) = L(\mathbf{q}_0) + \mathbf{w}_0^T \mathbf{v}_k \rho_k^{-1} \mathbf{v}_k^T \mathbf{w}_0 - \log(\varepsilon_k r_x \rho_k) - 2\varepsilon_k \log \frac{1-\lambda}{\lambda}. \quad (5.12)$$

Once the whole neighborhood of \mathbf{q}_0 has been explored, the sequence \mathbf{q}_k that maximizes L is chosen as a new starting point. To reduce the volume of the calculations, it is better not to completely re-evaluate the auxiliary quantities but to calculate them iteratively. From equation (5.10), we have the following formulae:

$$\mathbf{w}_k = \mathbf{w}_0 - \mathbf{k}_k \rho_k^{-1} \mathbf{v}_k^T \mathbf{w}_0 \quad (5.13)$$

$$\mathbf{A}_k = \mathbf{A}_0 - \mathbf{k}_k \rho_k^{-1} \mathbf{k}_k^T \quad (5.14)$$

and, after initialization, the two sets of equations above make up the complete BG deconvolution algorithm, (5.11)-(5.12) being used to explore the neighborhood of the current sequence and (5.13)-(5.14) to select a new sequence.

5.3.5. Other methods

The method presented above allows us to restore signals modeled as BG processes with the hypothesis of linear distortion and white, Gaussian, stationary observation noise. There are several other techniques for dealing with this problem and various extensions have also been proposed. Below, we look briefly at the most important of these methods.

If we restrict ourselves to the framework adopted so far, several elements have to be chosen: type of representation of the system IR, exact nature of the likelihood (joint or marginal), and technique for maximizing the likelihood. We stress once again that the choice of the type of likelihood has little influence on the algorithmic questions and appropriate adjustment of the hyperparameters generally gives similar results in both cases. It should, however, be pointed out that, for joint likelihood, the fact that we have to give the hyperparameter values that are far from their “empirical” values can be troublesome in a *non-supervised* framework (see section 5.4.2). In compensation, the joint likelihood offers a little more flexibility in the development of the optimization procedures.

The choice of how to represent the IR has a significant influence on the development of the maximization algorithms in as much as it affects the way the input-output relation of the system is written. If we ignore the all-pole representations, which are little used because of the constraints they impose on the phase of h , the first works on BG deconvolution were based on a zero-pole representation of the IR [KOR 82, MEN 83]. Due to its parsimonious character, this representation limits the memory resources and, to a lesser extent, the computing power required for the optimization procedure. The importance of such considerations has obviously greatly decreased since that time. Furthermore, using such a representation can lead to algorithms of a type and algebraic structure remarkably similar to those of the procedure described in section 5.3.4, even though the details of the calculations are notably more complex. Finally, it should be noted that most of the works of the last 10 years have used a representation of h by a discretized IR, which simplifies the algebra to be handled and takes advantage of the structure of matrix \mathbf{H} .

It is using the likelihood optimization techniques that BG deconvolution methods can be distinguished from one another. These techniques have three component parts:

- ① an algorithm for evaluating the increment of the likelihood when there is an elementary modification of the Bernoulli sequence \mathbf{q} ;
- ② an algorithm for updating all the quantities when a new current sequence \mathbf{q} is accepted;
- ③ a strategy for partial exploration of these sequences.

In the case where all the observed data \mathbf{y} are available (off-line processing), several deterministic likelihood maximization procedures have been proposed: SMLR and MMLR methods and their variants [CHA 96, KOR 82, MEN 90] use restricted neighborhoods (two neighboring sequences differ by one, or at most two, samples) and explore all the sequences that are neighbors of the current sequence then select the one that maximizes the likelihood before going on to a new iteration; ICM [LAV 93] methods, based on a neighborhood system similar to the previous one, select *one* neighboring sequence of the current sequence in a predetermined or random way and accept it if it leads to an increase in the likelihood before iterating the procedure; IWM techniques [KAA 97] are based on more extensive neighborhoods and compensate for the resulting increased numerical complexity by maximizing the *joint* likelihood, not sequentially but alternately with respect to variables \mathbf{q} and r_e , which considerably lightens the calculations. We should stress that these deterministic optimization methods do not guarantee convergence towards the global maximum of the likelihood.

The choice among the various methods should be guided by the trade-off to be made between the more or less exhaustive character of the exploration of possible sequences \mathbf{q} on the one hand, and the numerical complexity of the method on the other. This last point depends on the strategy chosen and the distribution of the volume of calculations between components ① and ② of the method.

We should also mention that recursive techniques suitable for on-line processing of the data have also been proposed [CHI 85, GOU 89, IDI 90]. These have the same three components as above but the recursive processing imposes significant constraints on the strategy for exploring sequences \mathbf{q} . In general, only the components of \mathbf{q} corresponding to the current or very recent samples of \mathbf{y} can vary, the more distant past of the Bernoulli sequence being fixed. As decisions concerning this more distant past are not called into question, the exploration of all sequences \mathbf{q} is more partial than in the previous case. This leads to a generally more modest performance, the price to be paid for recursive processing.

Note that components ① and ② of the deterministic BG deconvolution methods also lie at the heart of stochastic optimization methods of the simulated annealing type. The (theoretical) interest of these approaches is that they guarantee convergence towards the global maximum of the likelihood but, unfortunately, at the cost of very heavy calculations. Formulae such as (5.11)-(5.12) and (5.13)-(5.14) lend themselves well to the development of such approaches but they have been used very little in practice as the increase in volume of the calculations relative to deterministic optimization methods is out of all proportion with the potential improvement in the results.

An interesting extension would be to replace the BG model employed up to now by a “BG + Gaussian” model. In applications such as echography, this type of model enables us, among other things, to take account of low-amplitude reflectors located between high-amplitude reflectors. The influence of the use of this model on the

way the problem is formulated and the way the likelihood is expressed is relatively limited. Several deconvolution methods similar to those mentioned so far and based on deterministic [LAV 93, MEN 90] and stochastic [LAV 93] optimization procedures have been put forward. The case of a non-Gaussian distribution of the amplitudes has also been treated formally [KAA 97], but does not appear to have led to any practical applications.

Finally, BG deconvolution techniques based not on maximum likelihood but on posterior mean estimators have recently been proposed [CHE 96, DOU 97]. Such estimators require samples of $(\mathbf{q}, \mathbf{r}_e)$ to be drawn at random according to the distribution $p(\mathbf{r}_e | \mathbf{q}, \mathbf{y}) \Pr(\mathbf{q} | \mathbf{y})$, which is carried out using pseudo-random sampling procedures known as *Markov Chain Monte Carlo* (MCMC, see Chapter 7). The main interest of these approaches is that they allow the blind deconvolution problem (see section 5.5.2) to be treated elegantly. We should insist here that components ① and ② of the deterministic methods of BG deconvolution mentioned earlier are again at the heart of the algorithms for drawing samples according to the distribution $p(\mathbf{r}_e | \mathbf{q}, \mathbf{y}) \Pr(\mathbf{q} | \mathbf{y})$, and thus numerically highly efficient algorithms such as that of formulae (5.11)-(5.12) and (5.13)-(5.14) are once again very useful.

5.4. Examples of processing and discussion

In this section, we will discuss the comparative natures of the solutions obtained by BG deconvolution and L2Hy deconvolution in more detail and look at the practical considerations that distinguish these two methods.

5.4.1. Nature of the solutions

Figure 5.3a gives the result of a BG deconvolution by SMLR on the data of Figure 5.1b. The values used for the hyperparameters are $(\lambda, r_x, r_b) = (0.07, 0.01, 5.10^{-5})$, close to the “true” values used for the synthesis: $(\lambda^*, r_x^*, r_b^*) = (0.05, 0.01, 5.4.10^{-5})$. This result is better than that of Figure 5.3b obtained by L2Hy convex regularization with $\phi(r) = \sqrt{r^2 + T^2}$ ($\mu = 0.02$ and $T = 10^{-4}$), which is already very good. Appropriate thresholding of the latter gives results very close to the SMLR; only the group of four spikes around index 170 is still better restored by SMLR. In addition, three spikes are doubtless impossible to restore with the data used: the very low-amplitude one situated around 140 and the last two, which are minimally represented in the data because of the boundary hypothesis used. The BG deconvolution result obviously has a more markedly spiky character than the result from convex regularization. In general, the quality of the estimation of the spike amplitudes is lower with L2Hy deconvolution, as it is spread over several points. Finally, relative to the case of quadratic regularization, the results of BG/L2Hy deconvolution bring a clear improvement (compare Figures 5.3 and 5.2).

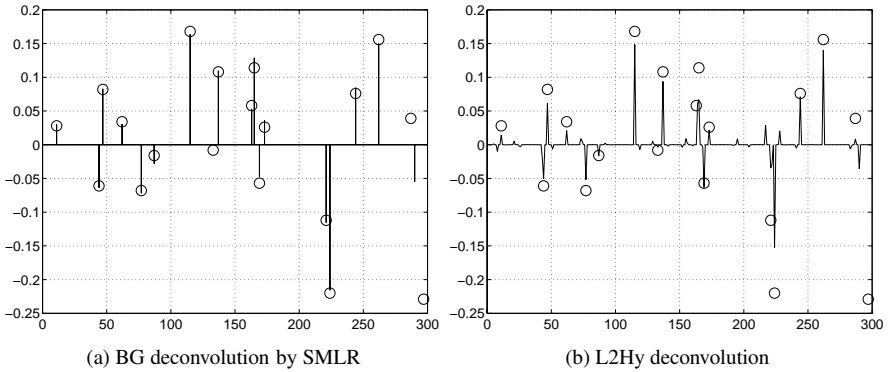


Figure 5.3. Comparison of behavior of BG and L2Hy deconvolution methods on the data of Figure 5.1b

To complete this analysis, we reran the two algorithms, with the same set of parameters, on data that only differed from those of Figure 5.1b by the realization of the Gaussian white noise that was added. We thus obtained Figure 5.4, which should be compared with Figure 5.3. The results of L2Hy deconvolution, Figures 5.3b and 5.4b, are qualitatively much closer to one another than the results of BG deconvolution, Figures 5.3a and 5.4a. Thresholding of 0.01 of both L2Hy solutions would give identical spike positions. This better stability of the L2Hy solution corresponds to the robustness expected for this method.

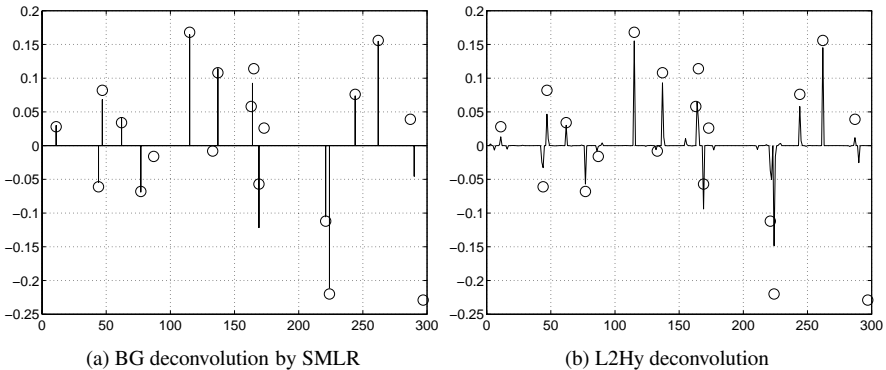
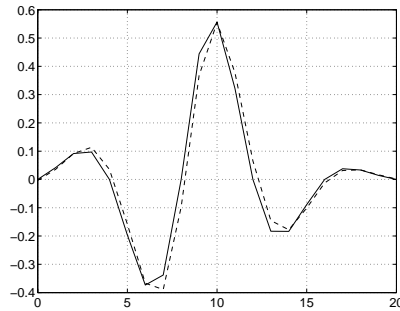
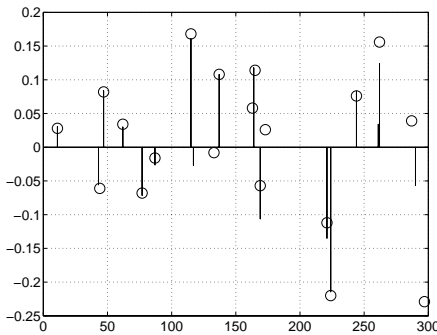


Figure 5.4. Test of robustness of the BG and L2Hy methods with respect to noise realization

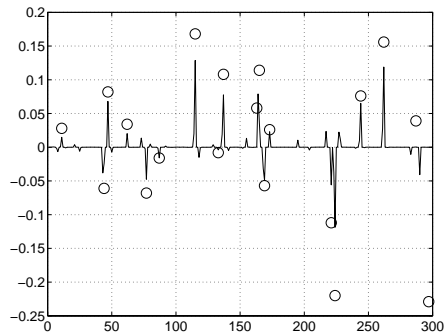
The quality of the spike deconvolution mainly depends on the bandwidth of the IR and on the SNR, parameters that condition the quality of the deconvolution in general. Let us add a characteristic of BG deconvolution: a high sensitivity to imperfect knowledge of the IR. As an illustration, the data of Figure 5.1b were processed again, for the same hyperparameters, but a perturbed IR. The perturbation was a *phase rotation* of ten degrees (see Chapter 9, section 9.4.3), the effect of which is shown in Figure 5.5a. It is a moderate perturbation that only affects the phase of the frequency response of the IR (its energy spectrum is unchanged). Comparing Figures 5.5b and c with Figures 5.3a and b, we note that the L2Hy deconvolution shows better robustness to perturbation generated by the inadequacy of the IR.



(a) original IR (—) and IR after a 10° phase rotation (---)



(b) BG deconvolution by SMLR



(c) L2Hy deconvolution

Figure 5.5. Test of robustness of the BG and L2Hy methods with respect to poor specification of the IR

5.4.2. Setting the parameters

In addition to the observations, both methods require initialization, here taken as zero, and specification of the numerical values for the IR and the hyperparameters. There are two hyperparameters for the L2Hy regularization and three for the BG

deconvolution described here (the method recommended by Kaaresen only needs two; see section 5.3.5). In fact, the L2Hy deconvolution does not take any decisions itself, so to obtain a result of the same kind as the BG, a threshold parameter would be needed. To these parameters must be added parameters having less influence but necessary for the test to stop the descent method used to minimize the criterion. In comparison, the SMLR deconvolution presented here does not use a parametrized stopping test as the optimization is performed in a discrete state space and a local minimum is obtained in a finite time.

In practice neither of these methods requires fine adjustment of the parameters. There is no adjustment method that is at the same time low-cost, universal and statistically well founded. The interested reader will find some empirical, common sense recipes in the articles mentioned above.

The statistically sound hyperparameter estimators are also the most cumbersome to use [CHA 96, GOU 92]. They bring in MCMC techniques of a similar nature to those mentioned in section 5.5.2. As an example, the method known as SEM, applied to BG deconvolution in [CHA 96], gives the estimates $(\hat{\lambda}, \hat{r}_x, \hat{r}_b) = (0.08, 0.008, 5.8 \cdot 10^{-5})$ in the case of the data of Figure 5.1b. Application of BG deconvolution with these parameters gives a result of quality between those of Figures 5.3a and 5.5b.

5.4.3. Numerical complexity

It is quite hard to precisely evaluate the numerical complexity of the BG or L2Hy deconvolution algorithms because of the iterative character of the methods and the difficulty of predicting the number of iterations needed for convergence. The BG methods get their efficiency from specialized techniques that exploit the sparsity of the spikes and only keep their advantage when the spikes come at a low rate (≤ 0.1). The specialization of these techniques also makes them more complex and more difficult to implement than the standard descent techniques used for L2Hy deconvolution, which are numerically more costly. The cost nevertheless remains very reasonable: processing the 300 samples of the synthetic example only takes about one second of a CG algorithm using the linesearch strategy proposed in [LAB 08], written in Matlab and run on a PC (Intel Pentium 4, 2 GHz, 1 GB).

5.5. Extensions

The methods presented in this chapter admit numerous extensions concerning:

- the structures of noise covariance matrix \mathbf{R} and observation matrix \mathbf{H} , which have so far been taken as diagonal and Toeplitz respectively;
- the estimation of the IR on the basis of observations \mathbf{y} , of unknown reflectivity;

- multichannel deconvolution, to take advantage of lateral correlations among signals received in neighboring positions, a standard context in NDE (see Chapter 9) and seismology. This theme has many connections with image restoration and will not be developed here. For further information, see [IDI 93, KAA 98b, LAV 91].

5.5.1. Generalization of structures of \mathbf{R} and \mathbf{H}

All the observations on methodology remain valid for any matrices \mathbf{R} and \mathbf{H} . Due to the matrix standpoint adopted here, a change in their structure has little effect on the algorithms – on a macroscopic scale – but the computational complexity may increase by a factor of N . However, when matrices \mathbf{R} and \mathbf{H} of the direct problem are strongly structured, the techniques presented can be generalized without any notable loss of numerical efficiency.

In the case of colored noise, effective implementations of BG deconvolution have been proposed in the case of noise modeled in autoregressive form, provided that the order of the model is not too high [CHA 93, MEN 83]. In fact, the most interesting extensions concern other structures of \mathbf{H} ; some particularly noteworthy extensions are:

- spectral ray analysis where \mathbf{H} represents a Fourier matrix. The spike train restoration problem is dealt with using extensions of BG techniques in [DUB 97, BOU 06], and by convex and non-convex regularization in [BOU 07] and [SAC 98], respectively;

- basis selection in decomposition into packets of wavelets: \mathbf{H} thus represents a wavelet transform [PES 96];

- “double BG” [CHA 93] and “double L2Hy” [GAU 01] deconvolution in which \mathbf{H} models any non-homogeneous phase rotations of the convolution kernel (see Chapter 9, section 9.4.2).

Note that, in these examples, matrix \mathbf{H} generally has many more columns than rows and the apparently very under-determined aspect of the underlying problem is overcome using impulse priors.

5.5.2. Estimation of the impulse response

In almost all applications, the IR – and thus the resulting matrix \mathbf{H} – is not an input of the problem: because the convolution is often only a crude physical model, the IR has no existence in itself and must be estimated. If possible, specific auxiliary measurements are used as in instrument calibration. If not, the problem can be tackled by *blind* deconvolution, i.e., both the IR and the reflectivity are estimated from the observed signal. This estimation is obviously valid up to an amplitude factor and a

time-shift factor, which cannot be identified from the data. The phase of the frequency response of the IR cannot be identified either if the input signal is Gaussian [LII 82]. From this point of view, a spiky input signal corresponds to a more favorable situation.

The main methods for estimating the IR are based on the various ways of defining, interpreting and exploiting the non-Gaussian character of the reflectivity. An account of these methods is given in Chapter 9.

In this section, we will only discuss the methods specifically using BG models or their variants, such as a Gaussian mixture. The simplest blind BG methods [GOU 86, KAA 98b] are based on maximization of the generalized likelihood (GML), defined as the probability distribution of all random quantities (observation, reflectivity, h) conditionally on all deterministic parameters (noise variances, λ , etc.). The generalized likelihood is formally defined by equation (3.9) in Chapter 3. Unlike maximization of the *exact* likelihood – defined as the probability of the observations alone, knowing the parameters; see Chapter 3, equation (3.3) – generalized maximum likelihood (GML) techniques have no asymptotic convergence properties. On the other hand, they are the only ones that can be implemented by simple iterative deterministic algorithms such as, for example, alternating an SMLR BG deconvolution step with a step to re-estimate h and the hyperparameters. In the implementation proposed by Kaarensen [KAA 98b], the GML gives good results for an observed signal of 1,000 samples synthesized with a wavelet of narrower bandwidth than that used in our simulations and a favorable SNR of 15 dB, *but with an impulse density parameter fixed in advance* (this parameter probably cannot be identified by GML). In addition, the method seems to be convergent for the estimation of the IR in the sense where a decrease in the SNR to 7 dB can be compensated for by processing 10 times as much data.

The most recent contributions on blind BG deconvolution [ROS 03, LAB 06] are based on MCMC techniques following Cheng *et al.* [CHE 96]. The MCMC techniques are the most statistically sound, since they give access to the exact likelihood, but they require intensive calculation. They consist of probabilizing the IR and hyperparameters and sampling the posterior law of the reflectivity, the IR and the hyperparameters conditionally to the observations. If the estimators chosen for the reflectivity and the IR are conditional expectations, they can be approached by averages of pseudo-random realizations drawn according to the posterior law. The example proposed by Cheng *et al.* [CHE 96] is composed of 2,000 samples obtained with three values of the SNR (26 dB, 18.6 dB and 4 dB) and a wavelet of bandwidth comparable to that of our simulations. The examples at 18.6 dB and 26 dB give good results on the IR and input, whereas the 4 dB case only gives acceptable results for the IR. It is difficult to compare the respective merits of the methods put forward by Cheng *et al.* and Kaarensen using simulation results, as none of the experimental conditions coincide. However, Cheng *et al.*'s method estimates the density of the impulses unlike Kaarensen's, for which this parameter is fixed. This illustrates the fact that a method

giving access to the exact likelihood offers more possibilities than a method based on maximization of the generalized likelihood. As regards real data processing, the potential of the MCMC method by Cheng *et al.* as been investigated by Rosec *et al.* [ROS 03] on marine seismic data. They show that these techniques enable us to improve seismic image resolution in at least two respects: better interface localization and layer detection.

Finally, although it remains implicit in the above mentioned articles, only the multichannel blind approaches seem to provide satisfactory results for the processing of real data.

5.6. Conclusion

We find ourselves faced with a problem of spike train deconvolution whenever two point sources become indiscernible in the observed signal because of the limited resolution of the sensor and the small distance between the sources. If we use the hypotheses of linearity and shift-invariance, after discretization the problem comes down to a discrete deconvolution in which the input is a discrete-time sparse spike train. The limited bandwidth of the sensor usually makes inverse filtering methods ineffective to restore the broadband character of the input.

We have presented two families of techniques aimed specifically at restoring spiky signals:

- the first favors estimation of the input by optimization of a criterion regularized on \mathbb{R}^n which penalizes non-impulse solutions;
- the second insists particularly on the source detection aspect and places a reduced number of spikes by optimizing a criterion that depends on the position of the impulses. As these positions are discrete, these methods solve a problem of combinatorial optimization, processed by suboptimal algorithms.

Finally, we have described two methods, L2Hy deconvolution by convex penalization and Bernoulli-Gaussian deconvolution by SMLR. These methods are representative of the two families respectively.

L2Hy deconvolution gives solutions with a marked spiky character, in the sense that they are composed of a majority of samples with a low value and a few samples with a modulus that is much higher. Impulse detection, if necessary, has to take place in a second step by thresholding these solutions. Due to the convexity of the optimized criterion, the solution obtained is continuous and depends on the observations, the IR and the hyperparameters, which ensures that the solution is less sensitive to these parameters. In comparison, BG deconvolution produces a solution that includes spike detection. The solutions provided are thus discontinuous with respect to the parameters. These characteristics of the solutions provided by the two methods are illustrated

by a synthetic example that allows them to be compared. In the conditions where the hypotheses of the model are properly fulfilled, BG deconvolution gives solutions of a quality better than or equivalent to L2Hy deconvolution, depending on the noise realization. In contrast, it is not as robust as the latter to modeling errors such as poor knowledge of the IR.

This crucial problem of modeling errors will be considered again in Chapter 9 where it will be analyzed by means of an example on real data from non-destructive testing using ultrasound. Chapter 9 illustrates the gap between the application of the techniques presented here and the practical processing of data, which, in addition to the observed signal, requires the specification of the quantities that were assumed to be known here but which are often unknown in practice, such as the IR and the hyperparameters. These quantities are all the more difficult to adjust in practice when the assumed convolutional model is only a rough, first-order representation of the underlying physics. Processing real data often makes it necessary to implement extensions of the methods presented here. Chapter 9 describes one of these extensions, sections 5.3.5 and 5.5 point the interested reader towards the variants and extensions that are not covered in detail in this book.

5.7. Bibliography

- [ATA 82] ATAL B. S., REMDE J. R., “A new method of LPC excitation for producing natural sounding speech at low bit rates”, in *Proc. IEEE ICASSP*, vol. 1, Paris, France, p. 614-617, May 1982.
- [BOU 06] BOURGUIGNON S., CARFANTAN H., “Spectral analysis of irregularly sampled data using a Bernoulli-Gaussian model with free frequencies”, in *Proc. IEEE ICASSP*, Toulouse, France, May 2006.
- [BOU 07] BOURGUIGNON S., CARFANTAN H., IDIER J., “A sparsity-based method for the estimation of spectral lines from irregularly sampled data”, *IEEE J. Selected Topics Sig. Proc.*, vol. 1, num. 4, p. 575-585, Dec. 2007, Issue: Convex Optimization Methods for Signal Processing.
- [CHA 93] CHAMPAGNAT F., IDIER J., DEMOMENT G., “Deconvolution of sparse spike trains accounting for wavelet phase shifts and colored noise”, in *Proc. IEEE ICASSP*, Minneapolis, MN, p. 452-455, 1993.
- [CHA 96] CHAMPAGNAT F., GOUSSARD Y., IDIER J., “Unsupervised deconvolution of sparse spike trains using stochastic approximation”, *IEEE Trans. Signal Processing*, vol. 44, num. 12, p. 2988-2998, Dec. 1996.
- [CHE 96] CHENG Q., CHEN R., LI T.-H., “Simultaneous wavelet estimation and deconvolution of reflection seismic signals”, *IEEE Trans. Geosci. Remote Sensing*, vol. 34, p. 377-384, Mar. 1996.
- [CHI 85] CHI C. Y., GOUTSIAS J., MENDEL J. M., “A fast maximum-likelihood estimation and detection algorithm for Bernoulli-Gaussian processes”, in *Proc. IEEE ICASSP*, Tampa, FL, p. 1297-1300, Apr. 1985.

- [CRU 74] CRUMP N. D., "A Kalman filter approach to the deconvolution of seismic signals", *Geophysics*, vol. 39, p. 1-13, 1974.
- [DEM 84] DEMOMENT G., REYNAUD R., HERMENT A., "Range resolution improvement by a fast deconvolution method", *Ultrasonic Imaging*, vol. 6, p. 435-451, 1984.
- [DJU 96] DJURIC P., "A model selection rule for sinusoids in white Gaussian noise", *IEEE Trans. Signal Processing*, vol. 44, num. 7, p. 1744-1751, July 1996.
- [DOU 97] DOUCET A., DUVAUT P., "Bayesian estimation of state space models applied to deconvolution of Bernoulli-Gaussian processes", *Signal Processing*, vol. 57, p. 147-161, 1997.
- [DUB 97] DUBLANCHET F., IDIER J., DUVAUT P., "Direction-of-arrival and frequency estimation using Poisson-Gaussian modeling", in *Proc. IEEE ICASSP*, Munich, Germany, p. 3501-3504, Apr. 1997.
- [EFR 04] EFRON B., HASTIE T., JOHNSTONE I., TIBSHIRANI R., "Least angle regression", *Annals Statist.*, vol. 32, num. 2, p. 407-451, 2004.
- [FAT 80] FATEMI M., KAK A. C., "Ultrasonic B-scan imaging: Theory of image formation and a technique for restoration", *Ultrasonic Imaging*, vol. 2, p. 1-47, 1980.
- [GAU 95] GAUTIER S., LE BESNERAIS G., MOHAMMAD-DJAFARI A., LAVAYSSIÈRE B., "Data fusion in the field of non destructive testing", in K. HANSON (Ed.), *Maximum Entropy and Bayesian Methods*, Santa Fe, NM, Kluwer Academic Publ., p. 311-316, 1995.
- [GAU 01] GAUTIER S., IDIER J., CHAMPAGNAT F., VILLARD D., "Restoring separate discontinuities from ultrasonic data", in *Review of Progress in Quantitative Nondestructive Evaluation*, AIP Conf. Proc. Vol 615(1), Brunswick, ME, p. 686-690, July 2001.
- [GOO 77] GOODWIN G. C., PAYNE R. L., *Dynamic System Identification. Experiment Design and Data Analysis*, Academic Press, 1977.
- [GOU 86] GOUTSIAS J. K., MENDEL J. M., "Maximum-likelihood deconvolution: An optimization theory perspective", *Geophysics*, vol. 51, p. 1206-1220, 1986.
- [GOU 89] GOUSSARD Y., DEMOMENT G., "Recursive deconvolution of Bernoulli-Gaussian processes using a MA representation", *IEEE Trans. Geosci. Remote Sensing*, vol. GE-27, p. 384-394, 1989.
- [GOU 92] GOUSSARD Y., "Blind Deconvolution of sparse spike trains using stochastic optimization", in *Proc. IEEE ICASSP*, vol. IV, San Francisco, CA, p. 593-596, Mar. 1992.
- [HOG 74] HOGBOM J., "Aperture synthesis with a non-regular distribution of interferometer baselines", *Astron. Astrophys. Suppl.*, vol. 15, p. 417-426, 1974.
- [IDI 90] IDIER J., GOUSSARD Y., "Stack algorithm for recursive deconvolution of Bernoulli-Gaussian processes", *IEEE Trans. Geosci. Remote Sensing*, vol. 28, num. 5, p. 975-978, Sep. 1990.
- [IDI 93] IDIER J., GOUSSARD Y., "Multichannel seismic deconvolution", *IEEE Trans. Geosci. Remote Sensing*, vol. 31, num. 5, p. 961-979, Sep. 1993.
- [KAA 97] KAARESEN K. F., "Deconvolution of sparse spike trains by iterated window maximization", *IEEE Trans. Signal Processing*, vol. 45, num. 5, p. 1173-1183, May 1997.

- [KAA 98a] KAARESEN K. F., "Evaluation and applications of the iterated window maximization method for sparse deconvolution", *IEEE Trans. Signal Processing*, vol. 46, num. 3, p. 609-624, Mar. 1998.
- [KAA 98b] KAARESEN K. F., "Multichannel blind deconvolution of seismic signals", *Geophysics*, vol. 63, num. 6, p. 2093-2107, Nov.-Dec. 1998.
- [KOR 82] KORMYLO J. J., MENDEL J. M., "Maximum-likelihood detection and estimation of Bernoulli-Gaussian processes", *IEEE Trans. Inf. Theory*, vol. 28, p. 482-488, 1982.
- [KWA 80] KWAKERNAAK H., "Estimation of pulse heights and arrival times", *Automatica*, vol. 16, p. 367-377, 1980.
- [LAB 06] LABAT C., IDIER J., "Sparse blind deconvolution accounting for time-shift ambiguity", in *Proc. IEEE ICASSP*, vol. III, Toulouse, France, p. 616-619, May 2006.
- [LAB 08] LABAT C., IDIER J., "Convergence of conjugate gradient methods with a closed-form stepsize formula", *J. Optim. Theory Appl.*, vol. 136, num. 1, Jan. 2008.
- [LAV 91] LAVIELLE M., "2-D Bayesian deconvolution", *Geophysics*, vol. 56, p. 2008-2018, 1991.
- [LAV 93] LAVIELLE M., "Bayesian deconvolution of Bernoulli-Gaussian processes", *Signal Processing*, vol. 33, p. 67-79, 1993.
- [LEC 89] LECLERC Y. G., "Constructing simple stable description for image partitioning", *Int. J. Computer Vision*, vol. 3, p. 73-102, 1989.
- [LII 82] LII K. S., ROSENBLATT M., "Deconvolution and estimation of transfer function phase and coefficients for non Gaussian linear processes", *Annals Statist.*, vol. 10, num. 4, p. 1195-1208, 1982.
- [MEN 83] MENDEL J. M., *Optimal Seismic Deconvolution*, Academic Press, New York, NY, 1983.
- [MEN 90] MENDEL J. M., *Maximum-Likelihood Deconvolution – A Journey into Model-Based Signal Processing*, Springer Verlag, New York, NY, 1990.
- [O'B 94] O'BRIEN M. S., SINCLAIR A. N., KRAMER S. M., "Recovery of a sparse spike time series by L_1 norm deconvolution", *IEEE Trans. Signal Processing*, vol. 42, num. 12, p. 3353-3365, Dec. 1994.
- [OLD 86] OLDENBURG D. W., LEVY S., STINSON K., "Inversion of band-limited reflection seismograms: Theory and practice", *Proc. IEEE*, vol. 74, p. 487-497, 1986.
- [OSB 00] OSBORNE M. R., PRESNELL B., TURLACH B. A., "A new approach to variable selection in least squares problems", *IMA J. Numer. Anal.*, vol. 20, num. 3, p. 389-403, 2000.
- [PES 96] PESQUET J.-C., KRIM H., LEPORINI D., HAMMAN E., "Bayesian approach to best basis selection", in *Proc. IEEE ICASSP*, Atlanta, GA, p. 2634-2637, May 1996.
- [ROS 03] ROSEC O., BOUCHER J.-M., NSIRI B., CHONAVEL T., "Blind marine seismic deconvolution using statistical MCMC methods", *IEEE Trans. Ocean. Eng.*, vol. 28, num. 3, p. 502-512, 2003.

- [SAC 98] SACCHI M. D., ULRYCH T. J., WALKER C. J., “Interpolation and extrapolation using a high-resolution discrete Fourier transform”, *IEEE Trans. Signal Processing*, vol. 46, num. 1, p. 31-38, Jan. 1998.
- [SAI 90] SAITO N., “Superresolution of noisy band-limited data by data adaptive regularization and its application to seismic trace inversion”, in *Proc. IEEE ICASSP*, Albuquerque, NM, p. 1237-1240, Apr. 1990.
- [STO 89] STOICA P., MOSES R. L., FREIDLANDER B., SÖDERSTRÖM T., “Maximum likelihood estimation of the parameters of multiple sinusoids from noisy measurements”, *IEEE Trans. Acoust. Speech, Signal Processing*, vol. 37, num. 3, p. 378-392, Mar. 1989.
- [TAY 79] TAYLOR H., BANKS S., MCCOY F., “Deconvolution with the L_1 norm”, *Geophysics*, vol. 44, num. 1, p. 39-52, 1979.
- [VAN 68] VAN TREES H. L., *Detection, Estimation and Modulation Theory*, Part 1, John Wiley, New York, NY, 1968.
- [WAL 97] WALTER E., PRONZATO L., *Identification of Parametric Models from Experimental Data*, Springer-Verlag, Heidelberg, Germany, 1997.
- [WON 92] WONG K. M., REILLY J. P., WU Q., QIAO S., “Estimation of directions of arrival of signals in unknown correlated noise, part I: The MAP approach and its implementation”, *IEEE Trans. Signal Processing*, vol. 40, num. 8, p. 2007-2017, Aug. 1992.
- [WOO 75] WOOD J. C., TREITEL S., “Seismic signal processing”, *Proc. IEEE*, vol. 63, p. 649-661, 1975.

Chapter 6

Deconvolution of Images

6.1. Introduction

As explained in Chapter 1, constructing admissible solutions for ill-posed problems such as image deconvolution necessarily implies restricting ourselves to a limited class of solutions, given some prior knowledge. In the case of images, the generally expected solution possesses a certain degree of *local regularity*, measurable by norms of derivatives or finite directional differences. More precisely, it is legitimate to assume that the variations of intensity are limited except in transitions between regions, if we exclude the case of strongly textured regions.

This chapter is above all devoted to how to take account of this qualitative property of regularity “almost everywhere” as prior knowledge. Research work on the subject is plentiful and the target field of application particularly vast. In Chapters 12 to 14 of this book we will see that this property is also an essential regularizing tool in imaging for inverse problems other than deconvolution.

Looking for a discrete solution by minimizing a penalized criterion is one of the simplest techniques. In some cases, such a discrete solution can be presented as an approximation of a continuous solution defined in a functional framework. However, the conditions for a functional solution to exist and be unique are mathematically more difficult to establish. In practice, immersion in the functional framework is not an indispensable prerequisite. It is touched on in this chapter, without the in-depth mathematical treatment that would be necessary to guarantee the existence of functional

solutions. Similarly, the connection between penalization of the intensity variables and calculation of the solution by *isotropic* or *anisotropic diffusion* is mentioned.

The Bayesian probabilistic interpretation could also have served as the mathematical framework, justifying the minimization of penalized criteria by looking for the *maximum a posteriori* estimator (see Chapter 3). There again, immersion in this statistical framework is not an indispensable precondition for handling the basic tools given in this chapter. On the other hand, it becomes necessary when certain “advanced” tools are introduced, such as resampling techniques, simulated annealing, and the estimation of hyperparameters by maximum likelihood. This is why these two chapters will refer much more explicitly to the Bayesian framework.

This chapter is principally divided into three and progresses in approximately chronological order:

- Their structural simplicity and ease of implementation justify the fact that the linear solutions obtained by the minimization of “Tikhonov-type” penalized least squares criteria play a central, historical role in the “inverse filtering” domain. The definition, and also the limitations, of these solutions are discussed in section 6.2. As far as their practical calculation is concerned, Chapter 4 is partly devoted to the subject.

- In the 1980s, following on from work such as that by the Geman brothers [GEM 84], a more sophisticated and ambitious approach appeared, which consisted not only of estimating an image from imperfect data, but also of associating with it a step for *detection* of discrete *hidden* variables materializing outlines or regions. Section 6.3 is devoted to the *detection-estimation* methods which result from this.

- Since the 1990s, matters have started to evolve towards greater simplicity. This has led to penalty functions being chosen from a large family of convex, non-quadratic functions and the detection step has been abandoned in favor of solving the problem by *convex optimization* (section 6.4).

For the sake of completeness, we should add that minimizing penalized criteria is not the only possible option in image deconvolution. More specifically, *multiresolution* methods, based on decomposing the observed image in a wavelet domain, have been developed in recent years and applied to astronomical [STA 02] and satellite [KAL 03] imaging in particular.

6.2. Regularization in the Tikhonov sense

6.2.1. Principle

6.2.1.1. Case of a monovariate signal

Following the approach developed by Tikhonov in [TIK 77] and earlier articles (the oldest date from 1963), the penalty function for candidate solutions is a quadratic

norm on the signal and its derivatives, measuring their regularity. Let us suppose first of all that a discrete or continuous *monovariate function* (a 1D signal rather than a 2D image) x^* is to be estimated from imperfect data:

$$\mathbf{y} = Hx^* + \text{noise}, \quad (6.1)$$

where H is a (bounded) linear operator. The approach introduced by Tikhonov requires the choice of a regularizing function $\|Dx\|^2$, where D is also a linear operator, e.g. a differential operator. The estimated solution \hat{x} is then defined as the minimizer of:

$$\mathcal{J}(x) = \|\mathbf{y} - Hx\|^2 + \lambda \|Dx\|^2, \quad (6.2)$$

where λ is a regularization parameter (strictly positive). In practice, it is possible to impose conditions that are not very restrictive and ensure that \mathcal{J} is strictly convex and has a unique minimizer. In particular, this is the case if $\|Dx\|$ is a norm for x .

The first term of \mathcal{J} in (6.2) is a quadratic norm that penalizes the difference between the data \mathbf{y} and the observation model for an admissible function x . In what follows, we will concern ourselves more with the second term of \mathcal{J} , i.e., the construction of the regularizing functional. In Tikhonov's original contribution, x is a function of a continuous variable of an interval $\Omega \subset \mathbb{R}$ in \mathbb{R} , and:

$$\|Dx\|^2 = \sum_{r=0}^R \int_{\Omega} c_r(s) (x^{(r)}(s))^2 ds, \quad (6.3)$$

where the weights c_r are strictly positive functions and $x^{(r)}$ is the r th order derivative of x . Qualitatively, it is clear that such a choice corresponds to a prior hypothesis on the smoothness of the signal to be estimated x^* .

6.2.1.2. Multivariate extensions

Multivariate extensions have been proposed. In dimension d , to penalize only the gradient of x (noted ∇x here), we can define:

$$\|Dx\|^2 = \int_{\Omega} \|\nabla x(\mathbf{s})\|^2 d\mathbf{s} = \int_{\Omega} \sum_{i=1}^d \left(\frac{\partial x}{\partial s_i}(\mathbf{s}) \right)^2 d\mathbf{s}. \quad (6.4)$$

The penalty functionals can also involve partial derivatives of higher order. In imaging ($\Omega \subset \mathbb{R}^2$), several contributions bring in partial derivatives of order two (e.g. [TER 83]).

If functional \mathcal{J} defined by equation (6.2) has a minimizer, this minimizer is the solution of a Euler-Lagrange equation [AUB 06]). In the case of equation (6.4), this can be written:

$$H^* Hx - \lambda \Delta x = H^* \mathbf{y} \quad (6.5)$$

(where $\Delta = \sum_{i=1}^d \partial^2 / \partial s_i^2$ is the Laplacian with respect to the space coordinates and H^* is the adjoint operator of H), with the boundary condition:

$$\left. \frac{\partial x}{\partial n} \right|_{\partial\Omega} = 0, \quad (6.6)$$

where $\partial\Omega$ represents the boundary of Ω and n a normal vector outward to $\partial\Omega$.

6.2.1.3. Discrete framework

Discrete equivalents of functionals such as (6.3) or (6.4) appear in two types of work.

Some try to approach the minimizer of functional \mathcal{J} by discrete approximation, usually by means of a scheme involving finite differences. For example, the penalizing functional (6.4) can be approached by:

$$\sum_{m=2}^M \sum_{n=1}^N (x_{m,n} - x_{m-1,n})^2 + \sum_{m=1}^M \sum_{n=2}^N (x_{m,n} - x_{m,n-1})^2 \quad (6.7)$$

in the case of a rectangular domain Ω , divided into squares using $M \times N$ sites (m, n) having the same step distance for the rows and columns. For conventional discretization schemes (finite differences, and also finite elements), there are results showing the convergence of the discrete minimizers towards the function that minimizes \mathcal{J} when the grid becomes finer (see, for example, [TER 83]).

Others directly position themselves in the discrete framework or discretize the signal to be restored over a fixed number of points right from the start, before posing the inversion problem. Penalty functions such as (6.7) are then often found, e.g. in [TIT 85], without reference to the regularization functions they come from.

6.2.2. Connection with image processing by linear PDE

Partial derivative equations (PDE) were introduced for the restoration of noisy images by Koenderink [KOE 84], on the basis of a mathematical analogy between the calculation of regular images and the diffusion of heat. In this formulation, it is assumed that a noisy image $y = \{y(\mathbf{s}), \mathbf{s} \in \Omega\}$ has been observed over a *continuum* Ω (typically a block \mathbb{R}^2 , which is obviously not realistic in practice). A series of images $x(\mathbf{s}, t)$ is defined depending on an outside parameter t (the time or the *scale*) and an equation for the evolution of x with time is introduced. The first diffusion equation applied to images, known as the *heat equation*, is the following linear parabolic PDE:

$$\frac{\partial x(\mathbf{s}, t)}{\partial t} = \Delta x(\mathbf{s}, t), \quad (6.8)$$

with the conditions (6.6) and

$$x(\mathbf{s}, 0) = y(\mathbf{s}), \quad (6.9)$$

The system formed of (6.6), (6.8) and (6.9) applies to image restoration from a noisy version y (i.e., $H = Id$ in (6.1)). The idea is to “diffuse” the intensity of the image isotropically as time goes on. It can be shown that this diffusion process corresponds to the application of a circular, Gaussian, linear convolution operator of variance $2t$ on y :

$$x(\mathbf{s}, t) = (G_t \star y)(\mathbf{s}) \quad \text{with} \quad G_t(\mathbf{s}) = \exp(-\|\mathbf{s}\|^2 / 4t) / 4\pi t.$$

In this formulation by PDE, the time parameter can be considered as a scale parameter. In the image $x(\mathbf{s}, t)$, the information at scales smaller than t is blurred but the information at coarser scales is preserved. For $t \rightarrow \infty$, $x(\mathbf{s}, t)$ tends towards a constant (the mean of y): the diffusion process must be stopped after a certain time, and this plays the role of a regularization parameter. To avoid the solution converging towards a constant solution, and to be able to take account of a degradation process H , (6.8) can be replaced by a “biased” PDE [NOR 90]:

$$\frac{\partial x(\mathbf{s}, t)}{\partial t} = \mu H^* (y(\mathbf{s}) - Hx(\mathbf{s}, t)) + \Delta x(\mathbf{s}, t). \quad (6.10)$$

A stationary state of this equation is obtained by making the left hand side zero, thus by solving the Euler-Lagrange equation (6.5) for $\mu = 1/\lambda$. Also note that equation (6.10) can be written in the form:

$$\frac{\partial x(\mathbf{s}, t)}{\partial t} = -\nabla \mathcal{J}(x),$$

which, once discretized (in time), corresponds to a gradient algorithm for the minimization of functional (6.2).

6.2.3. Limits of Tikhonov’s approach

When the image to be restored is made up of distinct regions, or when the aim of the imaging procedure is to display isolated defects in a homogeneous medium (in non-destructive testing), Tikhonov’s method proves limited in its capacity to detect discontinuities, or even to indicate their approximate positions.

As an illustration, let us consider the following simulated experiment: let $\mathbf{y} = [y_1, \dots, y_N]^T$ be a vector of noisy data, regularly sampled using a function x^* on $[0, 1]$ which is univariate and piecewise smooth: $y_n = x^*(n/N) + b_n$. The function x^* and the vector \mathbf{y} are represented in Figure 6.1.

With $R = 2$, $c_0 = c_1 = 0$, $c_2 = 1$ and $\|\mathbf{y} - Hx\|^2 = \sum_{n=0}^N (y_n - x(n/N))^2$, equation (6.2) can be written:

$$\mathcal{J}(x) = \|\mathbf{y} - Hx\|^2 + \lambda \int_0^1 (x''(s))^2 ds. \quad (6.11)$$

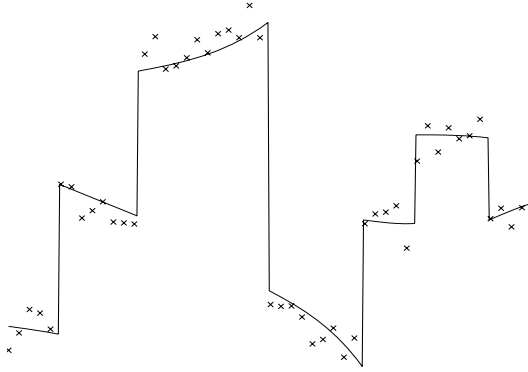


Figure 6.1. A piecewise smooth monovariate function x^* and 50 data items containing noise y_n , $n = 1, \dots, N = 50$

Let us introduce a discrete approximation for (6.11), by finite differences:

$$J_M(\mathbf{x}) = \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 + \lambda M^3 \sum_{m=2}^{M-1} (2x_m - x_{m-1} - x_{m+1})^2, \quad (6.12)$$

where $\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 = \sum_{n=1}^N (y_n - x_{nM/N})^2$, $\mathbf{x} = [x_1, \dots, x_M]^T$, and M is a multiple of N . We then define the estimated vector $\hat{\mathbf{x}}^\lambda$ as the minimizer of J_M . The point to point convergence of $\hat{\mathbf{x}}^\lambda$ towards the unique minimizer of (6.11) when $M \rightarrow \infty$ is a traditional result [NAS 81]. Figure 6.2 represents $\hat{\mathbf{x}}^\lambda$, a vector of length $M = 400$, obtained for the “best” value of λ in the L_1 sense, i.e., the value that minimizes

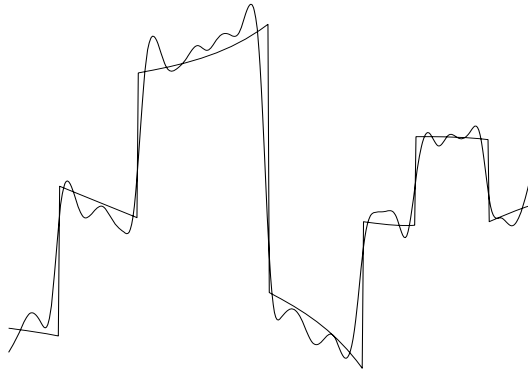


Figure 6.2. Smooth linear estimate $\hat{\mathbf{x}}^\lambda$ obtained as the minimizer of equation (6.12), for $M = 400$ and for the optimum value of λ in the L_1 sense: $C(\hat{\mathbf{x}}^\lambda, x^*) = 18.16\%$

$C(\widehat{\mathbf{x}}^\lambda, x^\star)$, with:

$$C(\mathbf{x}, x^\star) = \sum_{m=1}^M |x_m - x^\star(m/M)| / \sum_{m=1}^M |x^\star(m/M)|.$$

This procedure for choosing hyperparameters is artificial, since it requires x^\star to be known, but it allows relatively fair comparisons among methods for estimating the signal we are looking for.

The solution $\widehat{\mathbf{x}}^\lambda$ (Figure 6.2) is not satisfactory, as it is uniformly smooth. In comparison, a simple piecewise linear interpolation of successive data gives an error of the norm L_1 of 17.04%, which is notably lower.

If the number I and the positions $\boldsymbol{\tau} = [\tau_1, \dots, \tau_I]$ of the discontinuities were known, a suitable approach would be to replace criterion \mathcal{J} by:

$$\mathcal{J}_{\boldsymbol{\tau}}(x) = \|\mathbf{y} - Hx\|^2 + \lambda \sum_{i=0}^I \int_{\tau_i}^{\tau_{i+1}} (x''(s))^2 ds,$$

with $\tau_0 = 0$ and $\tau_{I+1} = 1$, a criterion to be minimized in x without the hypothesis of derivability in τ_1, \dots, τ_I . (This is why we do not write the sum of the integrals in the form $\int_0^1 (x(s''))^2 ds$.) Similarly, in 2 dimensions, if the discontinuities form a known set Γ of curves in the plane, regularity is imposed everywhere except in Γ . This idea can be generalized to any dimension d for a set $\Gamma \subset \Omega$ of dimension $d - 1$.

The discrete equivalent can be expressed as:

$$J_{\boldsymbol{\ell}}(\mathbf{x}) = \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 + \lambda M^3 \sum_{m=2}^{M-1} (1 - \ell_m) (2x_m - x_{m-1} - x_{m+1})^2, \quad (6.13)$$

where $\boldsymbol{\ell} = [\ell_2, \dots, \ell_{M-1}]$ is a binary vector of *edge variables*: $\ell_m = 1$ corresponds to the presence of a discontinuity¹ at position m .

In practice, this approach is very limited as our ignorance of the positions of the discontinuities is an integral part of the problem. The following sections are devoted to the main ideas and tools introduced in the signal and image processing community to deal with the question of restoration of piecewise regular functions.

1. Note that these “second order” discontinuities correspond to breaks in the slope. The breaks in the intensity of x^\star (Figure 6.1) thus correspond to two consecutive slope breaks: for example, $\ell_m = \ell_{m+1} = 1$ for a break in intensity between x_m and x_{m+1} .

6.3. Detection-estimation

6.3.1. Principle

In the mid-1980s, the controlled management of discontinuities made a big step forward, both theoretically and practically [BLA 87, GEM 84, MUM 85, TER 83]. The idea was put forward of considering the problem of estimating x and that of detecting the discontinuities (in the form of τ , ℓ or, more generally, set Γ of discontinuities, according to the context) *jointly*. To do this, joint minimization of criteria such as $\mathcal{J}_\tau(x)$ in (x, τ) or $J_\ell(x)$ in (x, ℓ) is not adequate. It is not difficult to see that this strategy leads to a maximum number of discontinuities (i.e., $\Gamma = \Omega$ in continuous and $\ell = [1, \dots, 1]^T$ in discrete). This does not happen if a “price to be paid” $\alpha > 0$ is imposed per discontinuity introduced [EVA 92, MUM 85]. The result is an *augmented criterion*:

$$K(x, \ell) = J_\ell(x) + \alpha \sum_{m=2}^{M-1} \ell_m \quad (6.14)$$

in the discrete univariate case and, in the continuous univariate case,

$$\mathcal{K}(x, \tau) = \mathcal{J}_\tau(x) + \alpha I. \quad (6.15)$$

In the functional case of dimension 2, the penalization of the discontinuities becomes proportional to the total length of the curves making up Γ . More generally, in dimension d , it can be written $\alpha \mathcal{H}^{d-1}(\Gamma)$, where \mathcal{H}^{d-1} is the Hausdorff measure of dimension $d - 1$ [EVA 92, MUM 85].

In equations (6.14) and (6.15), the penalty depends only on the number of discontinuities, not on their relative positions: they are said to be *uncoupled*. Variants can

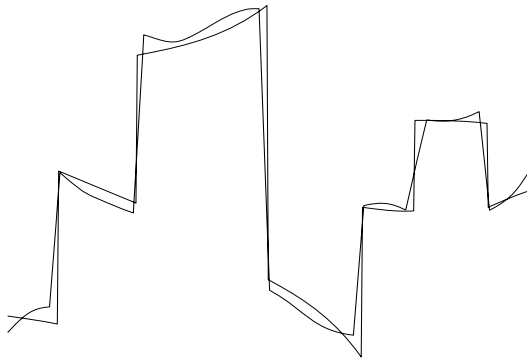


Figure 6.3. Smooth piecewise estimate $\hat{x}^{\alpha;\lambda}$ obtained as the joint minimizer of equation (6.14) for optimum values of λ, α in the L_1 sense: $C(\hat{x}^{\alpha;\lambda}, x^*) = 16.22\%$

be envisaged; for example, it could be decided that nearby discontinuities should be over-penalized by introducing a sliding price to be paid that falls with increasing distance between neighboring discontinuities. When the penalty depends on the relative positions of the discontinuities, the edge variables are said to be *interactive*. In image segmentation, specific models of interactive variables have been introduced (see Chapter 7).

Finally, we can point out several characteristics common to works such as [BLA 87, GEM 84, MUM 85, TER 86] in image restoration and computer vision:

- The image discontinuities are modeled explicitly, in the form of a set in the continuous setting, and more simply by Boolean variables in the discrete setting. These variables can be qualified as *hidden* (with respect to the observation procedure) in as far as they are excluded from observation equation (6.1). Conceptually and practically, the use of hidden variables provides a wide variety of tools for taking prior knowledge into consideration in a penalized form.

- In most cases, the augmented criterion is *half-quadratic* (HQ): a function K is said to be HQ if it depends on two sets of variables, say \mathbf{x} and ℓ , in such a way that K is quadratic in \mathbf{x} but not in (\mathbf{x}, ℓ) . Numerous works in image restoration have exploited this characteristic recently and section 6.5 is devoted to it.

6.3.2. Disadvantages

The calculating cost is the main practical disadvantage of detection-estimation. Without mentioning the continuous case, numerically heavy methods are already necessary to correctly handle the combinatory problem induced by the introduction of the binary variables in the discrete case. Most of these methods are based on the principle of *relaxation*, in a stochastic (*simulated annealing*; see [GEM 84] and Chapter 7) or deterministic (*continuation methods* such as *graduated non-convexity* – [BLA 87]; see also [NIK 98, NIK 99]) framework. Figure 6.3 was calculated by GNC deterministic relaxation.

Several authors [BOU 93, LI 95] have stressed the lack of stability of the estimate $\hat{\mathbf{x}}^{\alpha;\lambda}$ as another weakness of detection-estimation. This lack of stability comes from the fact that $\hat{\mathbf{x}}^{\alpha;\lambda}$ is not a continuous function of the data. In other words, Hadamard's third condition is not satisfied by $\hat{\mathbf{x}}^{\alpha;\lambda}$, and the problem is still ill-posed despite the penalization. Figure 6.4 illustrates this aspect, following an example given in [LI 95].

In fact, $\hat{\mathbf{x}}^{\alpha;\lambda}$ is a piecewise continuous function of the data, a behavior that is intrinsic in the discontinuity detection capacity of such an approach. In fact, solving a decision problem with multiple hypotheses comes down to partitioning the data space (here, \mathbb{R}^N) into as many regions E_k as there are competing hypotheses [BRÉ 94]. E_k is defined as the subset of the values of \mathbf{y} for which the hypothesis adopted is the

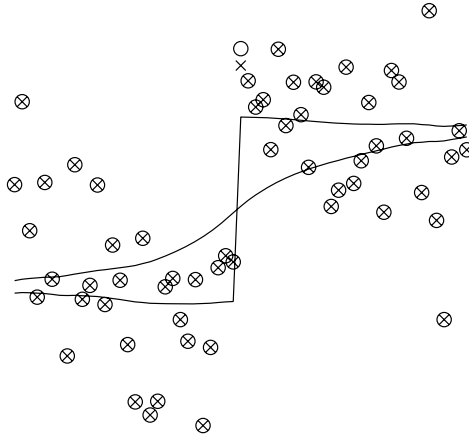


Figure 6.4. *Instability of $\hat{x}^{\alpha;\lambda}$ as a function of the data: continuous line, two estimates $\hat{x}^{\alpha;\lambda}$ coming from the same data set, except for one value. The data are represented by circles for one set and by crosses for the other. Parameters λ and α are kept constant*

k th according to the decision rule considered. Here, there are as many hypotheses as possible values of ℓ , i.e., 2^{M-2} , and $\hat{x}^{\alpha;\lambda}$ is continuous when \mathbf{y} is inside any E_k (for fixed ℓ , $\hat{x}^{\alpha;\lambda}$ minimizes quadratic criterion (6.13)). In other words, the discontinuities occur only at the edges of subsets, i.e., when the variation of the data causes a change of decision. In certain cases, far from being a disadvantage, this characteristic is desirable since it testifies to automatic decision making. On the other hand, if the signal restoration is only a decision-making aid for a human expert, detection-estimation is doubtless not to be recommended as it is both risky and numerically costly.

6.4. Non-quadratic approach

Penalized non-quadratic approaches took on increasing importance in signal and image restoration just before the mid 1990s [BOU 93, KÜN 94, LI 95]. The principle is to replace Tikhonov quadratic penalization by another even function better able to preserve discontinuities. For example,

$$\mathcal{J}_\phi(x) = \|\mathbf{y} - Hx\|^2 + \lambda \int_0^1 \phi(x''(s)) ds \quad (6.16)$$

is a generalization of (6.11) for the continuous 1D case, and similarly:

$$J_\phi(x) = \|\mathbf{y} - \mathbf{H}x\|^2 + \frac{\lambda}{M} \sum_{m=2}^{M-1} \phi\left(\frac{2x_m - x_{m-1} - x_{m+1}}{1/M^2}\right) \quad (6.17)$$

generalizes (6.12) for the discrete 1D case. Or again,

$$\int_{\Omega} \phi(\|\nabla x(\mathbf{s})\|) d\mathbf{s} \quad (6.18)$$

generalizes (6.4) for the continuous 2D case, and:

$$\nu^2 \sum_{m=2}^M \sum_{n=1}^N \phi\left(\frac{x_{m,n} - x_{m-1,n}}{\nu}\right) + \nu^2 \sum_{m=1}^M \sum_{n=2}^N \phi\left(\frac{x_{m,n} - x_{m,n-1}}{\nu}\right) \quad (6.19)$$

is a possible generalization of (6.7) for the discrete 2D case (if necessary, parameter ν is a discretization step). Penalty function structure (6.19) is the most usual in image restoration practice. As [AUB 97, section 2] points out, it does not correspond to a discretized version (with finite differences) of equation (6.18) but rather to:

$$\int_{\Omega} \phi\left(\frac{\partial x(\mathbf{s})}{\partial s_1}\right) d\mathbf{s} + \int_{\Omega} \phi\left(\frac{\partial x(\mathbf{s})}{\partial s_2}\right) d\mathbf{s},$$

which is not invariant under rotation of the image axes. As for equation (6.18), an associated discrete scheme can be written:

$$\nu^2 \sum_{m=2}^M \sum_{n=2}^N \phi\left(\frac{1}{\nu} \sqrt{(x_{m,n} - x_{m-1,n})^2 + (x_{m,n} - x_{m,n-1})^2}\right). \quad (6.20)$$

Starting from (6.19) and to approach a model that is invariant under rotation, it is more usual to add supplementary diagonal terms [HUR 96]:

$$\nu^2 \sum_{m=2}^M \sum_{n=2}^N \phi\left(\frac{x_{m,n} - x_{m-1,n-1}}{\nu\sqrt{2}}\right) + \nu^2 \sum_{m=2}^M \sum_{n=2}^N \phi\left(\frac{x_{m-1,n} - x_{m,n-1}}{\nu\sqrt{2}}\right), \quad (6.21)$$

than to opt for criterion (6.20). On this subject, see [BLA 87, section 6.1.1].

The essential thing is thus to choose the behavior of function ϕ . In order to preserve the edges between homogeneous regions and thus to authorize large variations of the estimate at some points, ϕ must increase more slowly than a parabola. The choice of ϕ and the related arguments are finally very close to the use of *robust norms* in statistics [HUB 81, REY 83]. Two groups of functions have mainly been proposed in the literature.

L_2L_1 functions

These are even, non-constant functions that are convex C^1 , C^2 at 0 and asymptotically linear. A simple, often used example is the branch of the hyperbola (Figure 6.6c and [CHA 97]):

$$\phi(u) = \sqrt{\eta^2 + u^2}, \quad \eta > 0.$$

The minimizer of (6.17) that corresponds to this choice is represented in Figure 6.5. It provides a fairly good compromise between restoration of the smooth areas and preservation of the edges.

Other families of convex functions give a qualitatively close result, for example the norms L_p [BOU 93]:

$$\phi(u) = |u|^p, \quad 1 \leq p < 2$$

(which are not C^2 at 0, and even not differentiable for $p = 1$, and increase faster than the linear regime), or again, the “fair” function [REY 83], which, although it remains convex, increases a little more slowly than the linear regime:

$$\phi(u) = |u/\eta| - \log(1 + |u/\eta|), \quad \eta > 0.$$

Certain, more rare contributions propose penalizations that are convex but do not act on the derivatives or finite differences of the function to be estimated. Worth noting are functions inspired from entropy [O’S 95], such as, for the restoration of positive images:

$$\sum_{m=2}^M \sum_{n=1}^N (x_{m,n} - x_{m-1,n}) \log \frac{x_{m-1,n}}{x_{m,n}} + \sum_{m=1}^M \sum_{n=2}^N (x_{m,n} - x_{m,n-1}) \log \frac{x_{m,n-1}}{x_{m,n}}. \quad (6.22)$$

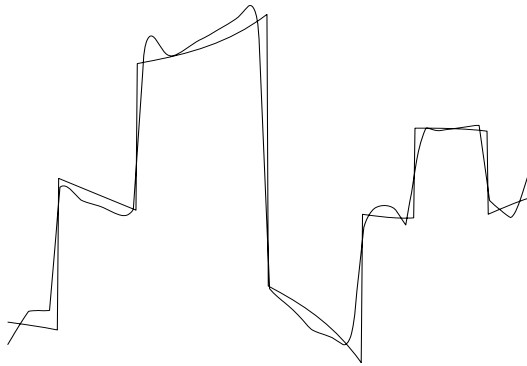


Figure 6.5. Smooth, nonlinear estimate $\hat{x}^{\eta,\lambda}$ obtained as the minimizer of equation (6.17), with $\phi(u) = \sqrt{\eta^2 + u^2}$ for optimal values of λ and s in the L_1 sense:
 $C(\hat{x}^{\eta,\lambda}, x^*) = 16.57\%$

L_2L_0 functions

These are non-constant, even functions, C^2 at 0, increasing on \mathbb{R}_+ and asymptotically constant. In consequence, the price to be paid for a variation of intensity

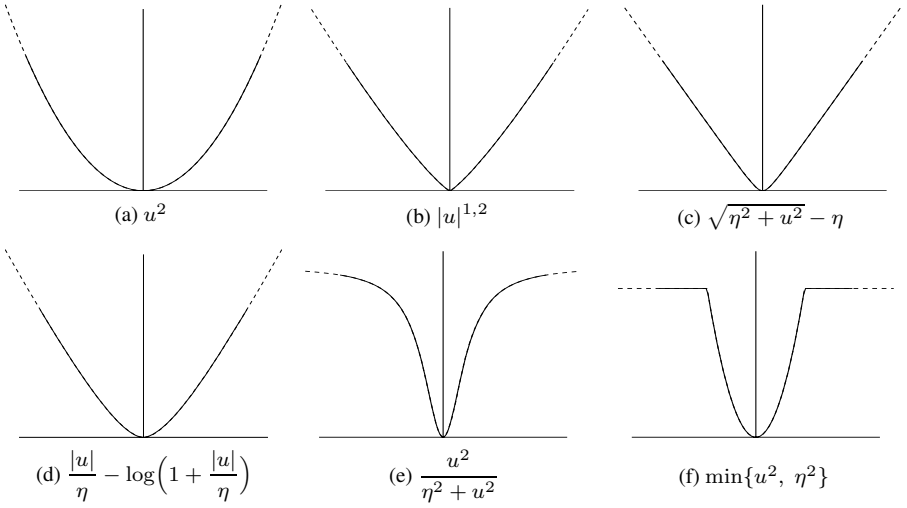


Figure 6.6. Examples of penalizing functions, classified by rapidity of increase at infinity.
Among the first four, which are convex, (c) and (d) are L_2L_1 .
The last two, non-convex, are L_2L_0

increases with its amplitude, but tends towards a constant. A typical example, given in [GEM 87] and shown in Figure 6.6e, can be written:

$$\phi(u) = \frac{u^2}{\eta^2 + u^2}, \quad \eta > 0.$$

L_2L_0 functions are not convex and the global minimization of functional (6.16) does not necessarily have a mathematical sense. However, in the discrete setting, the global minimization of function (6.17) is always mathematically well-defined, as it is posed in a space of finite dimension. From a practical point of view, of course, the possible existence of local minima makes global minimization of function (6.17) much more difficult. The minimizer of (6.17) is not represented here for the example treated; its behavior is very similar to that of the solution obtained by detection-estimation (Figure 6.3).

Variants are sometimes recommended in the literature, e.g. [GEM 92] advocates even, convex functions, increasing on \mathbb{R}_+ , which are thus non-derivable at 0, such that:

$$\phi(u) = \frac{|u|}{\eta + |u|}, \quad \eta > 0.$$

Comparisons

In practice, it is observed that the two classes, L_2L_1 and L_2L_0 , give very different solutions in terms of behavior and computing cost.

On the one hand, the L_2L_1 approach ensures the convexity of criteria \mathcal{J}_ϕ and J_ϕ . This property ensures the existence of a unique global minimum for equation (6.16) (in an appropriate space, such as the space $BV(\Omega)$ of functions with bounded variations [EVA 92]). The same is true for equation (6.17). Moreover, the convex criteria do not admit local minima, the convergence towards a global minimizer of equation (6.17) is thus proved for standard minimization algorithms (of the gradient type or for descent coordinate by coordinate, e.g. [BER 95]). Another interesting property of the solution thus obtained is its “stability” [BOU 93, KÜN 94, LI 95], that is to say, it satisfies Hadamard’s third condition, unlike the solutions obtained by detection-estimation.

On the other hand, the L_2L_0 approach shares the characteristics of the “detection-estimation” approach: the restored boundaries are very clear, the solution is only piecewise stable depending on the data and hyperparameters – see Figure 6.7; and the computing cost of algorithms that avoid the possible local minima is high. The similarity of the characteristics of the L_2L_0 and “detection-estimation” approaches is not an accident, as the next section shows.

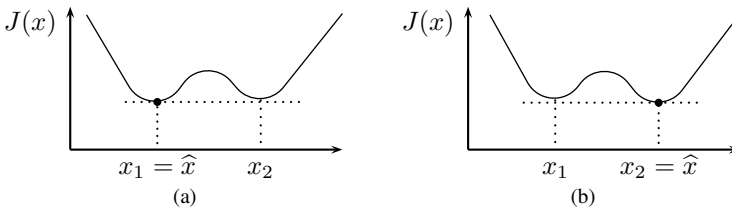


Figure 6.7. A small deformation of a multimodal criterion can completely change the position of the minimizer, when a local minimum becomes global. That is why the behavior of an estimate defined as the minimizer of a non-convex criterion can change abruptly after a slight modification of the data or the setting of the hyperparameters

6.4.1. Detection-estimation and non-convex penalization

Blake and Zisserman [BLA 87] observed that a HQ criterion (see section 6.3.1) with uncoupled edge variables such as equation (6.14) could be considered as an *augmented equivalent* of a certain non-quadratic criterion $J(x)$, in the sense that:

$$\min_{\ell \in \{0,1\}^{M-2}} K(\cdot, \ell) = J.$$

Qualifying criterion K as *augmented* stresses the fact that it depends not only on \mathbf{x} , but also on *auxiliary* variables ℓ , whereas J only depends on \mathbf{x} . More precisely, from equations (6.13) and (6.14), we have:

$$\begin{aligned}
 & \min_{\ell \in \{0,1\}^{M-2}} K(\mathbf{x}, \ell) \\
 &= \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 + \sum_{m=2}^{M-1} \min_{\ell_m \in \{0,1\}} \left(\lambda M^3 (1 - \ell_m) (2x_m - x_{m-1} - x_{m+1})^2 + \alpha \ell_m \right) \\
 &= \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 + \sum_{m=2}^{M-1} \min \left\{ \alpha, \lambda M^3 (2x_m - x_{m-1} - x_{m+1})^2 \right\} \\
 &= J_{\phi_\eta}(\mathbf{x}),
 \end{aligned}$$

where J_{ϕ_η} is defined by equation (6.17) for $\phi_\eta(u) = \min\{u^2, \eta^2\}$, i.e., for the *truncated quadratic* function represented in Figure 6.6f, with $\eta = \sqrt{M\alpha/\lambda}$. Consequently, $J = J_{\phi_\eta}$ and K have the same minimum and, more importantly, the same *minimizer* $\hat{\mathbf{x}}$ on all closed X of \mathbb{R}^M , in the sense that:

$$\hat{\mathbf{x}} \text{ minimizes } J \text{ on } X \iff \exists \hat{\ell} / (\hat{\mathbf{x}}, \hat{\ell}) \text{ minimizes } K \text{ on } X \times \{0, 1\}^{M-2}.$$

To find $\hat{\mathbf{x}}$ in practice, we therefore have the choice of concentrating on the minimization of J or on that of K . Blake and Zisserman define K as the detection-estimation criterion, deducing the expression for J from that for K and dealing with the problem of minimizing the non-convex criterion J by *graduated non-convexity* rather than attacking the combinatory problem posed by the minimization of K .

6.4.2. Anisotropic diffusion by PDE

In the same way as a strong connection can be established between *isotropic diffusion by PDE* and quadratic penalization (see section 6.2.2), it is also possible to link anisotropic filtering of the image by PDE and the minimization of non-quadratic functional criteria (see [TEB 98] and other articles in the same special issue). The first anisotropic PDE of images, i.e., taking the edges into consideration, was that introduced by Perona and Malik [PER 90]. The idea is to encourage diffusion in zones with weak gradients (corresponding to homogeneous areas), while preserving the areas with strong gradients (corresponding to edges). This anisotropic diffusion is formally expressed as follows²:

$$\frac{\partial x(\mathbf{s}, t)}{\partial t} = \operatorname{div} \left(c(\|\nabla x(\mathbf{s}, t)\|) \nabla x(\mathbf{s}, t) \right),$$

2. $\operatorname{div} = \sum_{i=1}^d \partial/\partial s_i$ is the spatial divergence.

with conditions (6.6). The coefficient of conduction $c(\cdot)$ is chosen as one in the uniform zones (with weak gradients) and tends to zero in the zones of strong gradients. Diffusion is thus delayed at the edges. Here again, the process has to be stopped after a certain time and this plays the role of a regularization parameter.

If a fidelity to data term is introduced in the PDE, as proposed by Nordström [NOR 90], the solution is likely to converge towards a suitable stationary state, from any initial condition $x(\mathbf{s}, 0)$:

$$\frac{\partial x(\mathbf{s}, t)}{\partial t} = \mu H^*(y(\mathbf{s}) - Hx(\mathbf{s}, t)) + \operatorname{div}(c(\|\nabla x(\mathbf{s}, t)\|) \nabla x(\mathbf{s}, t)). \quad (6.23)$$

This dynamic equation has similarities with the use of the penalizing functional (6.18). Euler's equation expressing a necessary condition for obtaining an extremum of (6.18) can be written:

$$H^*(Hx - y) - \lambda \operatorname{div} \left(\frac{\phi'(\|\nabla x\|)}{2 \|\nabla x\|} \nabla x \right) = 0.$$

This equation is verified by any asymptotic solution $x(\mathbf{s}, +\infty)$ of (6.23) if:

$$\mu = 1/\lambda \text{ and } c(t) = \phi'(t)/2t. \quad (6.24)$$

Functions c proposed by Perona and Malik and the corresponding ϕ are:

$$c(t) = \exp(-t^2) \implies \phi(t) = 1 - \exp(-t^2) + \text{const.}$$

$$c(t) = \frac{1}{1+t^2} \implies \phi(t) = \log(1+t^2) + \text{const.}$$

The first function ϕ is L_2L_0 , the second is a non-convex function (but with no asymptote) proposed in [HEB 89] for regularizing a reconstruction problem in medical imaging. In the case of a non-convex function ϕ , note that the intervention of a PDE such as (6.23) does not provide a “miracle” method for ensuring the existence of a global minimizer and performing the calculation.

6.5. Half-quadratic augmented criteria

The use of HQ criteria has recently imposed itself as a powerful numerical tool for image restoration with preservation of discontinuities [BRE 96, CHA 94, CHA 97, VOG 98]. As mentioned above in the framework of the “detection-estimation” approach, the HQ criteria first included *exclusively binary* edge variables that were either interactive or uncoupled. Following [GEM 92], this section shows that the HQ criteria cover a much larger domain. More precisely, many penalized, non-quadratic approaches admit an equivalent formulation in the HQ framework. This has a double consequence: from a formal point of view, this equivalence leads to a better understanding of the real choices of signal and image models; from a practical point of view, the HQ formulation furnishes new algorithmic tools for minimizing non-quadratic criteria.

6.5.1. Duality between non-quadratic criteria and HQ criteria

D. Geman *et al.* [GEM 92, GEM 95] generalized Blake and Zisserman's construction (section 6.4.1) to non-discrete auxiliary uncoupled variables. By reversing the construction order, they showed the existence of augmented HQ criteria K for broad classes of non-quadratic criteria J , in the sense where:

$$\inf_{b \in \mathcal{B}} K(\cdot, b) = J \quad (6.25)$$

for a set \mathcal{B} to be defined, in general different from $\{0, 1\}$. In a form that was subsequently extended [CHA 97, IDI 01], this construction applies to functions ϕ that satisfy the following hypotheses:

$$\begin{cases} \phi \text{ is even,} \\ \phi(\sqrt{\cdot}) \text{ is concave on } \mathbb{R}_+, \\ \phi \text{ is continuous at zero and } C^1 \text{ on } \mathbb{R} \setminus \{0\} \end{cases} \quad (6.26)$$

Under these conditions, the following duality relations can be obtained from conventional convex analysis results [ROC 70]:

$$\phi(u) = \inf_{b \in \mathbb{R}_+} (bu^2 + \psi(b)), \quad (6.27)$$

$$\psi(b) = \sup_{u \in \mathbb{R}} (\phi(u) - bu^2).$$

From equations (6.17) and (6.27), it is easy to deduce that equation (6.25) is true for the augmented HQ criterion:

$$K(\mathbf{x}, \mathbf{b}) =$$

$$\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 + \lambda M^3 \sum_{m=2}^{M-1} b_m (2x_m - x_{m-1} - x_{m+1})^2 + \frac{\lambda}{M} \sum_{m=2}^{M-1} \psi(b_m). \quad (6.28)$$

Formally, we observe that the first part of this criterion can be identified with $J_{1-\mathbf{b}}(\mathbf{x})$, where J_ℓ has been defined for binary ℓ by equation (6.13). To within a multiplying factor, the auxiliary variables b_m play the role of $1 - \ell_m$ in equations (6.13)-(6.14), but the b_m are not binary.

Whether they are convex or not, most of the functions ϕ proposed in the literature satisfy (6.26). For example, we have:

$$\phi(u) = \sqrt{\eta^2 + u^2}, \eta > 0 \implies \psi(b) = \begin{cases} \eta^2 b + 1/4b & \text{if } b \in]0, b_\infty = 1/2\eta], \\ 1 & \text{if } b \geq b_\infty. \end{cases}$$

$$\phi(u) = \frac{u^2}{\eta^2 + u^2}, \eta > 0 \implies \psi(b) = \begin{cases} (1 - \eta\sqrt{b})^2 & \text{if } b \in]0, b_\infty = 1/\eta^2], \\ 1 & \text{if } b \geq b_\infty. \end{cases}$$

6.5.2. Minimization of HQ criteria

6.5.2.1. Principle of relaxation

[GEM 92] points out the structural advantages connected with half-quadratic properties for the minimization of the augmented criterion K rather than J . More precisely, we proceed by *relaxation*, i.e., by considering the sub-problem with fixed \mathbf{x} , then fixed \mathbf{b} alternately until convergence is obtained. These two sub-problems are simple since one is quadratic and the other separable. [GEM 92] is specifically concerned with the case where ϕ is non-convex and develops an HQ version of simulated annealing based on the principle of alternate sampling.

Several other authors [BRE 96, CHA 94, CHA 97, VOG 98] have taken an interest in the deterministic counterpart, which is well suited to the case where ϕ is convex. Starting from an initial pair (\mathbf{x}, \mathbf{b}) , they intertwine minimization of \mathbf{x} with \mathbf{b} fixed and minimization of \mathbf{b} with \mathbf{x} fixed:

① augmented criterion $K(\mathbf{x}, \mathbf{b})$ is a quadratic function of \mathbf{x} . Calculating its minimizer for \mathbf{x} with \mathbf{b} fixed is therefore a simple problem of inverting a linear system. It is worth noting that this step corresponds to an adaptive version of Tikhonov regularization, in as much as the regularization parameter in front of each quadratic penalty term in expression (6.28) is multiplied by a factor b_m ;

② criterion $K(\mathbf{x}, \mathbf{b})$ being a separable function of variables b_m , it can be minimized with respect to \mathbf{b} in a parallel form. Moreover, the updating equation for each b_m is explicit: it can be shown that the infimum of equation (6.27) is reached for³:

$$\forall u \neq 0, \hat{b}(u) = \phi'(u)/2u, \quad (6.29)$$

prolonged by continuity in 0; the expression for ψ is thus not required for expressing $\hat{b}(u)$.

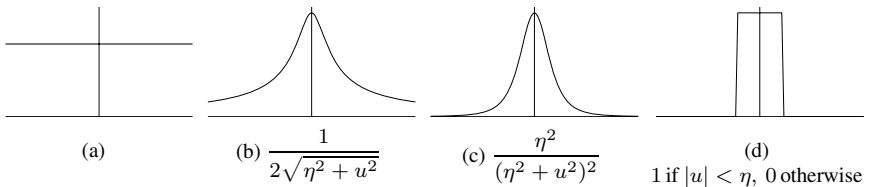


Figure 6.8. Function $\hat{b}(u) = \phi'(u)/2u$, respectively in the case where ϕ is: (a) quadratic (Figure 6.6a); (b) parabolic (Figure 6.6c); (c) the Geman and McClure function (Figure 6.6e); (d) truncated quadratic (Figure 6.6f)

3. Note that, in this expression of \hat{b} , the conduction coefficient (6.24) is retrieved.

6.5.2.2. Case of a convex function ϕ

When ϕ is convex, it has been shown that this procedure for descent *block of coordinates by block of coordinates*, given the name of ARTUR in [CHA 94], converges towards a unique minimum under broad conditions [CHA 97, IDI 01, NIK 05]. In a very slightly different formulation, belonging to robust estimation, a similar descent technique (identified as a *reweighted least squares* (RLS) method) and the associated convergence result have been known for a long time [YAR 85].

Although other descent techniques could be envisaged for calculating \hat{x} in the same conditions of regularity (e.g. conjugate gradient algorithm techniques [LAB 08]), the RLS approach has a remarkable advantage. If the inverse problem under consideration has previously been solved by Tikhonov quadratic penalization, the non-quadratic extensions can be obtained by trivial adaptation:

- introduce adaptive regularization into Tikhonov's method, which thus becomes step ① of the RLS method;
- add step ② to update auxiliary variables (6.29) (this is the only step that brings in the expression for ϕ); loop.

6.5.2.3. Case of a non-convex function ϕ

When the function ϕ is not convex, the RLS technique converges, if the minima are isolated, towards one of the local or global minima of J [DEL 98], such as other simple deterministic descent algorithms. It is thus necessary to use numerically heavier methods such as simulated annealing or GNC to avoid the local minima.

In the same way as there are versions of simulated annealing that take advantage of the half-quadratic nature of the augmented criterion [GEM 92], there is a certain interest in using an RLS version of GNC. Let us consider two functions ϕ_0 and ϕ_1 , which are L_2L_0 and L_2L_1 respectively. Let us take $\phi_\theta = (1 - \theta)\phi_0 + \theta\phi_1$ and suppose that θ goes from one to zero in small increments. We then find an interesting property: in step ② of the RLS, the expression for updating (6.29) based on ϕ_θ is a simple linear combination of the expressions valid for ϕ_0 and ϕ_1 respectively (whereas the dual function ψ_θ is not a linear combination of ψ_0 and ψ_1). Finally, as step ① does not depend on ϕ_θ , we thus immediately obtain an RLS version of GNC. The 1D estimate of Figure 6.3 was calculated using this principle.

6.6. Application in image deconvolution

6.6.1. Calculation of the solution

In this part, we will concentrate on the practical problem of finding the minimizer of a criterion of the type:

$$J(x) = \|y - Hx\|^2 + \lambda\Phi(x), \quad (6.30)$$

where \mathbf{x} is a discrete rectangular image composed of $M \times N$ pixels, represented in the form of a column vector (of length $M \times N$), concatenating the columns of the image. This arrangement is an algebraic device that is usual in imaging and allows the convolution of two vectors to be represented as the product of a matrix by a vector. The structure of convolution matrix \mathbf{H} is thus Toeplitz-block-Toeplitz (see Chapter 4, section 4.3.2).

Let us choose a penalization $\Phi(\mathbf{x})$ such as (6.19) or (6.20). If function ϕ is chosen to be quadratic, (6.19) and (6.20) are identical and their common minimizer in \mathbb{R}^{MN} is a linear function of the data (Chapter 2, section 2.2.2), the calculation of which is dealt with in Chapter 4. If the function ϕ is chosen to be L_2L_1 , (6.19) and (6.20) are different criteria, both convex. The minimization of (6.30) then comes under the case treated in section 2.2.3 of Chapter 2. We recall that the minimizer $\hat{\mathbf{x}}$ (if we assume it is unique, which is the case if (6.30) is strictly convex) is not, in general, an explicit function of the data, but that numerous iterative techniques based on successively decreasing the criterion engender algorithms that converge towards $\hat{\mathbf{x}}$.

Conceptually, descent coordinate by coordinate (also called the Gauss-Seidel method, or *relaxation*, or ICM for *iterated conditional modes*, or ICD for *iterative coordinate descent*) is certainly the simplest. It is a question of cyclicly solving the sub-problem of minimizing $J(\mathbf{x})$ as a function of pixel x_{mn} , the other pixels being fixed at their current values. One iteration of the algorithm corresponds to a complete scan of the image. Although hardly recommended as a generic minimization method [PRE 86, p. 303], this method is not always the slowest in the imaging context [SAU 93].

In general, the most efficient algorithms are those that exploit the convolutive nature of the problem. For example, in a gradient descent method, the gradient vector of the term $\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2$, which is written $2\mathbf{H}^T(\mathbf{H}\mathbf{x} - \mathbf{y})$, can be calculated by fast convolution in the Fourier domain, or in the direct domain by taking advantage of the sparse nature of \mathbf{H} . In the same way, it is possible to use the fact that a finite difference operator is a convolution. More generally, “modern” algorithms draw profit from the fast techniques developed for inverting linear systems with specific structures, particularly of the convolutive type. These techniques are sometimes used as they stand, or use “elementary bricks” in an iterative structure, e.g. of the weighted least squares type such as the ARTUR algorithm presented in section 6.5.2.2. Worthy of special mention is the recent appearance of *preconditioning* in image restoration [VOG 98].

LEGEND is an algorithmic structure close to ARTUR [CHA 94, IDI 01, ALL 06] (see also [YAR 85] in the robust estimation framework, [AUB 97, AUB 06] in the functional case and [GEM 95] for a simulated annealing version). Like ARTUR, LEGEND minimizes a HQ augmented criterion. It is in its special structure that the interest

of LEGEND lies. Let us consider a penalized criterion written:

$$J(\mathbf{x}) = \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 + \lambda \sum_{i \in \mathcal{I}} \phi(\mathbf{v}_i^T \mathbf{x}), \quad (6.31)$$

where $\mathcal{I} = \{1, \dots, I\}$ ($I < +\infty$) allows us to arbitrarily number the column vectors \mathbf{v}_i fixed, for example, to engender finite differences such as (6.17), (6.19), (6.21). An iteration of the LEGEND algorithm is written:

$$\mathbf{x}^{(k)} = (2\mathbf{H}^T \mathbf{H} + \lambda \mathbf{V} \mathbf{V}^T)^{-1} (2\mathbf{H}^T \mathbf{y} + \lambda \mathbf{V} \mathbf{b}^{(k)}), \quad (6.32)$$

where $\mathbf{b}^{(k)}$ is a column vector of size I calculated component by component:

$$b_i^{(k)} = \mathbf{v}_i^T \mathbf{x}^{(k-1)} - \phi'(\mathbf{v}_i^T \mathbf{x}^{(k-1)}), \quad (6.33)$$

and $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_I]$. The normal matrix $\mathbf{M} = 2\mathbf{H}^T \mathbf{H} + \lambda \mathbf{V} \mathbf{V}^T$ is taken to be invertible. Note that one iteration (6.32) performed for $\mathbf{b}^{(k)} = \mathbf{0}$ is sufficient to minimize (6.31) if ϕ is quadratic. If ϕ is non-quadratic, several iterations will be necessary to converge to a fixed point. The computing cost of (6.33) is assumed to be negligible, so the cost of each iteration remains that of the inversion of the linear system (6.32), the normal matrix of which does not vary during the iterations, a property that is specific to structures of the LEGEND type.

Depending on the case, matrix \mathbf{V} is Toeplitz in 1D, or Toeplitz-block-Toeplitz in 2D, or made up of blocks having these structures (the exact structure of \mathbf{V} also depends on the boundary conditions at $\partial\Omega$). Similarly, the convolution matrix \mathbf{H} is Toeplitz or Toeplitz-block-Toeplitz. Matrix \mathbf{M} thus has a structure *close to Toeplitz*, and the fast calculation of (6.32) can be envisaged.

Convergence towards the minimizer of (6.30), is guaranteed for all $\mathbf{x}^{(0)}$ if ϕ is a convex function such as an L_2L_1 function [IDI 01, NIK 05]. Local convergence is also ensured in a broader context [ALL 06].

6.6.2. Example

To illustrate this, let us calculate the minimizer of criterion (6.30), penalized in the form (6.19) (with $\nu = 1$), in the case of an L_2L_1 ϕ function, for an image deconvolution problem. The original image \mathbf{x}^* (202 rows, 99 columns, Figure 6.9c) is quantified on 256 grey levels. Blurring is Gaussian circular, with a standard deviation equivalent to one pixel, and the convolved image (190 rows, 87 columns, Figure 6.9a) has additive white Gaussian noise with a standard deviation of 3, corresponding to a SNR of 33 dB. ϕ is the Huber function [HUB 81], a C^1 function, defined piecewise:

$$\phi_\eta^{\text{Huber}}(u) = \begin{cases} u^2 & \text{if } |u| \leq \eta, \\ 2\eta|u| - \eta^2 & \text{otherwise.} \end{cases} \quad (6.34)$$

LEGEND is an appropriate algorithm for calculating the minimizer $\hat{\mathbf{x}}^{\eta, \lambda}$. As convolution by a circular Gaussian kernel is separable (i.e., can be decomposed into two 1D convolutions, one horizontal and one vertical), the algorithm proposed in [BRE 96], which is similar to ARTUR whilst making use of the separability, is also suitable. It is this algorithm that served to calculate the restored images of Figures 6.9d to f. Note that a conjugate gradient algorithm would also be a valid option [LAB 08].

For a quantitative evaluation of the restoration quality according to the hyperparameters (η, λ) , it is known that the norm L_1 of the error is often preferable to the norm L_2 . Empirically, the quasi-norm $L_{1/2}$:

$$c(\eta, \lambda) \propto \|\hat{\mathbf{x}}^{\eta, \lambda} - \mathbf{x}^*\|_{1/2} = \left(\sum_{i,j} \sqrt{\hat{x}_{i,j}^{\eta, \lambda} - x_{i,j}^*} \right)^2$$

seemed even better to us. In itself, this choice is not essential. The stakes are limited in comparison with the problem of *non-supervised* adjustment of the parameters, considered in Chapter 8. In addition, good restitution of some local details that are important from a visual standpoint is not necessarily taken into account in a global quality measure.

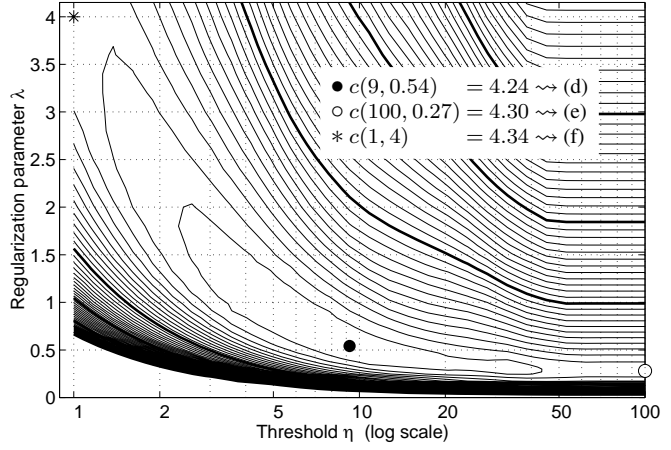
The interest of the trace of $c(\eta, \lambda)$, Figure 6.9b, is therefore primarily didactic. To obtain this, we proceeded by successive lines (λ constant) and from right to left (s decreasing), initializing each minimization with the result of the previous one. For $\eta = 100$, the calculation was initialized by linear inversion with the circulant approximation. When s is large (on the scale of differences of intensity between neighboring pixels), criterion (6.19) becomes quadratic, and thus easily minimizable. When s decreases, the criterion is less and less convex and would eventually become non-differentiable for $\eta = 0$. As a method for minimizing non-differentiable criteria, this procedure is known as a *regularization method* [GLO 81].

On the bivariate function $c(\eta, \lambda)$, the “right” settings form a relatively narrow valley that separates the two over-regularized solutions. The points of the valley are not all equivalent: the restored edges are all the sharper when threshold η is chosen to be small. The optimum threshold in the sense of the norm $L_{1/2}$ is $\eta = 9$. This setting corresponds quite well to a visual judgement: the restored image contains few artefacts except for the blocky effect on the derrick, hoist and the V-shaped mooring cable, oriented along the diagonals of the image. A penalization less sensitive to the orientation, such as (6.20) or the sum of (6.19) and (6.21), would doubtless correct this fault.

The best setting for λ for a very small threshold ($\eta = 1$) produces a slightly “over-segmented” restoration which seems less “natural”. This effect is not flagrant on the image tested as it shows up in shaded areas, which are few in this image (inflatable



(a) Blurred and noisy image y



(b) $c(\eta, \lambda) \propto \|\hat{x}^{\eta, \lambda} - x^*\|_{1/2}$



(c) Original image x^*



(d) $\hat{x}^{9, 0.54}$



(e) $\hat{x}^{100, 0.27} \simeq \hat{x}^{\infty, 0.27}$



(f) $\hat{x}^{1, 4}$

Figure 6.9. Synthetic image-deconvolution problem, approached by minimization of a convex criterion such as (6.30), where Φ is defined by (6.19) for $\nu = 1$ and ϕ is the Huber function (6.34). The error norm in the $L_{1/2}$ sense, image edges excluded, was adopted to assess the quality of the restoration; (b) represents the level lines of $c(\eta, \lambda) = \|\hat{x}^{\eta, \lambda} - x^*\|_{1/2}/(\text{number of pixels})$, spaced at 0.5. The minimum value, indicated by a black spot, is 4.24; it corresponds to (d). The quadratic case is obtained for $\eta \rightarrow \infty$; in this case, the minimum value, indicated by a white spot, is 4.30; it corresponds to (e). In contrast, we obtain the case L_1 for $\eta = 0$; for the smallest grid value, $\eta = 1$, the minimum value, indicated by a star, is 4.34; this corresponds to (f)

dinghy and lower part of the hull). We could envisage correcting this defect by penalizing differences of order two rather than order one. Nevertheless, the choice of a threshold η that is systematically small, or even zero, is not really desirable unless the aim is to segment the image rather than restore it.

6.7. Conclusion

Within the framework of penalized methods for the restoration of objects that are regular “almost everywhere” such as natural images, three types of approach have been presented.

Tikhonov’s original approach corresponds to *quadratic* penalization. This is both the oldest and the least costly, as it provides a solution by simple inversion of a linear system. Except in favorable situations (highly redundant data with little noise), it is not satisfactory for restoring piecewise homogeneous objects. In the examples illustrating this chapter, the method reaches, at best, an error norm that is not very competitive, of 18.16% (Figure 6.2).

Convex non-quadratic penalization gives robust estimators that demonstrate a much better capacity for preserving the boundaries between regions. The error norm attained becomes 16.57% (Figure 6.5). Moreover, these estimators can be calculated iteratively by fixed point methods, e.g. gradient methods. Other fixed point descent methods were proposed in the 1990s [BRE 96, CHA 94, CHA 97]. They work on half-quadratic augmented forms of the criterion to be minimized. They have the interesting structure of quasi-Newton, fixed-step algorithms [VOG 98], with the guarantee of convergence [CHA 97, IDI 01, NIK 05, ALL 06].

The *detection-estimation* approach is the equivalent half-quadratic formulation for a non-convex penalized approach (more accurately, using truncated quadratic penalization functions). This is the only one that has a real capacity for detecting boundaries between distinct regions. In the 1D example that illustrates this chapter, this capacity is expressed by a slight decrease in the L_1 error norm: 16.22% (Figure 6.3) instead of 16.57% for convex penalization. However, this decrease is obtained at an increased computing cost. Moreover, it becomes more difficult to set the hyperparameters empirically because of the instability of the solution (as it is a discontinuous function of the data, it is not generally continuous as a function of the hyperparameters).

It is interesting to end on a “historical” note. Since the pioneering work of the 1960s and 1970s, the evolution of the methods commonly studied and applied in signal and image restoration has not always been towards increasing complexity. Linear filtering methods (seen in more detail in Chapter 4), associated with the minimization of quadratic criteria, gave way, after the Geman brothers’ work [GEM 84], to much more costly methods based on discrete hidden variables and taking simulated

annealing as the optimization technique. Only in the 1990s did the much less expensive use of minimization of *convex non-quadratic* criteria become preponderant thanks to works such as [BOU 93], although already present in specific forms in the 1980s (e.g. in the maximum entropy framework [MOH 88]).

6.8. Bibliography

- [ALL 06] ALLAIN M., IDIER J., GOUSSARD Y., “On global and local convergence of half-quadratic algorithms”, *IEEE Trans. Image Processing*, vol. 15, num. 5, p. 1130-1142, May 2006.
- [AUB 97] AUBERT G., VESE L., “A variational method in image recovery”, *SIAM J. Num. Anal.*, vol. 34, num. 5, p. 1948-1979, Oct. 1997.
- [AUB 06] AUBERT G., KORNPBOST P., *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*, vol. 147 of *Applied Mathematical Sciences*, Springer Verlag, New York, NY, 2nd edition, 2006.
- [BER 95] BERTSEKAS D. P., *Nonlinear Programming*, Athena Scientific, Belmont, MA, 1995.
- [BLA 87] BLAKE A., ZISSERMAN A., *Visual Reconstruction*, The MIT Press, Cambridge, MA, 1987.
- [BOU 93] BOUMAN C. A., SAUER K. D., “A generalized Gaussian image model for edge-preserving MAP estimation”, *IEEE Trans. Image Processing*, vol. 2, num. 3, p. 296-310, July 1993.
- [BRÉ 94] BRÉMAUD P., *An Introduction to Probabilistic Modeling*, Undergraduate Texts in Mathematics, Springer Verlag, Berlin, Germany, 2nd edition, 1994.
- [BRE 96] BRETTE S., IDIER J., “Optimized single site update algorithms for image deblurring”, in *Proc. IEEE ICIP*, Lausanne, Switzerland, p. 65-68, 1996.
- [CHA 94] CHARBONNIER P., BLANC-FERAUD L., AUBERT G., BARLAUD M., “Two deterministic half-quadratic regularization algorithms for computed imaging”, in *Proc. IEEE ICIP*, vol. 2, Austin, TX, p. 168-172, Nov. 1994.
- [CHA 97] CHARBONNIER P., BLANC-FERAUD L., AUBERT G., BARLAUD M., “Deterministic edge-preserving regularization in computed imaging”, *IEEE Trans. Image Processing*, vol. 6, num. 2, p. 298-311, Feb. 1997.
- [DEL 98] DELANEY A. H., BRESLER Y., “Globally convergent edge-preserving regularized reconstruction: an application to limited-angle tomography”, *IEEE Trans. Image Processing*, vol. 7, num. 2, p. 204-221, Feb. 1998.
- [EVA 92] EVANS L. C., GARIEPY R. F., *Measure Theory and Fine Properties of Functions*, CRC Press, Berlin, Germany, 1992.
- [GEM 84] GEMAN S., GEMAN D., “Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images”, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, num. 6, p. 721-741, Nov. 1984.

- [GEM 87] GEMAN S., MCCLURE D., "Statistical methods for tomographic image reconstruction", in *Proc. 46th Session of the ICI, Bulletin of the ICI*, vol. 52, p. 5-21, 1987.
- [GEM 92] GEMAN D., REYNOLDS G., "Constrained restoration and the recovery of discontinuities", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, num. 3, p. 367-383, Mar. 1992.
- [GEM 95] GEMAN D., YANG C., "Nonlinear image recovery with half-quadratic regularization", *IEEE Trans. Image Processing*, vol. 4, num. 7, p. 932-946, July 1995.
- [GLO 81] GLOWINSKI R., LIONS J. L., TRÉMOLIÈRES R., *Numerical Analysis of Variational Inequalities*, Elsevier Science Publishers (North-Holland), Amsterdam, The Netherlands, 1981.
- [HEB 89] HEBERT T., LEAHY R., "A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors", *IEEE Trans. Medical Imaging*, vol. 8, num. 2, p. 194-202, June 1989.
- [HUB 81] HUBER P. J., *Robust Statistics*, John Wiley, New York, NY, 1981.
- [HUR 96] HURN M., JENNISON C., "An extension of Geman and Reynolds' approach to constrained restoration and the recovery of discontinuities", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-18, num. 6, p. 657-662, June 1996.
- [IDI 01] IDIER J., "Convex half-quadratic criteria and interacting auxiliary variables for image restoration", *IEEE Trans. Image Processing*, vol. 10, num. 7, p. 1001-1009, July 2001.
- [KAL 03] KALIFA J., MALLAT S., ROUGÉ B., "Deconvolution by thresholding in mirror wavelet bases", *IEEE Trans. Image Processing*, vol. 12, num. 4, p. 446-457, Apr. 2003.
- [KOE 84] KOENDERINK J. J., "The structure of images", *Biological Cybernetics*, vol. 50, p. 363-370, 1984.
- [KÜN 94] KÜNSCH H. R., "Robust priors for smoothing and image restoration", *Ann. Inst. Stat. Math.*, vol. 46, num. 1, p. 1-19, 1994.
- [LAB 08] LABAT C., IDIER J., "Convergence of conjugate gradient methods with a closed-form stepsize formula", *J. Optim. Theory Appl.*, vol. 136, num. 1, Jan. 2008.
- [LI 95] LI S. Z., HUANG Y. H., FU J. S., "Convex MRF potential functions", in *Proc. IEEE ICIP*, vol. 2, Washington DC, p. 296-299, 1995.
- [MOH 88] MOHAMMAD-DJAFARI A., DEMOMENT G., "Image restoration and reconstruction using entropy as a regularization functional", in ERIKSON G. J., RAY S. C. (Eds.), *Maximum Entropy and Bayesian Methods in Science and Engineering*, vol. 2, Dordrecht, The Netherlands, MaxEnt Workshops, Kluwer Academic Publishers, p. 341-355, 1988.
- [MUM 85] MUMFORD D., SHAH J., "Boundary detection by minimizing functionals", in *IEEE Conf. Comp. Vision Pattern Recogn.*, San Francisco, CA, p. 22-26, 1985.
- [NAS 81] NASHED M. Z., "Operator-theoretic and computational approaches to ill-posed problems with applications to antenna theory", *IEEE Trans. Ant. Propag.*, vol. 29, p. 220-231, 1981.
- [NIK 98] NIKOLOVA M., IDIER J., MOHAMMAD-DJAFARI A., "Inversion of large-support ill-posed linear operators using a piecewise Gaussian MRF", *IEEE Trans. Image Processing*, vol. 7, num. 4, p. 571-585, Apr. 1998.

- [NIK 99] NIKOLOVA M., "Markovian reconstruction using a GNC approach", *IEEE Trans. Image Processing*, vol. 8, num. 9, p. 1204-1220, Sep. 1999.
- [NIK 05] NIKOLOVA M., NG M., "Analysis of half-quadratic minimization methods for signal and image recovery", *SIAM J. Sci. Comput.*, vol. 27, p. 937-966, 2005.
- [NOR 90] NORDSTROM N., "Biased anisotropic diffusion: a unified regularization and diffusion approach to edge detection", *Image and Vision Computing*, vol. 8, num. 4, p. 318-327, 1990.
- [O'S 95] O'SULLIVAN J. A., "Roughness penalties on finite domains", *IEEE Trans. Image Processing*, vol. 4, num. 9, p. 1258-1268, Sep. 1995.
- [PER 90] PERONA P., MALIK J., "Scale-space and edge detection using anisotropic diffusion", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-12, p. 629-639, July 1990.
- [PRE 86] PRESS W. H., FLANNERY B. P., TEUKOLSKY S. A., VETTERLING W. T., *Numerical Recipes, the Art of Scientific Computing*, Cambridge University Press, Cambridge, MA, 1986.
- [REY 83] REY W. J., *Introduction to Robust and Quasi-robust Statistical Methods*, Springer Verlag, Berlin, 1983.
- [ROC 70] ROCKAFELLAR R. T., *Convex Analysis*, Princeton University Press, 1970.
- [SAU 93] SAUER K. D., BOUMAN C. A., "A local update strategy for iterative reconstruction from projections", *IEEE Trans. Signal Processing*, vol. 41, num. 2, p. 534-548, Feb. 1993.
- [STA 02] STARK J.-L., PANTIN E., MURTAGH F., "Deconvolution in astronomy: a review", *Publ. Astr. Soc. Pac.*, vol. 114, p. 1051-1069, 2002.
- [TEB 98] TEBOUL S., BLANC-FERAUD L., AUBERT G., BARLAUD M., "Variational approach for edge-preserving regularization using coupled PDE's", *IEEE Trans. Image Processing*, vol. 7, num. 3, p. 387-397, Mar. 1998, Special Issue on Partial Differential Equations and Geometry Driven Diffusion in Image Processing and Analysis.
- [TER 83] TERZOPOULOS D., "Multilevel computational process for visual surface reconstruction", *Comput. Vision Graphics Image Process.*, vol. 24, p. 52-96, 1983.
- [TER 86] TERZOPOULOS D., "Regularization of inverse visual problems involving discontinuities", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, num. 4, p. 413-424, July 1986.
- [TIK 77] TIKHONOV A., ARSENIN V., *Solutions of Ill-Posed Problems*, Winston, Washington, DC, 1977.
- [TIT 85] TITTERINGTON D. M., "General structure of regularization procedures in image reconstruction", *Astron. Astrophys.*, vol. 144, p. 381-387, 1985.
- [VOG 98] VOGEL R. V., OMAN M. E., "Fast, robust total variation-based reconstruction of noisy, blurred images", *IEEE Trans. Image Processing*, vol. 7, num. 6, p. 813-823, June 1998.
- [YAR 85] YARLAGADDA R., BEDNAR J. B., WATT T. L., "Fast algorithms for l_p deconvolution", *IEEE Trans. Acoust. Speech, Signal Processing*, vol. ASSP-33, num. 1, p. 174-182, Feb. 1985.

PART III

Advanced Problems and Tools

Chapter 7

Gibbs-Markov Image Models

7.1. Introduction

Chapter 6 introduced an important class of tools for signal and image restoration using solutions constructed as minimizers of penalized criteria. For a direct model $\mathbf{y} = H(x^*) + \text{noise}$, this approach first requires choices to be made for:

- a function measuring *goodness of fit* $\Psi(\mathbf{y} - H(x))$, which was taken as quadratic in Chapter 6;
- a regularizing function $\Phi(x)$;
- a value for hyperparameter λ that governs the compromise between regularization and goodness of fit.

With the proviso that it exists and is unique, solution \hat{x} is thus defined as the minimizer of a criterion:

$$\mathcal{J}(x) = \Psi(\mathbf{y} - H(x)) + \lambda\Phi(x). \quad (7.1)$$

Several “big” questions come to mind concerning \hat{x} :

- the choice of Φ was discussed in Chapter 6, essentially on the basis of qualitative arguments: a signal to be estimated may be “smooth”, another “regular almost everywhere”. Are there more “objective” arguments on which to base the choice of Φ (and Ψ)?

- since the solution \hat{x} depends on the hyperparameter λ and possibly additional parameters (e.g. the threshold parameter for L_2L_0 or L_2L_1 functions), is it possible to estimate these hyperparameters rather than choosing them empirically?;
- how much confidence can be given to \hat{x} ? Can a quantified uncertainty be associated with it?

Generally, one of the aims of statistical modeling is precisely to take vague or uncertain information into account. More specifically, we saw in Chapter 3 that Bayesian inference is a natural framework for approaching inverse problems. Many contributions in image restoration explicitly use this statistical framework [BES 86, BOU 93, DEM 89, GEM 84]. It is thus legitimate to give more details of how Bayesian inference contributes to signal and image restoration and to look more particularly at the extent to which it provides answers to questions that are difficult to deal with in the deterministic framework of the previous chapter.

7.2. Bayesian statistical framework

Proposing a probabilistic model for a signal means assuming that the signal is the realization of a random process in the hope of better structuring the available information by expressing it in the form of statistical characteristics of this process. Strictly speaking, and in spite of what often seems to be suggested, this does not mean that we assume the signal itself is random.

Under certain conditions, the *Bayesian* statistical framework is appropriate for interpreting (7.1). Adopting this framework means supposing that both the uncertainties (the *noise*) and signal x^* come from random quantities.

From now on, unless stated otherwise, we will take it that the observation, noise and signal are real vectors, respectively $\mathbf{y} \in \mathbb{R}^N$, $\mathbf{b} \in \mathbb{R}^N$ and $x^* = \mathbf{x}^* \in \mathbb{R}^M$:

$$\mathbf{y} = H(\mathbf{x}^*) + \mathbf{b}. \quad (7.2)$$

The cases in which the signal or observation belongs to a space of infinite dimension lend themselves only partially to a statistical formulation. It has long been known that Tikhonov's approach can be interpreted in Bayesian statistics, even without discretization [FRA 70]. In contrast, in the case of non-quadratic functionals Ψ or Φ , the minimization of criterion (7.1) has no systematic statistical interpretation. In substance, the difficulty stems from the fact that the probability of a random process indexed on a space of infinite dimension is characterized by an infinite collection of finite-dimensional distributions. This collection has no direct relation with (7.1) and does not allow a likelihood function to be defined naturally.

Let us also assume that functions Φ and Ψ verify the normalization condition:

$$\exists T_x > 0, \quad Z_x(T_x) = \int_{\mathbb{R}^M} e^{-\Phi(\mathbf{x})/T_x} d\mathbf{x} < +\infty, \quad (7.3)$$

$$\exists T_b > 0, \quad Z_b(T_b) = \int_{\mathbb{R}^N} e^{-\Psi(\mathbf{b})/T_b} d\mathbf{b} < +\infty \quad (7.4)$$

(T_x and T_b are so-called *temperature* parameters; their role will become clear later). We can thus define functions that have the characteristics of probability densities:

$$p_{\mathbf{X}}(\mathbf{x}) = p_{\mathbf{X}}(\mathbf{x}; T_x) = \frac{1}{Z_x(T_x)} e^{-\Phi(\mathbf{x})/T_x}, \quad (7.5)$$

$$p_{\mathbf{B}}(\mathbf{b}) = p_{\mathbf{B}}(\mathbf{b}; T_b) = \frac{1}{Z_b(T_b)} e^{-\Psi(\mathbf{b})/T_b}. \quad (7.6)$$

We then call $p_{\mathbf{X}}$ the prior law.

$$p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{x}) = p_{\mathbf{B}}(\mathbf{y} - H(\mathbf{x})) \quad (7.7)$$

designates the likelihood of the data, knowing \mathbf{x} , and

$$p_{\mathbf{X}|\mathbf{Y}}(\mathbf{x} | \mathbf{y}) = \frac{p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{x})p_{\mathbf{X}}(\mathbf{x})}{p_{\mathbf{Y}}(\mathbf{y})} = \frac{p_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y})}{\int_{\mathbb{R}^M} p_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y}) d\mathbf{x}}, \quad (7.8)$$

the posterior likelihood. The joint law for the pair (\mathbf{X}, \mathbf{Y}) can be written:

$$p_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y}) = p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{x})p_{\mathbf{X}}(\mathbf{x}). \quad (7.9)$$

The maximum *a posteriori* (MAP) is the Bayesian estimator most commonly used in imaging:

$$\begin{aligned} \hat{\mathbf{x}}^{\text{MAP}} &= \arg \max_{\mathbf{x}} p_{\mathbf{X}|\mathbf{Y}}(\mathbf{x} | \mathbf{y}) \\ &= \arg \min_{\mathbf{x}} \Psi(\mathbf{y} - H(\mathbf{x}))/T_b + \Phi(\mathbf{x})/T_x \end{aligned} \quad (7.10)$$

from equations (7.5)-(7.9). Consequently, the MAP is the same as the minimizer of the penalized criterion (7.1) for $\lambda = T_b/T_x$. This traditional identity has already been demonstrated and commented on in section 3.4.

7.3. Gibbs-Markov fields

Penalty functions appeared in the previous chapter in equations (6.7), (6.12), (6.17), (6.19), (6.21) and even (6.22), sharing a common characteristic: in the statistical interpretation of the previous section, they all correspond to a Gibbs-Markov field prior law. We can therefore conclude that “Markov” tools, and tools that can

be interpreted as such, are both very widely used in signal and image restoration and relatively simple to handle despite their reputation, since it is possible to “do Markov” without being aware of it.

Nevertheless, to be able to discuss more advanced questions such as those posed in the introduction to this chapter, it is necessary to make the definitions and fundamental properties of Gibbs-Markov fields explicit. Let us start with the notion of a Gibbs field on a finite set and the associated vocabulary.

7.3.1. Gibbs fields

7.3.1.1. Definition

Consider a finite set of variables $\mathbf{x} = \{x_1, \dots, x_S\}$, e.g. the pixels of an image, having values in a set \mathcal{E} . Depending on the case, we may be interested in $\mathcal{E} = \mathbb{R}$, $\mathcal{E} = \mathbb{R}^+$, $\mathcal{E} = \{1, \dots, 256\}$, etc. We define a set of *sites* $\mathcal{S} = \{1, \dots, S\}$ by bringing together the indices of the elements of \mathbf{x} . In the case of a rectangular image of size $I \times J$, the sites are the pairs (i, j) , artificially ordered from 1 to $S = IJ$.

Let \mathcal{C} be a set of *C cliques*, i.e., subsets of \mathcal{S} and the same number of real functions of \mathbf{x} , noted W_c , $c \in \mathcal{C}$, respecting the following constraint: although noted $W_c(\mathbf{x})$, the value taken by W_c , this function only depends on the variables $\mathbf{x}_c = \{x_s, s \in c\}$ and not on all of \mathbf{x} . The set of functions W_c is called a *Gibbs potential*.

If \mathcal{E} is \mathbb{R} or a subset of \mathbb{R} of non-zero measure (typically, \mathbb{R}_+ or an interval of \mathbb{R}), situations that can be globally designated as *the continuous case*¹, we define a *Gibbs field* \mathbf{X} having a value in \mathcal{E} , of energy:

$$\Phi = \sum_{c \in \mathcal{C}} W_c, \quad (7.11)$$

at temperature T , as the random process with the following pdf:

$$p_{\mathbf{X}}(\mathbf{x}) = \frac{1}{Z(T)} e^{-\Phi(\mathbf{x})/T}, \quad (7.12)$$

where the normalization constant:

$$Z(T) = \int_{\mathcal{E}^S} e^{-\Phi(\mathbf{x})/T} d\mathbf{x},$$

called the *partition function* here, is assumed to be finite.

1. In this chapter, the opposition between *continuous case* and *discrete case* concerns the pixel values, whereas Chapter 6 distinguished between *continuous* functions and *discrete* functions according to their support, i.e., their domain of definition instead of their range.

In addition, a Gibbs field is said to be *homogeneous* if the definition of functions W_c respects invariance under translation (which implicitly assumes that \mathbf{x} is spatially organized): for any pair of cliques c, c' that are superimposable by translation:

$$W_c(\mathbf{x}) = W_{c'}(\mathbf{x}'), \quad \forall \mathbf{x}, \mathbf{x}' \text{ such that } \mathbf{x}_c = \mathbf{x}'_{c'}.$$

The definition of a Gibbs field is comparable to equations (7.5) and (7.6): the process inducing a probability law from an “energy” is identical in the three cases. On the other hand, a Gibbs field differs from a generic random process by the “local” structure of its energy (7.11), a “macroscopic” quantity defined as the sum of “microscopic” contributions.

In the case where the variables $\mathbf{x} = \{x_1, \dots, x_S\}$ take discrete values (*the discrete case*), e.g. pixels with quantified grey levels, we define the probability of a Gibbs field \mathbf{X} with discrete values in a similar way:

$$\Pr(\mathbf{X} = \mathbf{x}) = \frac{1}{Z(T)} e^{-\Phi(\mathbf{x})/T}, \quad (7.13)$$

where Φ is again defined by (7.11) and the partition function is now written:

$$Z(T) = \sum_{\mathbf{x} \in \mathcal{E}^S} e^{-\Phi(\mathbf{x})/T}.$$

By a variety of construction procedures, ranging from the simplest (all-pole filtering of white, multivariate noise; unilateral fields) to the most sophisticated (study of (7.13) at the limit $S \rightarrow \infty$), we can also define Gibbs fields on networks of infinite size [GUY 95], or with continuous indices, e.g. Gaussian fields [MOU 97] or spatial point processes [BAD 89]. As the use of these extended models remains infrequent in image restoration, the definition of Gibbs fields on finite networks will be quite sufficient for our purposes.

7.3.1.2. Trivial examples

The special case of a Gibbs field admitting only singleton cliques $c = \{s\}$:

$$p_{\mathbf{X}}(\mathbf{x}) = \frac{1}{Z(T)} \exp \left\{ - \sum_{s \in S} W_s(\mathbf{x})/T \right\} = \frac{1}{Z(T)} \prod_{s \in S} e^{-W_s(\mathbf{x})/T} = \prod_{s \in S} p_{X_s}(x_s),$$

corresponds to a hypothesis that the components of the field are independent. At the other extreme, when \mathcal{C} contains set S itself, all the random processes on S become Gibbs fields. However, “Gibbsian” analysis is obviously of no interest in these two trivial cases.

7.3.1.3. Pairwise interactions, improper laws

As a Gibbs energy, a penalty function such as:

$$\sum_{i=2}^I \sum_{j=1}^J (x_{i,j} - x_{i-1,j})^2 + \sum_{i=1}^I \sum_{j=2}^J (x_{i,j} - x_{i,j-1})^2, \quad (7.14)$$

can be decomposed over horizontal and vertical cliques composed of pairs of adjacent sites, $\{(i-1, j), (i, j)\}$ and $\{(i, j-1), (i, j)\}$.

Similarly, many other penalty functions used in image restoration can be written $\sum_{r \sim s} \phi_{rs}(x_r, x_s)$, where $r \sim s$ describes a set of pairs of sites that are “spatial neighbors”. These functions are Gibbs energies for which the interactions are limited to pairs.

Among these functions, the most used can be written $\Phi(\mathbf{x}) = \sum_{r \sim s} \phi(x_r - x_s)$, following the example of (7.14). As the energy of a (homogeneous) Gibbs field, this expression poses a problem if $\mathcal{E} = \mathbb{R}$, since it is easy to show that $\forall T$, $Z(T) = +\infty$ by a linear invertible change of variables $t_1 = x_1$, $t_2 = x_2 - x_1$, $t_3 = x_3 - x_1$, \dots , $t_S = x_S - x_1$. The prior law $\exp\{-\Phi(\mathbf{x})/T\}$, which is not normalizable, is said to be improper.

When the prior law is improper, the posterior likelihood $p_{\mathbf{X}|\mathbf{Y}}$ can still be the pdf of a proper law since, from (7.8) it can be written:

$$p_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}) = e^{-\Psi(\mathbf{y}-H(\mathbf{x}))/T_b - \Phi(\mathbf{x})/T_x} / Z_{x|z}(T_x, T_b) \quad (7.15)$$

where $Z_{x|z}(T_x, T_b) = \int_{\mathbb{R}^M} e^{-\Psi(\mathbf{y}-H(\mathbf{x}))/T_b - \Phi(\mathbf{x})/T_x} d\mathbf{x} < +\infty$ is a less restrictive condition than (7.3)-(7.4). As shown in [SAQ 98, p. 1031, footnote], an improper law can be “managed” as the limit of a family of pdfs, e.g. by introducing the augmented energy:

$$\Phi_\varepsilon(\mathbf{x}) = \sum_{r \sim s} \phi_{rs}(x_r - x_s) + \varepsilon \|\mathbf{x}\|^2 \xrightarrow{\varepsilon \rightarrow 0^+} \Phi(\mathbf{x}).$$

7.3.1.4. Markov chains

In this subsection and the rest of the chapter, we will use the shortened notation $p(a|b)$ for the conditional density $p_{A|B}(a|b)$, unless it is ambiguous. In the same way, we will use $p(a)$ for $p_A(a)$.

In the continuous case, a *Markov chain* is defined as a collection of random variables $\mathbf{X} = (X_1, \dots, X_S)$ that admit a joint probability density $p_{\mathbf{X}}$ such that:

$$\forall s \in \{2, \dots, S\}, (x_1, \dots, x_s) \in \mathcal{E}^s, p(x_s | x_1, \dots, x_{s-1}) = p(x_s | x_{s-1}).$$

For example, an autoregressive signal of order one is a Markov chain. In the case when s is a time index, we can say that the random evolution with time of such a process depends only on the most recent past.

Markov chains are Gibbs fields. In fact, we have:

$$p(\mathbf{x}) = p(x_1) \prod_{s>1} p(x_s | x_1, \dots, x_{s-1}) = p(x_1) \prod_{s>1} p(x_s | x_{s-1}) \quad (7.16)$$

from the sequential Bayes' rule and the definition of a Markov chain; or again:

$$p(\mathbf{x}) = \exp \left\{ \log p(x_1) + \sum_{s>1} \log p(x_s | x_{s-1}) \right\}. \quad (7.17)$$

Inversely, it can be shown by descending recurrence that a Gibbs field, the cliques of which are the pairs $\{s-1, s\}$, $s = 2, \dots, S$, is a Markov chain [QIA 90]. In a much more general context, there is again a “Gibbs-Markov” equivalence, which we will look at in the next section.

7.3.1.5. Minimum cliques, non-uniqueness of potential

Is the singleton $\{1\}$ a clique for Gibbs field (7.17)? Yes, for a potential composed of the S elements $-\log p(x_1)$, $-\log p(x_2 | x_1)$... No, if the first two terms are replaced by their sum $-\log p(x_1) - \log p(x_2 | x_1)$. More generally, we can consider that any non-empty subset of a clique is a clique; or, looking at things the opposite way, we can restrict ourselves to *maximum* cliques only, i.e., those that are not subsets of any other set.

For a given Gibbs energy, both conventions “maximum cliques and subsets” and “maximum cliques only” make the list of cliques unique. However, there remain several potentials to describe the same energy (except in the trivial cases of disjoint cliques, in the convention “maximum cliques only”), since the sum $W_c + W_d$ of the contributions associated with two cliques c and d having a non-empty intersection can also be written $(W_c + W'_{c \cap d}) + (W_d - W'_{c \cap d})$ for any function $W'_{c \cap d}$ defined on $c \cap d$.

7.3.2. Gibbs-Markov equivalence

7.3.2.1. Neighborhood relationship

Let \mathbf{X} be a random field defined on $\mathcal{S} = \{1, \dots, S\}$, with values in \mathcal{E} , admitting a density $p_{\mathbf{X}}$ strictly positive on $\mathcal{E}^{\mathcal{S}}$. Let \mathcal{V}_s , $s \in \mathcal{S}$ be subsets of \mathcal{S} , \mathcal{V}_s being called the *set of neighbors of s* . We will only concern ourselves with sets \mathcal{V}_s that induce a neighborhood relationship that is antireflexive (s is not its own neighbor: $s \notin \mathcal{V}_s$) and symmetric (if s is a neighbor of r , r is a neighbor of s : $s \in \mathcal{V}_r \iff r \in \mathcal{V}_s$). In what

follows, we will note the neighborhood relationship as \sim in the sense where $r \sim s$ is equivalent to $r \in \mathcal{V}_s$.

This definition of the relation \sim is abstract. In practice when the idea of neighborhood pre-exists in \mathcal{S} , it is frequent (but not obligatory) to make the neighborhood relationship coincide with it. For example, if the pre-existing idea of neighborhood is induced by a distance Δ , we can define $\mathcal{V}_s = \{r \neq s, \Delta(r, s) \leq D\}$, by choosing the neighborhood “range” $D > 0$. If $\mathcal{S} \subset \mathbb{Z}^p$, if Δ is the usual Euclidean distance and if D^2 is an integer, we thus define the neighborhood relation of order D^2 .

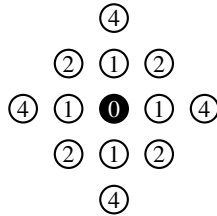


Figure 7.1. A site of \mathbb{Z}^2 admits four neighbors of order one, four more neighbors of order two, no neighbor of order three, four of order four. More generally, a site of \mathbb{Z}^p has $2^1 C_p^1 = 2p$ neighbors of order one, $2^2 C_p^2 = 2p(p-1)$ of order two, $2^3 C_p^3 = 4p(p-1)(p-2)/3$ of order three; but the rest of the sequence shows that the general term is not $2^q C_p^q$

7.3.2.2. Definition of a Markov field

\mathbf{X} is said to be a Markov field for the neighborhood relationship \sim if:

$$\forall s \in \mathcal{S}, \mathbf{x} \in \mathcal{E}^{\mathcal{S}}, p(x_s | \mathbf{x}_{\mathcal{S} \setminus s}) = p(x_s | \mathbf{x}_{\mathcal{V}_s}). \quad (7.18)$$

In other words, the probability of X_s , conditionally on all the other variables of \mathbf{X} , does not depend on S variables, but only on the values taken by X_s and its neighbors. Once again, we find the possibility of specifying the model of a signal or image through its local properties.

This definition concerns the continuous case. In the discrete case it is transposed in the form: $\forall s \in \mathcal{S}, \mathbf{x} \in \mathcal{E}^{\mathcal{S}}$,

$$\Pr(X_s = x_s | \mathbf{X}_{\mathcal{S} \setminus s} = \mathbf{x}_{\mathcal{S} \setminus s}) = \Pr(X_s = x_s | \mathbf{X}_{\mathcal{V}_s} = \mathbf{x}_{\mathcal{V}_s}). \quad (7.19)$$

Specifying the *local characteristics* of \mathbf{X} , i.e., the S conditional laws “ X_s knowing $\mathbf{X}_{\mathcal{S} \setminus s}$ ”, is sufficient to define the law of \mathbf{X} . In the discrete case, this assertion is proved

by the identity:

$$\frac{\Pr(\mathbf{x})}{\Pr(\mathbf{w})} = \prod_{s=1}^S \frac{\Pr(x_s | x_1, \dots, x_{s-1}, w_{s+1}, \dots, w_S)}{\Pr(w_s | x_1, \dots, x_{s-1}, w_{s+1}, \dots, w_S)},$$

valid for any pair $\mathbf{x}, \mathbf{w} \in \mathcal{E}^S$ [BES 74] and transposable to the continuous case.

It could be thought that this identity allows Markov fields to be constructed from freely specified conditional laws (7.19). This is not so. Except for “lucky accidents”, arbitrarily specified conditional laws do not correspond to any joint law. In this sense, (7.18) and (7.19) are not constructive definitions but, rather, characteristic properties. To specify a Markov field, it is necessary to proceed indirectly, first specifying a joint law, then checking that it satisfies the Markov condition (7.18) or (7.19) as the case may be. This condition is automatically satisfied by Gibbs fields, as the next subsection shows.

7.3.2.3. A Gibbs field is a Markov field

We consider the continuous case here. The discrete case lends itself to a development that is equivalent throughout. Let us examine the structure of the conditional laws $p(x_s | \mathbf{x}_{S \setminus s})$ when \mathbf{X} is a Gibbs field:

$$\begin{aligned} p(x_s | \mathbf{x}_{S \setminus s}) &= \frac{p(\mathbf{x})}{p(\mathbf{x}_{S \setminus s})} = \frac{p(\mathbf{x})}{\int_{\mathcal{E}} p(\mathbf{x}) dx_s} \\ &= \frac{\exp \left\{ - \sum_{c \in \mathcal{C}} W_c(\mathbf{x})/T \right\} / Z(T)}{\int_{\mathcal{E}} \exp \left\{ - \sum_{c \in \mathcal{C}} W_c(\mathbf{x})/T \right\} / Z(T) dx_s}. \end{aligned}$$

The equations of the first line are true whether \mathbf{X} is a Gibbs field or not. The Gibbsian structure appears on the second line and entails simplifications for all the multiplying terms that do not depend on x_s , i.e., $1/Z(T)$ and any $\exp \{-W_c(\mathbf{x})/T\}$ term indexed by a clique c that does not contain s . After simplification, we obtain:

$$p(x_s | \mathbf{x}_{S \setminus s}) = \frac{\exp \left\{ - \sum_{c \ni s} W_c(\mathbf{x})/T \right\}}{\int_{\mathcal{E}} \exp \left\{ - \sum_{c \ni s} W_c(\mathbf{x})/T \right\} dx_s},$$

where the sums are restricted to the cliques that contain s . It is important to note that this expression does not depend on all the variables x_r , $r \in \mathcal{S}$: those that do not belong to any clique containing s are not included. It results that property (7.18) is true: a Gibbs field is a Markov field for a neighborhood structure described by $\mathcal{V}_s = \{r, \exists c / r \in c \ni s\}$.

7.3.2.4. Hammersley-Clifford theorem

Are there Markov fields that are not Gibbs fields? This question was answered by an equivalence theorem stated by Hammersley and Clifford in an unpublished report

in 1968. This theorem says that, in the discrete case, Markov fields verifying the positivity condition $\forall \mathbf{x}, \Pr(\mathbf{X} = \mathbf{x}) > 0$ are Gibbs fields, whose cliques are singletons or sets of sites that are two-by-two neighbors. The demonstration, even when modernized, is too technical to be presented here. It can be found in [WIN 03, section 3.3] and in [BRÉ 99, section 7.2], for example. This result is important from a theoretical point of view since it indicates that Gibbs and Markov fields are essentially the same mathematical objects, called Gibbs-Markov random fields (GMRFs) in what follows.

7.3.3. Posterior law of a GMRF

Analysis of the posterior law of a GMRF \mathbf{X} brings us a result that is useful in image restoration: under not very restrictive conditions, this law is itself that of a GMRF, for which it is interesting to study the neighborhood structure. For a GMRF \mathbf{X} of energy (7.11) and in the observation conditions of section 7.2, it has already been pointed out, in equation (7.15), that the energy of the posterior law can be written

$$\Psi(\mathbf{y} - H(\mathbf{x}))/T_b + \Phi(\mathbf{x})/T_x. \quad (7.20)$$

This corresponds to the Gibbs potential of the prior law augmented by terms expressing the co-log-likelihood $-\log p_{\mathbf{Y}|\mathbf{X}} = -\log p_{\mathbf{B}}(\mathbf{y} - H(\mathbf{x}))$. Thus, in general, we can say that the posterior law is that of a Markov field for which the neighborhood graph comes from the decomposition of the log-likelihood over cliques. A finer analysis is possible case by case.

Consider, for example, the common situation of a Gaussian linear observation $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{b}$, where \mathbf{b} is the realization of Gaussian noise independent of \mathbf{x} , of mean $\mathbf{m}_{\mathbf{B}}$ and invertible covariance $\mathbf{R}_{\mathbf{B}}$. The conditional co-log-likelihood of the observations is thus quadratic. As a function of \mathbf{x} , it can be decomposed in the form:

$$\begin{aligned} -\log p_{\mathbf{B}}(\mathbf{y} - \mathbf{H}\mathbf{x}) &= (\mathbf{y} - \mathbf{H}\mathbf{x})^T \mathbf{R}_{\mathbf{B}}^{-1} (\mathbf{y} - \mathbf{H}\mathbf{x})/2 + \text{const.} \\ &= -\mathbf{y}^T \mathbf{R}_{\mathbf{B}}^{-1} \mathbf{H}\mathbf{x} + \mathbf{x}^T \mathbf{H}^T \mathbf{R}_{\mathbf{B}}^{-1} \mathbf{H}\mathbf{x}/2 + \text{const.} \\ &= -\sum_s \alpha_s x_s + \sum_{r,s} \beta_{rs} x_r x_s /2 + \text{const.} \end{aligned} \quad (7.21)$$

where α_s and β_{rs} are the components of vector $\mathbf{H}\mathbf{R}_{\mathbf{B}}^{-1}\mathbf{y}$ and matrix $\mathbf{H}^T \mathbf{R}_{\mathbf{B}}^{-1} \mathbf{H}$ respectively. The only cliques involved in expression (7.21) are singletons and pairs. All the sites become neighbors if $\mathbf{H}^T \mathbf{R}_{\mathbf{B}}^{-1} \mathbf{H}$ is a full matrix. The more sparse it is, the fewer new connections appear in the posterior-law neighborhood graph. In the extreme, if \mathbf{H} and $\mathbf{R}_{\mathbf{B}}$ are diagonal matrices (which is the case of denoising or interpolation), the prior and posterior neighborhood structures are identical.

7.3.4. Gibbs-Markov models for images

In this section, we will draw up a schematic classification of GMRF published in the image processing literature, which takes its inspiration from [BES 86]. Note that most of these models are mathematical expressions of the qualitative notion of “regularity almost everywhere”, already frequently mentioned in the previous chapter.

7.3.4.1. Pixels with discrete values and label fields for classification

The oldest models use discrete values, no doubt for two reasons: first, for a long time, the capacity of computers limited image coding to one byte per pixel or even less; second, estimating an image on a limited number of hues is a way of grouping the pixels together in *classes*, taking their neighborhood into account, i.e., approaching a *contextual classification* problem.

Ising’s model is the simplest and the oldest of non-trivial, discrete-valued GMRFs [PIC 87]. It is a binary model with interactions in pairs having a neighborhood of order one. Defined on \mathbb{Z}^2 , it was proposed by Ising in 1925 to explain the behavior of ferromagnetic materials, in particular *phase transition*. This occurs below a certain temperature, called the critical temperature, by non-zero correlation among infinitely distant pixels. In imaging, the models are defined on networks of finite size so the role of phase transitions has been little studied [MOR 96]. The energy of an Ising field is written:

$$\Phi(\mathbf{x}) = \alpha \sum_s x_s + \beta \sum_{r \sim s} x_r x_s, \quad (7.22)$$

with $\mathcal{E} = \{-1, 1\}$. If $\alpha \neq 0$, the first term favors states of opposite sign to α . If $\beta > 0$, the low-energy (so most probable) configurations are composed of dissimilar neighbors and the behavior is *repulsive*. In contrast, $\beta < 0$ produces *attractive* behavior between neighbors. This is the desired behavior for classification since it favors homogeneous regions.

Ising’s model can easily be generalized to K colors, e.g. in the form of an energy written:

$$\Phi(\mathbf{x}) = \sum_{k=1}^K \alpha_k n_k + \sum_{k=1}^K \sum_{l=k}^K \beta_{kl} v_{kl},$$

n_k being the number of pixels of color k and $v_{kl} = v_{lk}$ the number of pairs of neighbors of colors respectively k and l . We find (7.22) for $\alpha_1 = -\alpha_{-1} = \alpha$ and $\beta_{1,1} = \beta_{-1,-1} = -\beta_{-1,1} = \beta$.

In the case of *ordered* colors, such as grey levels, the β_{kl} can be chosen as increasing functions of $|k - l|$ to encourage gradual changes of shade. Minimizing a criterion such as $\|\mathbf{y} - \mathbf{x}\|^2 + \lambda \Phi(\mathbf{x})$, where \mathbf{y} is an image observed over a number of shades at least equal to K , then gives a version of \mathbf{y} reduced to K levels, favoring homogeneous regions.

In the case of non-ordered colors, the model can be simplified by considering, for example, $\beta_{kl} = \beta < 0$ if $k = l$, 0 otherwise. We thus obtain Potts' model. The case of "non-ordered colors" may seem marginal but it brings us to the concept of the *label field*, which is fundamental in the modern approaches to image classification. From a very general point of view, we can consider that contextual classification consists of distributing the pixels into classes that are homogeneous in a sense to be defined, and marked by labels that are not colors or grey levels. Choosing a spatial model for the labels, such as Potts' model (or an improved version such as the *chien-model* [DES 95]), encourages the formation of aggregates of pixels assigned to the same class (this is known as *contextual classification*). A second level of K models (one model per class), enables the pixel law to be specified *region by region*, and thus the homogeneity for each class to be defined. In this way, it is possible to distinguish regions by their mean intensity, their variance or their texture. In the case of textures, the second-level models can be specialized GMRF (such as the *autobinomial* model [BES 74]; see also [CRO 83]) but also marked point processes, assemblies of primitives, fractal models, etc. [WIN 03].

Image modeling for classification can also have the aspect of a sophisticated stack of models, which contrasts with the relatively rustic character of the models the most used in image restoration. There is an explanation for this difference: whereas classification seeks to give a simplified representation of a complex but known object (i.e., an image), the aim of restoration is to give the most faithful possible image starting from imperfect knowledge of it. The poorer the data, the less useful it is to model details, such as texture, which it will, in any case, be impossible to restore.

In what follows, we will take the continuous case ($\mathcal{E} = \mathbb{R}$) and give greater place to GMRFs which are more suited to image restoration than to classification. In essence, the Gibbs energies of these models are the penalty functions presented in the previous chapter: first quadratic energies, then half-quadratic energies associated with detection-estimation, and finally non-quadratic energies, convex or non-convex.

7.3.4.2. Gaussian GMRF

Penalty (7.14) is the energy of a Gaussian GMRF (improper). More generally, a quadratic energy defined as positive (\mathbf{M} symmetric positive matrix):

$$\Phi(\mathbf{x}) = \mathbf{x}^T \mathbf{M} \mathbf{x} - 2\mathbf{m}^T \mathbf{x} + \mu = \sum_{r=1}^S \sum_{s=1}^S M_{rs} x_r x_s - \sum_{s=1}^S m_r x_s + \mu \quad (7.23)$$

corresponds, at any temperature T , to a Gaussian GMRF. The improper case is reached when \mathbf{M} is positive in the broad sense only.

By identifying $\Phi(\mathbf{x})/T$ with $(\mathbf{x} - \mathbf{m}_x)^T \mathbf{R}_x^{-1} (\mathbf{x} - \mathbf{m}_x)/2$ to within an additive constant, the mean and covariance of \mathbf{X} can be found:

$$\mathbf{m}_x = \mathbf{M}^{-1} \mathbf{m}, \quad \mathbf{R}_x = T\mathbf{M}^{-1}/2.$$

Moreover, it is clear from (7.23) that the cliques of a Gaussian GMRF are singletons and pairs and that the neighborhood relationship can be “read” directly in the matrix \mathbf{M} (or in \mathbf{R}_x^{-1}): r is a neighbor of s if $s \neq r$ and $M_{rs} \neq 0$. Gaussian GMRFs possess many other structural properties, including the explicit character of the partition function on \mathbb{R}^S :

$$\int_{\mathbb{R}^S} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \mathbf{m}_x)^T \mathbf{R}_x^{-1} (\mathbf{x} - \mathbf{m}_x) \right\} d\mathbf{x} = (2\pi)^{S/2} |\mathbf{R}_x|^{1/2}. \quad (7.24)$$

7.3.4.3. Edge variables, composite GMRF

The detection-estimation approach of Chapter 6 (section 6.3) favors half-quadratic penalty functions, the simplest of which can be written:

$$\Phi(\mathbf{x}, \ell) = \sum_{r \sim s} (1 - \ell_{rs})(x_r - x_s)^2 + \alpha \sum_{r \sim s} \ell_{rs}, \quad (7.25)$$

where the neighborhood is of order one; more rarely, it is of order two as in (6.13). Each variable ℓ_{rs} has the value 0 or 1 and models the presence ($\ell_{rs} = 1$) or absence ($\ell_{rs} = 0$) of a discontinuity between pixels r and s , and α is the price to be paid per discontinuity. Interpreted as a Gibbs energy, (7.25) defines a *composite* (\mathbf{x}, ℓ) GMRF, distributed over *pixel sites* and *edge sites*, and having a value in $\mathcal{E}^S \times \{0, 1\}^C$, where C is the number of pairs of neighboring pixel sites.

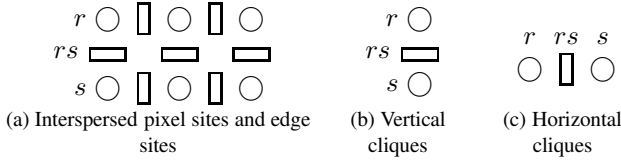


Figure 7.2. Spatial structure of composite GMRF of energy (7.25)

Composite GMRFs integrating continuous and discrete variables are defined neither by density function nor by discrete distribution but in the form of a mixed product:

$$p(\mathbf{x} | \ell) \Pr(\mathbf{L} = \ell) = \frac{1}{Z(T)} e^{-\Phi(\mathbf{x}, \ell)/T}, \quad Z(T) = \sum_{\ell} \int_{\mathbf{x}} e^{-\Phi(\mathbf{x}, \ell)/T} d\mathbf{x}. \quad (7.26)$$

In the case of (7.25), $Z(T) = +\infty$ if $\mathcal{E} = \mathbb{R}$. The law thus defined is improper.

It could be thought that the two terms of (7.25) correspond to $p(\mathbf{x} | \ell)$ and $\Pr(\mathbf{L} = \ell)$ respectively. However, to reason correctly, $\Pr(\mathbf{L} = \ell)$ of (7.26) should be deduced

by marginalization. Let us do this starting from the energy $\Phi_\varepsilon(\mathbf{x}, \ell) = \Phi(\mathbf{x}, \ell) + \varepsilon \|\mathbf{x}\|^2$, making the law of (\mathbf{X}, \mathbf{L}) proper if $\varepsilon > 0$:

$$\begin{aligned} \Pr(\mathbf{L} = \ell) &= \frac{1}{Z(T)} \int_{\mathcal{E}^S} e^{-\Phi_\varepsilon(\mathbf{x}, \ell)/T} d\mathbf{x} \\ &= \frac{1}{Z(T)} \exp\left\{-\frac{\alpha}{T} \sum_{r \sim s} \ell_{rs}\right\} \int_{\mathcal{E}^S} \exp\left\{-\frac{\varepsilon}{T} \|\mathbf{x}\|^2 - \frac{1}{T} \sum_{r \sim s} (1 - \ell_{rs})(x_r - x_s)^2\right\} d\mathbf{x}. \end{aligned}$$

The last integral is the partition function of the Gaussian GMRF $\mathbf{X} \mid \mathbf{L}$: from (7.24), $Z_{\mathbf{x}|\ell}(T) = (\pi T)^{S/2} |\mathbf{M}|^{-1/2}$ if $\mathcal{E} = \mathbb{R}$, where \mathbf{M} is defined by:

$$M_{ss} = \varepsilon + \sum_{r, r \sim s} (1 - \ell_{rs}); \quad M_{rs} = -2 \text{ if } r \sim s \text{ and } \ell_{rs} = 0, \quad 0 \text{ otherwise.}$$

Finally, the law of \mathbf{L} is not iid, since $Z_{\mathbf{x}|\ell}(T)$ depends on ℓ :

$$\Pr(\mathbf{L} = \ell) = \frac{Z_{\mathbf{x}|\ell}(T)}{Z(T)} \exp\left\{-\frac{\alpha}{T} \sum_{r \sim s} \ell_{rs}\right\} \propto \exp\left\{-\frac{\alpha}{T} \sum_{r \sim s} \ell_{rs}\right\}.$$

7.3.4.4. Interactive edge variables

The role of edge variables is to surround homogeneous regions, thus *segmenting* the image. Classification and segmentation are not identical operations as the second works without pre-specified classes.

The minimization of a criterion such as (7.25) does not guarantee that the outlines will be closed. To give more importance to closure, [GEM 84] introduced *interactive* edge variables. This does not signify that these variables become correlated; we noted above that they already are in the model induced by energy (7.25). The new feature is a set \mathcal{C}_L of cliques composed of four edge sites that are added to penalize interrupted outlines:

$$\Phi(\mathbf{x}, \ell) = \sum_{r \sim s} (1 - \ell_{rs})(x_r - x_s)^2 + \alpha \sum_{c \in \mathcal{C}_L} G_c(\ell). \quad (7.27)$$

The values of $G_c(\ell)$ proposed by [GEM 84] are indicated in Figure 7.3. The resulting criterion (7.27) is not exactly the one proposed in [GEM 84], since the model put forward in this article was based on pixels with discrete values (with *non-ordered colors*, see p. 182). To guarantee systematic closing of the outlines, G_c would have to be infinite for type ② configurations. However, this condition would not respect the positivity condition of the Hammersley-Clifford theorem and would make the posterior energy minimization problem (even more) tricky.

The interactivity of the edge variables not only encourages outline closure but also allows their appearance to be modified. Horizontal or vertical straight edges are

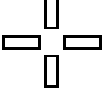
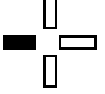
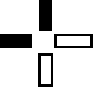
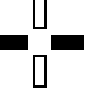
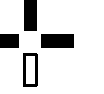
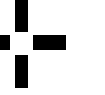
$c \in \mathcal{C}_L$						
Type	①	②	③	④	⑤	⑥
$G_c(\ell)$	0	2.7	1.8	0.9	1.8	2.7

Figure 7.3. Clique c composed of four edge sites and values of $G_c(\ell)$ for various configurations of ℓ . The “activated” variables $\ell_{rs} = 1$ are represented in black. $G_c(\ell)$ is invariant by symmetry and rotation

favorable: type ④ configurations are less costly than type ③. Other interactive edge models have been proposed in the same vein, e.g. to refine the outlines [MAR 87] or reduce directional effects [SIL 90].

7.3.4.5. Non-Gaussian GMRFs

Among the non-Gaussian GMRFs, those most used in image restoration are of order one and their energy is a function of interpixel differences:

$$\Phi(\mathbf{x}) = \sum_{r \sim s} \phi(x_r - x_s), \quad (7.28)$$

where ϕ is typically an L_2L_1 , L_2L_0 function [KÜN 94], or a related function such as $\phi(u) = |u|^p$ [BOU 93]. The use of these models as priors in a procedure for estimation in the MAP sense is the same as adopting the non-quadratic penalized approach of section 6.4 (which can usefully be looked at again).

7.4. Statistical tools, stochastic sampling

7.4.1. Statistical tools

Section 7.3 sets the penalty functions of the previous chapter in a Bayesian probabilistic framework. Some of the “big” questions posed in the introduction now find a natural answer, at least from the formal standpoint. In particular, it becomes possible to quantify the uncertainty via the density of the posterior law $p_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y})$ (see Chapter 3), and to extract simpler indices from this law, such as the posterior mean and covariance:

$$\hat{\mathbf{x}}^{\text{PM}} = \mathbb{E}(\mathbf{X}|\mathbf{y}) = \int_{\mathcal{E}^S} \mathbf{x} p_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}) d\mathbf{x} \quad (7.29)$$

$$\begin{aligned} \mathbf{R}^{\text{CP}} &= \text{Cov}(\mathbf{X}|\mathbf{y}) = \mathbb{E}((\mathbf{X} - \hat{\mathbf{x}}^{\text{PM}})(\mathbf{X} - \hat{\mathbf{x}}^{\text{PM}})^T | \mathbf{y}) \\ &= \int_{\mathcal{E}^S} (\mathbf{x} - \hat{\mathbf{x}}^{\text{PM}})(\mathbf{x} - \hat{\mathbf{x}}^{\text{PM}})^T p_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}) d\mathbf{x}. \end{aligned} \quad (7.30)$$

The posterior mean $\hat{\mathbf{x}}^{\text{PM}}$ has S components. It is a Bayesian estimator than could be a rival for the MAP. Its use in image restoration was advocated by Marroquin [MAR 87] but has failed to become widespread.

The posterior covariance $\mathbf{R}^{\text{CP}} = (R_{rs}^{\text{CP}})$ is a matrix of size $S \times S$. It can be used to generate a *confidence region* for the estimated intensities. For example, from the standard deviations $\sigma_s = (\text{var}(X_s | \mathbf{y}))^{1/2} = \sqrt{R_{ss}^{\text{CP}}}$, we can define a Cartesian product with intervals $\mathbf{I}(\mu) = \bigotimes_s \mathbf{I}_s(\mu)$ and extremities:

$$b_s^{\pm} = \hat{x}_s^{\text{PM}} \pm \mu \sigma_s, \quad \mu > 0.$$

This approach neglects the correlations in the posterior law but is attractive as it is relatively simple and visual: the b_s^{\pm} are error bars on each pixel, estimated by \hat{x}_s^{PM} .

Another fundamental statistical tool, this time concerning hyperparameter estimation, is the *likelihood* of the data, obtained by integrating \mathbf{X} out of the problem. In the continuous case, the result is the data probability density, which also depends on the hyperparameters of the problem:

$$p_{\mathbf{Y}}(\mathbf{y}; \boldsymbol{\theta}) = \int_{\mathcal{E}^S} p_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y}) d\mathbf{x} = \int_{\mathcal{E}^S} p_{\mathbf{B}}(\mathbf{y} - H(\mathbf{x})) p_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}. \quad (7.31)$$

Vector $\boldsymbol{\theta}$ designates the set of hyperparameters on which estimators $\hat{\mathbf{x}}$ constructed using the joint law (7.9), such as $\hat{\mathbf{x}}^{\text{MAP}}$ and $\hat{\mathbf{x}}^{\text{PM}}$, depend: temperature parameters T_x and T_b determining the regularization coefficient $\lambda = T_b/T_x$, together with all other degrees of freedom parametrizing Φ or Ψ (e.g. the threshold of the L_2L_1 or L_2L_0 functions). Finally, to recall that the density $p_{\mathbf{Y}}(\mathbf{y})$ depends on $\boldsymbol{\theta}$, it is noted $p_{\mathbf{Y}}(\mathbf{y}; \boldsymbol{\theta})$.

The *maximum likelihood* estimator $\hat{\boldsymbol{\theta}}^{\text{MV}}$ is defined as maximizing $p_{\mathbf{Y}}(\mathbf{y}; \boldsymbol{\theta})$ (see section 3.5). This is the preferred tool for tackling inverse problems in a non-supervised framework, which is the question covered in the next chapter.

The practical use of these statistical tools raises two important questions at two very different levels.

At the *information* level, what meaning should be given to these statistical quantities? For example, let us assume that μ defines a 95% confidence region:

$$\Pr(\mathbf{X} \in \mathbf{I}(\mu) | \mathbf{y}) = \int_{\mathbf{I}(\mu)} p(\mathbf{x} | \mathbf{y}) d\mathbf{x} = 0.95. \quad (7.32)$$

As the context is Bayesian, this is a *mean* confidence region, which takes account of the uncertainties connected with the measurements \mathbf{y} , quantified by the noise law, and also our uncertain knowledge of \mathbf{x} , quantified by the prior law. In accordance

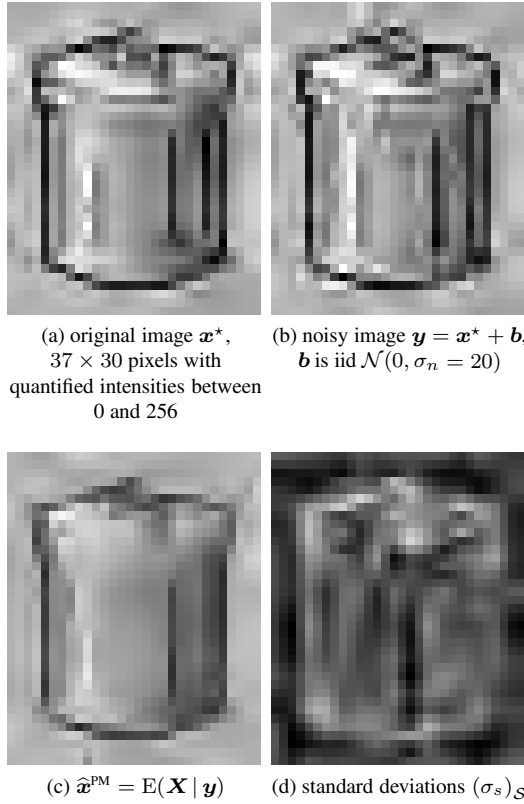


Figure 7.4. Mean and posterior standard deviation, (c) and (d) respectively, for a prior Gibbs-Markov energy law (7.28), with ϕ of type L_2L_1 . These quantities were calculated by an MCMC method (see section 7.4.2.2). They depend on three parameters: the standard deviation of the noise, fixed at its true value $\sigma_n = 20$, and (T, η) , the GMRF temperature and threshold of function ϕ , chosen empirically. The grey-level scale is dilated for the standard deviation, the maximum value of which does not exceed 16

with Bayesian principles, this prior should sum up a state of knowledge gained independently of the data. In these conditions, it can be observed *empirically* that equation (7.32) is true in the sense where the same experiment, repeated N times *in the same state of knowledge*, would find the event $\mathbf{X} \in \mathbf{I}(\mu)$ realized about $0.95 N$ times if N is large. In most practical situations, it is obvious that the Gibbs-Markov models of section 7.3.4 disobey these principles, in the sense that they only take account of a part of the prior information available. The information taken into consideration is sufficient to produce usable estimators and indications of the associated uncertainties but it does not allow these uncertainties to be accurately quantified. The meaning

of the standard deviations of Figure 7.4d is thus only *qualitative*. As expected, they decrease inside a homogeneous domain such as the background and increase at the edges. However, their precise value is not significant. For example, we can verify that $\mathbf{x}^* \notin \mathbf{I}(7)$, i.e., there are sites s such that $|\hat{x}_s^{\text{PM}} - x_s^*| > 7\sigma_s$. For these sites, the value of σ_s is clearly underestimated. Similarly, the posterior mean of Figure 7.4c is not sufficiently contrasted.

From an *operational* point of view, to evaluate quantities (7.29), (7.30) and (7.31), it is necessary to be able to calculate the integrals over \mathcal{E}^S . The first two can be written component by component as simple integrals:

$$\begin{aligned}\hat{x}_s^{\text{PM}} &= \mathbb{E}(X_s | \mathbf{y}) = \int_{\mathcal{E}} x_s p_{X_s | \mathbf{Y}}(x_s | \mathbf{y}) dx_s, \\ \text{var}(X_s | \mathbf{y}) &= \int_{\mathcal{E}} (x_s - \hat{x}_s^{\text{PM}})^2 p_{X_s | \mathbf{Y}}(x_s | \mathbf{y}) dx_s,\end{aligned}$$

but these expressions do not allow a direct calculation as the marginal densities $p_{X_s | \mathbf{Y}}$ are themselves integrals over \mathcal{E}^{S-1} . As for equation (7.31), we will concern ourselves with its maximization rather than its evaluation, which does not simplify the problem, at least at first glance.

Outside special cases such as the Gaussian case, the calculations may appear insurmountable, or at least incompatible with practical uses. Fortunately, special pseudo-random techniques, called *Markov Chain Monte Carlo* (MCMC) techniques, help us get around many of the calculation difficulties. These techniques are introduced in the next section. The reader may refer to [ROB 04] for a more comprehensive presentation in a generic inferential context, and to [BRÉ 99, WIN 03] in the context of spatial statistics and image analysis.

7.4.2. Stochastic sampling

By *sampling*, we mean the pseudo-random generation of realizations of a probability law. In imaging, and particularly as far as GMRFs are concerned, sampling has several roles. The most obvious is in image synthesis. GMRF realizations can be used to “dress” objects, by giving them a suitable textured appearance [CRO 83]. The realizations of GMRFs of energy (7.28), with ϕ of the L_2L_1 type, such as that represented in Figure 7.5, have a “grained” texture.

In image restoration, it is the posterior law $p_{\mathbf{X} | \mathbf{Y}=\mathbf{y}}$ that it is interesting to simulate, in the aim of calculating Bayesian estimators such as the posterior mean $\hat{\mathbf{x}}^{\text{PM}} = \mathbb{E}(\mathbf{X} | \mathbf{y})$. In fact, if $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \dots, \mathbf{X}^{(K)}$ is a series of random variables having

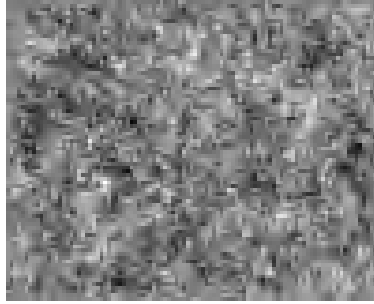


Figure 7.5. Simulation of a GMRF of energy (7.28), where ϕ is the Huber function (L_2L_1 function), on a 80×100 grid. The value of the boundary pixels is kept to zero to avoid the improper character of the law

the law $p_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}$, we have:

$$\mathbb{E}(F(\mathbf{X}) | \mathbf{Y} = \mathbf{y}) = \lim_{K \rightarrow \infty} \frac{1}{K} \sum_{k=1}^K F(\mathbf{X}^{(k)}) \quad (7.33)$$

almost surely for any function F for which this series is *ergodic*. In particular, we can approach $\hat{\mathbf{x}}^{\text{PM}}$ through the empirical mean of K realizations of the posterior law if these realizations are extracted from a series that is *ergodic for the mean*.

In the case of an independent series, the ergodicity is guaranteed provided that $\mathbb{E}(F(\mathbf{X}) | \mathbf{y})$ exists, which is the case if $\mathbb{E}(|F(\mathbf{X})| | \mathbf{y}) < \infty$. This is the strong law of large numbers. We could thus envisage applying equation (7.33) starting from an iid series $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \dots, \mathbf{X}^{(K)}$ of instantaneous law $p_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}$. The *standard* Monte Carlo method works precisely on this principle [HAM 64, Chapter 5]. Unfortunately, this method is not suitable as generating a single sample of the law $p_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}$ already involves an iterative process too costly to be repeated enough. We will nevertheless examine this procedure because it remains at the heart of the MCMC methods, which are better suited to the situation (see section 7.4.2.2).

7.4.2.1. Iterative sampling methods

Generating realizations of random quantities is a field of study in itself, which employs a vast range of techniques. When the quantities to be generated are of small dimensions (typically a real random variable) there are *direct* techniques for many families of probability laws. One of the most commonly used is the change of variable using a law with well-controlled sampling, the “generic” law in this respect being the uniform law [PRE 86, section 7.2]. For more complicated laws, rejection techniques are worthy of attention [PRE 86, section 7.3].

When the dimension of the quantity to be sampled exceeds a few units, direct techniques become inefficient, except in special cases such as the Gaussian case: a random vector \mathbf{X} with values in \mathbb{R}^S for which the law is $\mathcal{N}(\mathbf{m}, \mathbf{R})$ can be sampled in the form $\mathbf{A}\mathbf{E} + \mathbf{m}$, where \mathbf{E} is a normalized Gaussian vector (i.e., of law $\mathcal{N}(\mathbf{0}, \mathbf{I})$), provided that the matrix \mathbf{A} verifies $\mathbf{A}\mathbf{A}^T = \mathbf{R}$ (\mathbf{A} is not necessarily square). In other words, sampling a Gaussian vector of arbitrary covariance comes down to a problem of factorizing the covariance ($\mathbf{R} = \mathbf{A}\mathbf{A}^T$), generating independent Gaussian variables, then making linear combinations ($\mathbf{X} = \mathbf{A}\mathbf{E} + \mathbf{m}$).

In more general cases, sampling by Markov chain becomes inevitable. A series $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \dots, \mathbf{X}^{(k)}, \dots$ is generated, which is a Markov chain and converges in distribution towards the desired law. In the rest of this section, we assume that this law has a density p , making it implicit that this notation can designate a conditional density, such as $p_{\mathbf{X}|\mathbf{Y}}$.

The series $(\mathbf{X}^{(k)})$, Markovian by construction, is obtained as follows: a first quantity $\mathbf{X}^{(1)}$ with values in \mathcal{E}^S is generated from an arbitrary law, preferably a simple one. The following components are approached gradually by random transformation. This transformation is defined by a conditional density $t(\mathbf{x}' | \mathbf{x})$ called the *transition kernel*. There are several possible forms for the kernel. The oldest is the *Metropolis algorithm* [MET 53] (Table 7.1a). The more recent *Gibbs sampler* [GEM 84] is specifically adapted for sampling a GMRF, in as much as the transition takes place by sampling of the conditional laws $p(x_s | \mathbf{x}_{S \setminus s})$, that are local according to (7.18) (Table 7.1b). These are two particular cases of a more general structure, the Metropolis-Hastings sampler [HAS 70].

The first property expected of these structures is convergence *in distribution* towards p . Whether this property is satisfied obviously depends on the structure of the *transition kernel* $t(\mathbf{x}' | \mathbf{x})$, i.e., on the probability law that governs the generation of $\mathbf{X}^{(k+1)}$ when $\mathbf{X}^{(k)}$ is known. We have the following result, which is a simplified, restricted statement of [ROB 04, theorem 6.53].

Let $(\mathbf{X}^{(k)})_{k \in \mathbb{N}}$ be a *homogeneous, aperiodic* Markov chain, having a transition kernel t that admits the law π of density p for the invariant law:

$$\int_{\mathcal{E}^S} t(\mathbf{x}' | \mathbf{x}) p(\mathbf{x}) d\mathbf{x} = p(\mathbf{x}')$$

Thus, $\mathbf{X}^{(k)}$ converges in distribution towards π .

The homogeneity corresponds to the fact that the transition kernel does not evolve during the iterations. The meaning of the invariance of π is the following: if the law of $\mathbf{X}^{(k)}$ is π , that of $\mathbf{X}^{(k+1)}$ is also π . On this subject, note that the *equilibrium condition* $t(\mathbf{x}' | \mathbf{x})p(\mathbf{x}) = t(\mathbf{x} | \mathbf{x}')p(\mathbf{x}')$ implies the invariance of π . Aperiodicity is

<p style="text-align: center;">(a) Metropolis algorithm</p> <p>Initialization: choose $\mathbf{X}^{(1)}$ arbitrarily.</p> <ol style="list-style-type: none"> 1) Current configuration: $\mathbf{X}^{(k)} = \mathbf{x}$. 2) Propose \mathbf{x}' by sampling a symmetric <i>proposition kernel</i> $p(\mathbf{x}' \mathbf{x}) = p(\mathbf{x} \mathbf{x}')$. 3) If $p(\mathbf{x}') \geq p(\mathbf{x})$, $\mathbf{X}^{(k+1)} = \mathbf{x}'$. Otherwise, still accept $\mathbf{X}^{(k+1)} = \mathbf{x}'$ with the probability $p(\mathbf{x}')/p(\mathbf{x})$; in case of rejection, $\mathbf{X}^{(k+1)} = \mathbf{x}$.
<p style="text-align: center;">(b) Gibbs sampler</p> <p>Initialization: choose $\mathbf{X}^{(1)}$ arbitrarily.</p> <ol style="list-style-type: none"> 1) Current configuration: $\mathbf{X}^{(k)} = \mathbf{x}$. 2) Choose a site s “at random” (e.g. equiprobably). 3) Sample $X_s^{(k+1)}$ according to $p(x_s \mathbf{x}_{S \setminus s})$, the conditional law deduced from p. 4) $\mathbf{X}_{S \setminus s}^{(k+1)} = \mathbf{x}_{S \setminus s} = \mathbf{X}_{S \setminus s}^{(k)}$

Table 7.1. *Metropolis and Gibbs algorithms for sampling probability density p*

a more technical concept [ROB 04, section 6.3.3] that allows cyclic behavior to be excluded. These conditions are naturally respected by simple, intuitive choices. For example, for the proposition kernel p of the Metropolis algorithm (Table 7.1a, step 2), in the case of $S = \mathbb{R}$, the *random walk* $\mathbf{x}' = \mathbf{x} + \varepsilon$ ensures the symmetry of p if ε follows a centrosymmetric density law p_ε : $p_\varepsilon(-\varepsilon) = p_\varepsilon(\varepsilon)$.

For the choice of the site to be resampled in the Gibbs sampler (Table 7.1b, step 2), it is logical to favor equiprobability but the convergence in distribution also occurs for deterministic strategies of site exploration. This option makes the Gibbs sampler partially parallelizable: pixels that are not two-by-two neighbors can be resampled independently, and therefore simultaneously. In this way, Figure 7.5 was generated by “chequer-board” style updating, half the sites being resampled at each iteration (alternatively, all the “whites” or all the “blacks”).

Sampling and optimization are related problems. The simple fact of never accepting the proposition $\mathbf{X}^{(k+1)} = \mathbf{x}'$ if $p(\mathbf{x}') < p(\mathbf{x})$ (Table 7.1a, step 3) would transform the Metropolis algorithm into a *random search optimization* method. Similarly, the Gibbs sampler is a “stochastic version” of the Gauss-Seidel descent technique (section 2.2.2.2.1). In this respect, it can be said that Markov chain sampling methods are to sampling as iterative methods of criterion descent are to optimization. Like the

descent techniques, sampling techniques can have a multitude of variations. For instance, [GEM 92] and [GEM 95] stress global updating methods based on sampling semi-quadratic energy models that are supposed to be faster than methods proceeding pixel by pixel. These algorithms are “stochastic versions” of the descent algorithms presented in Chapter 6.

7.4.2.2. Monte Carlo method of the MCMC kind

Rather than in the convergence in distribution, the interest of iterative methods lies in a “generalized law of large numbers” for the calculation of stochastic quantities: Property (7.33) is true when $(\mathbf{X}^{(k)})_{k \in \mathbb{N}}$ is a Markov chain obtained in the conditions of the previous subsection. At first glance, this result is surprising since the density of the successive variables $\mathbf{X}^{(k)}$ is not necessarily p ; the important thing here is that it approaches p when k increases.

The MCMC methods take advantage of this result by approaching a statistical expectation through the empirical mean of a finite number of successive iterations of a sampler. This is how Figures 7.4c and d were obtained. First, $K = 500$ samples $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(K)}$ converging towards the posterior law of the GMRF were generated by a checker-board Gibbs sampler (going from k to $k + 1$ corresponds to a complete update, all the “whites” then all the “blacks”). We deduce from this:

$$\widehat{\mathbf{m}} = \frac{1}{K - k_0 + 1} \sum_{k=k_0}^K \mathbf{X}^{(k)}, \quad \widehat{\mathbf{v}} = \frac{1}{K - k_0 + 1} \sum_{k=k_0}^K (\mathbf{X}^{(k)})^{[2]},$$

and finally $\widehat{\sigma} = (\widehat{\mathbf{v}} - \widehat{\mathbf{m}}^{[2]})^{[1/2]}$, where $^{[p]}$ means taking the exponent term by term. The summations leading to $\widehat{\mathbf{m}}$ and $\widehat{\mathbf{v}}$ can be performed recursively during sampling. In addition, the introduction of a *burn-in time* k_0 means that the first k_0 samples, for which the distribution is the furthest from the objective, are not used [GEY 92, section 3.7]. The choice of k_0 for a fixed K is the result of a “bias-variance compromise”, which is empirically set at $k_0 = 10$ here.

A more thorough mathematical presentation of MCMC methods would lead us to consider the *speed of convergence* of the various structures. Like minimization algorithms, sampling algorithms lend themselves to analytical calculations of the convergence rate [ROB 04]. In fact, some sampling algorithms are extremely long. The most important thing is to come to a qualitative understanding of why. The main pit-fall to be avoided is cyclic sampling of conditional laws $X_s | \mathbf{X}_{\mathcal{S} \setminus s}$ when the random variables X_s are strongly dependent. Although it is true that a Gibbs sampler proceeding in this way is likely to converge towards the joint law, the convergence will be very slow, successive samples being strongly correlated and the empirical means calculated having a residual variance that decreases slowly. This problem is similar to the more common one posed by trying to find the minimum of a narrow valley by a Gauss-Seidel technique (Chapter 2, section 2.2.2.2.1), which is illustrated in [PRE 86, Figure 10.6.1].

7.4.2.3. Simulated annealing

The posterior expectation $\hat{\mathbf{x}}^{\text{PM}} = \mathbb{E}(\mathbf{X} \mid \mathbf{y})$ can be approached by empirical means coming from an MCMC method. More generally, as shown in [MAR 87], the same is true for other Bayesian estimators with separable cost. This is not the case for the MAP (7.10). When the posterior energy is convex (or even quadratic) in \mathbf{x} , it would be absurd to try to calculate $\hat{\mathbf{x}}^{\text{MAP}}$ using a stochastic algorithm. On the other hand, when this energy has local minima, or when the state space is a large *discrete* space, the idea of using a stochastic search method to find $\hat{\mathbf{x}}^{\text{MAP}}$ is attractive as traps such as local minima can then be avoided.

Simulated annealing possesses this property, which was first demonstrated in the discrete case [GEM 84], then extended to the continuous case (see the references given in [GEM 92]). Its basic structure is an iterative sampling method made *inhomogeneous* by the slow decrease of a temperature parameter during the iterations. The brief presentation below considers the continuous case. Let p be the density of the law to be maximized and let

$$p_T(\mathbf{x}) = \frac{1}{Z(T)} (p(\mathbf{x}))^{1/T}$$

with

$$Z(T) = \int_{\mathcal{E}^S} (p(\mathbf{x}))^{1/T} d\mathbf{x}$$

be the renormalized density at temperature T . Finally, let $t_T(\mathbf{x}' \mid \mathbf{x})$ be a transition kernel allowing p_T to be sampled.

Simulated annealing combines two properties. One concerns the behavior of the inhomogeneous sampling method that generates a random series $(\mathbf{X}^{(k)})_{k \in \mathbb{N}}$ following $t_{T(k)}(\mathbf{x}' \mid \mathbf{x})$, where $(T(k))_{k \in \mathbb{N}}$ is a deterministic, decreasing series tending towards zero. If the initial temperature is “sufficiently high” and if the decrease is “sufficiently slow”, it can be shown that the law of $\mathbf{X}^{(k)}$ has a density close to $p_{T(k)}$, when k is large. The other property concerns the behavior of p_T as a function of T . As illustrated in Figure 7.6, this density becomes uniform at high temperatures, whereas it is concentrated on $\hat{\mathbf{x}}^{\text{MAP}}$ at a temperature tending to zero.

The first simulated annealing algorithm in image processing (deconvolution, segmentation) proposed by [GEM 84] is based on a Gibbs sampler working in the discrete case with interactive edge variables. Faster forms were subsequently put forward, by moving on to the continuous case with sampling of models having half-quadratic energy [GEM 92, GEM 95]. Nevertheless, simulated annealing remains little used as an optimization method in imaging. It is clearly more costly to minimize a multimodal criterion by simulated annealing than a convex criterion by a well chosen descent algorithm. And when the convex or non-convex nature of a criterion depends on the choice of model, calculating cost and simplicity are often decisive arguments.

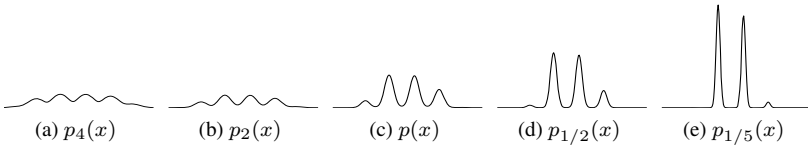


Figure 7.6. Probability densities deduced from one another by temperature changes: $p_T(x) \propto (p(x))^{1/T}$. When T increases, the density becomes uniform over its definition set. Conversely, a decrease in T accentuates the differences between the probabilities of the most probable and least probable events. When $T \searrow 0$, the probability becomes evenly distributed over the values of x maximizing $p(x)$. In the case of a unique maximizer, the limit law is a Dirac distribution

7.5. Conclusion

In the Bayesian probabilistic framework, local penalty functions can be interpreted as Gibbs potentials and criterion minimization as finding a maximum *a posteriori*. Does this probabilistic interpretation provide the tools for constructing more objective models than the qualitative approach of Chapter 6? The answer is “yes” in certain specific cases: the Kolmogorov model for imaging through atmospheric turbulence (Chapter 10) and the Poisson process for counting particles in corpuscular imaging (Chapter 14). However, for many imaging problems, the models are still “hand built”. Thus, Gibbs-Markov fields are often defined by specifying an ad hoc Gibbs energy. Although it may be true that the probabilistic framework enables, among other things, the characterization of Gibbs-Markov models by a conditional independence property, this property is not constructive, in the sense that it does not allow models to be effectively specified, apart from a few notable exceptions such as Markov chains and unilateral fields [IDI 01].

In addition, the Bayesian probabilistic framework gives a sense to mathematical operations such as marginalization and conditioning, which have no natural equivalents in the deterministic framework. This allows estimators other than the maximum *a posteriori* to be defined, and also enables uncertainties to be quantified (section 7.4.1) and “second-level problems” such as hyperparameter estimation to be formalized (Chapter 8).

Finally, on the subject of calculation aspects, it is incontestable that the emergence of MCMC methods in signal and image processing was an important practical advance. In particular, these numerical tools enable statistical indicators to be evaluated in a realistic environment (small amounts of data and presence of noise), whereas very few analytical results are available on the behavior of the estimators outside the asymptotic regime.

7.6. Bibliography

- [BAD 89] BADDELEY A. J., MØLLER J., “Nearest-neighbour Markov point processes and random sets”, *Int. Statist. Rev.*, vol. 57, p. 89-121, 1989.
- [BES 74] BESAG J. E., “Spatial interaction and the statistical analysis of lattice systems (with discussion)”, *J. R. Statist. Soc. B*, vol. 36, num. 2, p. 192-236, 1974.
- [BES 86] BESAG J. E., “On the statistical analysis of dirty pictures (with discussion)”, *J. R. Statist. Soc. B*, vol. 48, num. 3, p. 259-302, 1986.
- [BOU 93] BOUMAN C. A., SAUER K. D., “A generalized Gaussian image model for edge-preserving MAP estimation”, *IEEE Trans. Image Processing*, vol. 2, num. 3, p. 296-310, July 1993.
- [BRÉ 99] BRÉMAUD P., *Markov Chains. Gibbs Fields, Monte Carlo Simulation, and Queues*, Texts in Applied Mathematics 31, Springer, New York, NY, 1999.
- [CRO 83] CROSS G. R., JAIN A. K., “Markov random field texture models”, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-5, p. 25-39, 1983.
- [DEM 89] DEMOMENT G., “Image reconstruction and restoration: overview of common estimation structure and problems”, *IEEE Trans. Acoust. Speech, Signal Processing*, vol. ASSP-37, num. 12, p. 2024-2036, Dec. 1989.
- [DES 95] DESCOMBES X., MANGIN J.-F., PECHERSKY E., SIGELLE M., “Fine structure preserving Markov model for image processing”, in *9th Scand. Conf. Image Analysis SCIA'95*, Uppsala, Sweden, p. 349-356, 1995.
- [FRA 70] FRANKLIN J. N., “Well-posed stochastic extensions of ill-posed linear problems”, *J. Math. Anal. Appl.*, vol. 31, p. 682-716, 1970.
- [GEM 84] GEMAN S., GEMAN D., “Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images”, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, num. 6, p. 721-741, Nov. 1984.
- [GEM 92] GEMAN D., REYNOLDS G., “Constrained restoration and the recovery of discontinuities”, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, num. 3, p. 367-383, Mar. 1992.
- [GEM 95] GEMAN D., YANG C., “Nonlinear image recovery with half-quadratic regularization”, *IEEE Trans. Image Processing*, vol. 4, num. 7, p. 932-946, July 1995.
- [GEY 92] GEYER C. J., “Practical Markov chain Monte-Carlo (with discussion)”, *Statistical Science*, vol. 7, p. 473-511, 1992.
- [GUY 95] GUYON X., *Random Fields on a Network. Modeling, Statistics, and Applications*, Springer Verlag, New York, NY, 1995.
- [HAM 64] HAMMERSLEY J. M., HANDSCOMB D. C., *Monte Carlo Methods*, Methuen, London, UK, 1964.
- [HAS 70] HASTINGS W. K., “Monte Carlo sampling methods using Markov Chains and their applications”, *Biometrika*, vol. 57, p. 97, Jan. 1970.

- [IDI 01] IDIER J., GOUSSARD Y., RIDOLFI A., “Unsupervised image segmentation using a telegraph parameterization of Pickard random fields”, in MOORE M. (Ed.), *Spatial statistics: Methodological Aspects and Applications*, vol. 159 of *Lecture notes in Statistics*, p. 115-140, Springer Verlag, New York, NY, 2001.
- [KÜN 94] KÜNSCH H. R., “Robust priors for smoothing and image restoration”, *Ann. Inst. Stat. Math.*, vol. 46, num. 1, p. 1-19, 1994.
- [MAR 87] MARROQUIN J. L., MITTER S. K., POGGIO T. A., “Probabilistic solution of ill-posed problems in computational vision”, *J. Amer. Stat. Assoc.*, vol. 82, p. 76-89, 1987.
- [MET 53] METROPOLIS N., ROSENBLUTH A. W., ROSENBLUTH M. N., TELLER A. H., TELLER E., “Equations of state calculations by fast computing machines”, *J. Chem. Phys.*, vol. 21, p. 1087-1092, June 1953.
- [MOR 96] MORRIS R., DESCOMBES X., ZERUBIA J., An analysis of some models used in image segmentation, Research Report num. 3016, INRIA, Sophia Antipolis, France, Oct. 1996.
- [MOU 97] MOURA J. M. F., SAURAJ G., “Gauss-Markov random fields (GMrf) with continuous indices”, *IEEE Trans. Inf. Theory*, vol. 43, num. 5, p. 1560-1573, Sep. 1997.
- [PIC 87] PICKARD D. K., “Inference for discrete Markov fields: The simplest nontrivial case”, *J. Acoust. Soc. Am.*, vol. 82, p. 90-96, 1987.
- [PRE 86] PRESS W. H., FLANNERY B. P., TEUKOLSKY S. A., VETTERLING W. T., *Numerical Recipes, the Art of Scientific Computing*, Cambridge University Press, Cambridge, MA, 1986.
- [QIA 90] QIAN W., TITTERINGTON D. M., “Parameter estimation for hidden Gibbs chains”, *Statistics & Probability Letters*, vol. 10, p. 49-58, June 1990.
- [ROB 04] ROBERT C. P., CASELLA G., *Monte Carlo Statistical Methods*, Springer Texts in Statistics, Springer Verlag, New York, NY, 2nd edition, 2004.
- [SAQ 98] SAQUIB S. S., BOUMAN C. A., SAUER K. D., “ML parameter estimation for Markov random fields with applications to Bayesian tomography”, *IEEE Trans. Image Processing*, vol. 7, num. 7, p. 1029-1044, July 1998.
- [SIL 90] SILVERMAN B. W., JENNISON C., STANDER J., BROWN T. C., “The specification of edge penalties for regular and irregular pixel images”, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-12, num. 10, p. 1017-1024, Oct. 1990.
- [WIN 03] WINKLER G., *Image Analysis, Random Fields and Dynamic Monte Carlo Methods*, Springer Verlag, Berlin, Germany, 2nd edition, 2003.

Chapter 8

Unsupervised Problems

8.1. Introduction and statement of problem

We are going to consider the “generic” problem of estimating a physical quantity degraded by a linear process and corrupted by noise, which will be taken to be white. This formulation, with reasonable approximations, includes many problems commonly encountered in signal and image processing, such as segmentation, deconvolution and the reconstruction of one- or multi-dimensional quantities. Here, we will limit ourselves to the case where the quantities of interest are indexed by discrete, finite variables, a usual situation in the processing of sampled data of finite dimension. The system considered is thus governed by the equation:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{b} \quad (8.1)$$

where \mathbf{y} , \mathbf{x} and \mathbf{b} are vectors that contain the observed data, the unknown quantity to be estimated, and the samples of the noise respectively. \mathbf{H} is the matrix representing the linear degradation acting on \mathbf{x} .

As shown in the previous chapters, the estimation of \mathbf{x} is often an ill-posed problem and regularization is generally required if acceptable results are to be obtained. In this chapter, we adopt the Bayesian framework from the start, so the information about \mathbf{x} is summed up in the form of the posterior law:

$$p(\mathbf{x} | \mathbf{y}) \propto p(\mathbf{y} | \mathbf{x}) p(\mathbf{x}), \quad (8.2)$$

where $p(\mathbf{y} | \mathbf{x})$ and $p(\mathbf{x})$ denote the conditional density of \mathbf{y} when \mathbf{x} is known and the density of the prior law for \mathbf{x} respectively. Without loss of generality, we can write

$p(\mathbf{y} | \mathbf{x})$ and $p(\mathbf{x})$ in the form:

$$p(\mathbf{y} | \mathbf{x}) = p(\mathbf{y} | \mathbf{x}; \boldsymbol{\vartheta}) \propto \exp \{ -\Psi_{\boldsymbol{\vartheta}}(\mathbf{y} - \mathbf{H}\mathbf{x}) \}, \quad (8.3)$$

$$p(\mathbf{x}) = p(\mathbf{x}; \boldsymbol{\theta}) \propto \exp \{ -\Phi_{\boldsymbol{\theta}}(\mathbf{x}) \}. \quad (8.4)$$

This means that specifying laws $p(\mathbf{y} | \mathbf{x})$ and $p(\mathbf{x})$ is equivalent to specifying the functions Ψ and Φ , which, moreover, depend on parameters denoted $\boldsymbol{\vartheta}$ for Ψ and $\boldsymbol{\theta}$ for Φ . Once these functions have been specified, the quality of the results is strongly dependent on the accuracy with which the linear degradation \mathbf{H} is known and the values attributed to $\boldsymbol{\theta}$ and $\boldsymbol{\vartheta}$. In some situations, \mathbf{H} can be determined precisely, either from a knowledge of the phenomena under study or by preliminary tests. Similarly, it is often possible to specify the values of $\boldsymbol{\theta}$ and $\boldsymbol{\vartheta}$ empirically, by trial and error on simulated problems for example. However, approaches of this type are not always applicable, e.g. if \mathbf{H} is impossible to identify beforehand or undergoes large variations over time (case of communication channels), or if the method needs to be used by people who are not signal or image processing specialists to treat a variety of cases requiring the values $\boldsymbol{\theta}$ or $\boldsymbol{\vartheta}$ to be adapted. In such situations, it is desirable to estimate not only \mathbf{x} , but also \mathbf{H} , $\boldsymbol{\theta}$ and $\boldsymbol{\vartheta}$ (or some of these quantities) from the observed data \mathbf{y} . This is the problem of *unsupervised estimation*, which is the subject of this chapter.

Unsupervised estimation is a vast, difficult field. A variety of methods have been developed and it is impossible to present them fully here. For this reason, we will concentrate on a particular type of problem that will nevertheless allow us to bring out the difficulties intrinsic in unsupervised estimation and present a set of important methods in a consistent way. The type of problem we have chosen is that of a penalty function Φ that is *linear with respect to the parameters*¹ and *Markovian*, and can be expressed as:

$$\Phi_{\boldsymbol{\theta}}(\mathbf{x}) = \sum_i \theta_i N_i(\mathbf{x}) \quad (8.5)$$

where each $N_i(\mathbf{x})$ expresses *local* interactions between elements of \mathbf{x} that are neighbors (see Chapter 7). It should be stressed that many prior distributions in current use correspond to functions Φ having the above form. This is the case for Gaussian densities in particular.

First of all, we will consider that \mathbf{H} is perfectly known and equal to identity matrix \mathbf{I} , and that the noise, \mathbf{b} , is zero. This academic case of a *directly observed* random field will allow us to point out the difficulties in estimating $\boldsymbol{\theta}$ and present the fundamental concepts and basic techniques. Next, we will tackle the situation where \mathbf{H} is known but different from the identity matrix and where noise \mathbf{b} is present. This will make it necessary to adapt and extend the techniques introduced previously.

1. The probability laws thus defined form *exponential families*.

From the choices mentioned above, it appears that the techniques described in this chapter will be particularly suitable for the case where \mathbf{x} represents an image. It should be noted that the extension to 3D quantities does not pose any methodological problem; the extra difficulty comes essentially from the volume of data to be processed. In contrast, the 1D case can present great simplifications due to the causal structure of \mathbf{x} . Such simplifications will be pointed out as necessary as we go along.

8.2. Directly observed field

First, we consider the case of *complete data*, i.e., of a field \mathbf{X} , the realization of which, \mathbf{x}_0 , is known, and from which we intend to estimate parameters $\boldsymbol{\theta}$. Taking once again the model of an energy linearly dependent on the parameters, we have:

$$p(\mathbf{x}; \boldsymbol{\theta}) = \frac{1}{Z(\boldsymbol{\theta})} \exp \left\{ - \sum_i \theta_i N_i(\mathbf{x}) \right\} \quad (8.6)$$

where $Z(\boldsymbol{\theta})$ is the partition function (normalization constant):

$$Z(\boldsymbol{\theta}) = \int \exp \left\{ - \sum_i \theta_i N_i(\mathbf{x}) \right\} d\mathbf{x}. \quad (8.7)$$

It should be noted that the integral defined in equation (8.7) is taken with respect to variable \mathbf{x} , i.e., in the configuration space. It is therefore not possible to envisage its numerical evaluation by computer. Furthermore, it is impossible to formulate it analytically except in very special cases.

8.2.1. Likelihood properties

To make the derivations simpler, we prefer to handle the log-likelihood, which according to (8.6) is given by:

$$\log L(\boldsymbol{\theta}) = - \sum_i \theta_i N_i(\mathbf{x}_0) - \log Z(\boldsymbol{\theta}). \quad (8.8)$$

If we take the first derivative of this functional, we obtain:

$$\begin{aligned} \frac{\partial}{\partial \theta_i} \log L(\boldsymbol{\theta}) &= -N_i(\mathbf{x}_0) - \frac{1}{Z(\boldsymbol{\theta})} \frac{\partial}{\partial \theta_i} \int \exp \left\{ - \sum_i \theta_i N_i(\mathbf{x}) \right\} d\mathbf{x} \\ &= -N_i(\mathbf{x}_0) + \frac{1}{Z(\boldsymbol{\theta})} \int N_i(\mathbf{x}) \exp \left\{ - \sum_i \theta_i N_i(\mathbf{x}) \right\} d\mathbf{x} \\ &= -N_i(\mathbf{x}_0) + E_{\boldsymbol{\theta}}(N_i(\mathbf{x})) \end{aligned} \quad (8.9)$$

where $E_{\theta}(N_i(\mathbf{x}))$ represents the moment – i.e., the mathematical expectation – of $N_i(\mathbf{x})$ with respect to the law $p(\mathbf{x}; \theta)$. Similarly, we can obtain the expression for the elements of the Hessian \mathcal{H} , in the form:

$$\frac{\partial^2 \log L(\theta)}{\partial \theta_i \partial \theta_j} = E_{\theta}(N_i(\mathbf{x})) E_{\theta}(N_j(\mathbf{x})) - E_{\theta}(N_i(\mathbf{x}) N_j(\mathbf{x})) = -\text{cov}(N_i(\mathbf{x}) N_j(\mathbf{x})).$$

Matrix \mathcal{H} is negative, as it is the opposite of a covariance matrix. The log-likelihood is therefore concave. Algorithms of the gradient descent or conjugate gradient type are thus suitable for calculating the maximum likelihood (ML) for θ , as there are no local extrema. On the other hand, calculating the Hessian requires the evaluation of the moments $E_{\theta}(N_i(\mathbf{x}))$ and $E_{\theta}(N_i(\mathbf{x}) N_j(\mathbf{x}))$. As an analytical calculation of these quantities is not feasible, it is necessary to sample the distribution by Monte Carlo methods in order to evaluate them.

8.2.2. Optimization

8.2.2.1. Gradient descent

The maximization of the log-likelihood can be carried out using a Newton-type iterative scheme:

$$\theta^{n+1} = \theta^n - \mathcal{H}^{-1} \nabla_{\theta} \log L(\theta^n).$$

For the sake of simplicity, we can approach the expectations expressing the gradient and the Hessian by the corresponding value on a sample. In this vein, Younes has proved that the convergence of the following algorithm is almost sure [YOU 88]:

$$\theta_i^{n+1} = \theta_i^n + \frac{N_i(\mathbf{x}^{n+1}) - N_i(\mathbf{x}_0)}{(n+1)V} \quad (8.10)$$

where \mathbf{x}^{n+1} is a sample of the law $p(\mathbf{x}; \theta^n)$ and V a sufficiently large positive constant. Note that it is nevertheless necessary to sample the law *at each step of the algorithm*, which makes its use impossible in practice because of the computing time needed.

8.2.2.2. Importance sampling

The notion of importance sampling provides a less costly solution [DES 99, GEY92]. Let us consider the ratio of the partition functions for two distinct parameter vectors:

$$\frac{Z(\theta)}{Z(\omega)} = \frac{\int \exp\{-\sum_i \theta_i N_i(\mathbf{x})\} d\mathbf{x}}{\int \exp\{-\sum_i \omega_i N_i(\mathbf{x})\} d\mathbf{x}}. \quad (8.11)$$

We then have:

$$\begin{aligned}
 \frac{Z(\boldsymbol{\theta})}{Z(\boldsymbol{\omega})} &= \frac{\int \exp\{-\sum_i (\theta_i - \omega_i) N_i(\mathbf{x})\} \exp\{-\sum_i \omega_i N_i(\mathbf{x})\} d\mathbf{x}}{\int \exp\{-\sum_i \omega_i N_i(\mathbf{x})\} d\mathbf{x}} \\
 &= \frac{\int \exp\{-\sum_i (\theta_i - \omega_i) N_i(\mathbf{x})\} p(\mathbf{x}; \boldsymbol{\omega}) d\mathbf{x}}{\int \exp\{-\sum_i \omega_i N_i(\mathbf{x})\} d\mathbf{x}} \\
 &= E_{\boldsymbol{\omega}}(\exp\{-\sum_i (\theta_i - \omega_i) N_i(\mathbf{x})\}) .
 \end{aligned}$$

Let us slightly modify the definition of the log-likelihood by adding a constant (which leaves the argument of the minimum value unchanged):

$$\log L_{\boldsymbol{\omega}}(\boldsymbol{\theta}) = -\sum_i \theta_i N_i(\mathbf{x}_0) - \log \frac{Z(\boldsymbol{\theta})}{Z(\boldsymbol{\omega})} . \quad (8.12)$$

The first derivatives can then be written:

$$\begin{aligned}
 \frac{\partial \log L_{\boldsymbol{\omega}}(\boldsymbol{\theta})}{\partial \theta_i} &= -N_i(\mathbf{x}_0) - \frac{Z(\boldsymbol{\omega})}{Z(\boldsymbol{\theta})} \frac{1}{Z(\boldsymbol{\omega})} \frac{\partial}{\partial \theta_i} \int \exp\{-\sum_j (\theta_j - \omega_j) N_j(\mathbf{x})\} p(\mathbf{x}; \boldsymbol{\omega}) d\mathbf{x} \\
 &= -N_i(\mathbf{x}_0) + \frac{E_{\boldsymbol{\omega}}(N_i(\mathbf{x}) \exp\{-\sum_j (\theta_j - \omega_j) N_j(\mathbf{x})\})}{E_{\boldsymbol{\omega}}(\exp\{-\sum_j (\theta_j - \omega_j) N_j(\mathbf{x})\})} \quad (8.13)
 \end{aligned}$$

and the second derivatives can be expressed as:

$$\begin{aligned}
 \frac{\partial^2 \log L_{\boldsymbol{\omega}}(\boldsymbol{\theta})}{\partial \theta_j \partial \theta_i} &= \frac{E_{\boldsymbol{\omega}}(N_i(\mathbf{x}) e^{-\sum_k (\theta_k - \omega_k) N_k(\mathbf{x})}) E_{\boldsymbol{\omega}}(N_j(\mathbf{x}) e^{-\sum_k (\theta_k - \omega_k) N_k(\mathbf{x})})}{(E_{\boldsymbol{\omega}}(e^{-\sum_k (\theta_k - \omega_k) N_k(\mathbf{x})}))^2} \\
 &\quad - \frac{E_{\boldsymbol{\omega}}(N_i(\mathbf{x}) N_j(\mathbf{x}) e^{-\sum_k (\theta_k - \omega_k) N_k(\mathbf{x})}) E_{\boldsymbol{\omega}}(e^{-\sum_k (\theta_k - \omega_k) N_k(\mathbf{x})})}{(E_{\boldsymbol{\omega}}(e^{-\sum_k (\theta_k - \omega_k) N_k(\mathbf{x})}))^2} . \quad (8.14)
 \end{aligned}$$

The expressions obtained, although apparently more complex, lead to much faster algorithms. Here, the various statistical moments are taken with respect to the law $p(\mathbf{x}; \boldsymbol{\omega})$. In theory, it is thus sufficient to sample the law once with respect to a given set of parameters $\boldsymbol{\omega}$. The moments concerning the law for the various values $\boldsymbol{\theta}^n$ are then evaluated on the samples of the law $p(\mathbf{x}; \boldsymbol{\omega})$. In practice, some precautions should be taken however. If the two vectors of parameters $\boldsymbol{\theta}$ and $\boldsymbol{\omega}$ are too far apart, the estimation of the expectations will not be accurate. If the estimation of the expectations is to be robust, there must be enough overlap between the two distributions $p(\mathbf{x}; \boldsymbol{\omega})$ and $p(\mathbf{x}; \boldsymbol{\theta})$. During optimization, if the current value $\boldsymbol{\theta}^n$ is too far from $\boldsymbol{\omega}$, the law must be resampled taking $\boldsymbol{\omega} = \boldsymbol{\theta}^n$.

The algorithmic scheme then becomes the following:

- 1) calculate the image statistics $N_i(\mathbf{x}_0)$;
- 2) initialize $\boldsymbol{\theta}^0$ and set $n = 0$;
- 3) sample the distribution with the current values of the parameters $\boldsymbol{\theta}^n$;
- 4) estimate a confidence interval around $\boldsymbol{\theta}^n$;
- 5) find $\boldsymbol{\theta}^{n+1}$ as the maximizer of the log-likelihood defined by equation (8.12) with $\boldsymbol{\omega} = \boldsymbol{\theta}^n$ using formulae (8.13) and (8.14);
- 6) if $\boldsymbol{\theta}^{n+1}$ is at the edge of the confidence interval, set $n = n + 1$ and go back to 3; otherwise $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}^{n+1}$.

8.2.3. Approximations

Estimation in the ML sense thus proves very costly in computing time even if importance sampling is used. Other estimators, approximating ML, give algorithms that are sub-optimal but much more efficient from the computing time standpoint.

8.2.3.1. Encoding methods

The encoding method proposed by Besag [BES 74], is a first, simple alternative to ML. It consists of defining disjunct subsets of the image such that the pixels of a given subset are conditionally independent.

Let us assume that \mathbf{X} is a Gibbs-Markov field (GMRF):

$$\forall s \in S, p(x_s | x_t, t \neq s) = p(x_s | x_t, t \in \nu_s) \quad (8.15)$$

where ν_s is a neighborhood of s . Let S' be a subset of S such that:

$$\{s, t\} \subset S' \implies s \notin \nu_t, t \notin \nu_s. \quad (8.16)$$

The likelihood relative to the encoding set S' can then be written:

$$L^c(\boldsymbol{\theta}) = p(x_s, s \in S' | x_t, t \in S \setminus S') = \prod_{s \in S'} p(x_s | x_t, t \in \nu_s). \quad (8.17)$$

In the case of a GMRF, the $N_i(\mathbf{x})$ are written as a sum of local functions:

$$N_i(\mathbf{x}) = \sum_{c(i) \in C} n_i(x_s, s \in c(i)) \quad (8.18)$$

$c(i)$ is called a clique and contains a finite set of pixels such that:

$$\{s, t\} \subset c(i) \implies t \in \nu_s \quad (8.19)$$

The log-likelihood relative to the encoding set S' can then be written:

$$\begin{aligned}\log L^c(\theta) &= \sum_{s \in S'} \log p(x_s | x_t, t \in \nu_s) \\ &= \sum_{s \in S'} \left(- \sum_i \theta_i \left(\sum_{c(i) \ni s} n_i(r, r \in c(i)) \right) - \log Z_s(x_t, t \in \nu_s) \right).\end{aligned}$$

The advantage is that, here, we are dealing with a local partition function, which can be evaluated without resorting to sampling:

$$Z_s(\theta) = \int \exp \left\{ - \sum_i \theta_i \left(\sum_{c(i) \ni s} n_i(r, r \in c(i)) \right) \right\} dx_s. \quad (8.20)$$

Several estimators can be obtained by considering different encoding sets. However, for each estimator, only part of the data is used and the “optimum” way of combining the various resulting estimators is still an open question. In the literature, the estimator considered is usually defined by the mean of the estimators over the various encoding sets.

Let us take the example of 4-connexity. The neighborhood of a site $s = (u, v)$ is then defined as follows:

$$\nu_{(u,v)} = \{(u-1, v); (u+1, v); (u, v-1); (u, v+1)\}. \quad (8.21)$$

The probability of the realization of x can then be written:

$$\begin{aligned}f(x; \theta) &= \\ &= \frac{1}{Z_\theta} \exp \left(- \sum_{(u,v) \in S} \left(\theta_1 n_1(x_{(u,v)}) + \theta_2 n_2(x_{(u,v)}, (u+1, v)) + \theta_3 n_3(x_{(u,v)}, (u, v+1)) \right) \right).\end{aligned}$$

Two encoding sets can be defined:

$$\begin{aligned}S_1 &= \{(2u, 2v), (2u, 2v) \in S\} \cup \{(2u+1, 2v+1), (2u+1, 2v+1) \in S\}, \\ S_2 &= \{(2u+1, 2v), (2u+1, 2v) \in S\} \cup \{(2u, 2v+1), (2u, 2v+1) \in S\}.\end{aligned}$$

8.2.3.2. Pseudo-likelihood

At the cost of an approximation, pseudo-likelihood allows all the data to be taken into account [BES 74]. It is defined simply by extending equation (8.17) to the complete set of sites:

$$PL(\theta) = \prod_{s \in S} p(x_s | x_t, t \in \nu_s). \quad (8.22)$$

As for the encoding method, equation (8.22) only brings in local partition functions. If the pixels are conditionally independent, the pseudo-likelihood is equivalent to the likelihood. If not, it is just an approximation. It often gives more accurate results than the encoding method.

8.2.3.3. Mean field

The mean field approximation is a technique that was initially developed in statistical physics to study phase transition phenomena. It was then applied to image processing [CEL 03, GEI 91, ZER 93, ZHA 93] based on GMRF models. At each site, the mean field approximation neglects the fluctuations of the other sites, approximating their states by the mean. The random field is then approximated by a set of independent random variables.

The energy function defined in equation (8.5), approximated by the mean field, can be written:

$$\Phi_{\theta,s}^{\text{MF}}(x_s) = \Phi_{\theta}(\mathbf{x} = (x_s, m_t, t \neq s)), \quad (8.23)$$

where m_t represents the mean of site t . Let us write function $\Phi_{\theta,s}^{\text{MF}}(x_s)$ in the form:

$$\Phi_{\theta,s}^{\text{MF}}(x_s) = \Phi_{\theta,s}^{\text{MF},\nu}(x_s) + \bar{\Phi}_{\theta,s}^{\text{MF}}(m_t, t \neq s), \quad (8.24)$$

where $\bar{\Phi}_{\theta,s}^{\text{MF}}(x_t, t \neq s)$ does not depend on x_s . The Markov structure (local interactions) implies that function $\Phi_{\theta,s}^{\text{MF},\nu}(x_s)$ only depends on x_s and the means of its various neighbors $m_t, t \in \nu_s$.

The mean field approximation consists of approximating the marginal law in x_s , defined by:

$$p(x_s) = \frac{1}{Z(\theta)} \int_{x_t, t \neq s} \exp\{-\Phi_{\theta}(\mathbf{x})\} dx_t, \quad (8.25)$$

by the law:

$$p_s^{\text{MF}}(x_s) = \frac{1}{Z_s^{\text{MF}}(\theta)} \exp\{-\Phi_{\theta,s}^{\text{MF}}(x_s)\}, \quad (8.26)$$

which can also be written:

$$p_s^{\text{MF}}(x_s) = \frac{1}{Z_s^{\text{MF},\nu}(\theta)} \exp\{-\Phi_{\theta,s}^{\text{MF},\nu}(x_s)\}. \quad (8.27)$$

The Markov property then allows the mean field approximation to be written as the local conditional probability:

$$p_s^{\text{MF}}(x_s) = p(x_s | x_t = m_t, t \in \nu_s). \quad (8.28)$$

Thus, at each site, we have defined marginal laws associated with independent random variables. The mean field approximation of the random field under consideration

can then be written as follows:

$$p^{\text{MF}}(\mathbf{x}) = \prod_{s \in S} p_s^{\text{MF}}(x_s). \quad (8.29)$$

Mean field estimation thus consists of maximizing the likelihood defined by:

$$L^{\text{MF}}(\boldsymbol{\theta}) = p^{\text{MF}}(\mathbf{x}_0) = \prod_{s \in S} p_s^{\text{MF}}(\mathbf{x}_0(s)). \quad (8.30)$$

It is worth noting that the likelihood obtained by the mean field approximation is very close to the pseudo-likelihood. The difference lies in the conditional variables since, here, neighboring sites are fixed at their means whereas, for pseudo-likelihood, they are fixed at the values of the realization under consideration $x_t = x_0(t)$. The difficulty lies in estimating the m_t s as, unlike in the case of pseudo-likelihood, the values of the conditioning variables are unknown here. In order to estimate these values, we state that the mean of a site s , calculated for the approximation in s , must be equal to that used for the approximation of the laws of the sites that are neighbors of s . By definition, the mean of a site s , in the approximation sense, is given by:

$$m_s = E_s^{\text{MF}}(x_s) = \int x_s p_s^{\text{MF}}(x_s) dx_s = \frac{1}{Z_s^{\text{MF}, \nu}(\boldsymbol{\theta})} \int x_s \exp\{-\Phi_{\boldsymbol{\theta}, s}^{\text{MF}, \nu}(x_s)\} dx_s.$$

This expectation depends on the means of the neighbors of s . We thus have an equation of the type:

$$m_s = p^{\text{MF}}(m_t, t \in \nu_s). \quad (8.31)$$

If we consider each of the sites, we obtain a system of equations. Solving this system by an iterative algorithm gives an approximation of the mean values m_s . The likelihood, given by (8.30), is then easy to maximize as the terms of the product are independent.

8.3. Indirectly observed field

8.3.1. Statement of problem

We will now concern ourselves with the case where the system is governed by equation (8.1) in which the noise process is present. We also assume that the linear distortion \mathbf{H} is known but can be different from the identity matrix. The aim here is to estimate parameters $\boldsymbol{\theta}$ and $\boldsymbol{\vartheta}$ that control the probability distributions $p(\mathbf{x}; \boldsymbol{\theta})$ and $p(\mathbf{y} | \mathbf{x}; \boldsymbol{\vartheta})$ respectively. Here again, we will limit ourselves to ML estimators, so the estimates are defined by:

$$(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\vartheta}}) = \arg \max_{\boldsymbol{\theta}, \boldsymbol{\vartheta}} p(\mathbf{y}; \boldsymbol{\theta}, \boldsymbol{\vartheta}). \quad (8.32)$$

An initial difficulty arises from the fact that, in most cases, it is not possible to obtain the closed-form expression for $p(\mathbf{y}; \boldsymbol{\theta}, \boldsymbol{\vartheta})$, even to within a normalization factor. We only have the relation:

$$p(\mathbf{y}; \boldsymbol{\theta}, \boldsymbol{\vartheta}) = \int p(\mathbf{y}, \mathbf{x}; \boldsymbol{\theta}, \boldsymbol{\vartheta}) d\mathbf{x} = \int p(\mathbf{y} | \mathbf{x}; \boldsymbol{\vartheta}) p(\mathbf{x}; \boldsymbol{\theta}) d\mathbf{x} \quad (8.33)$$

where $p(\mathbf{y} | \mathbf{x}; \boldsymbol{\vartheta})$ and $p(\mathbf{x}; \boldsymbol{\theta})$ are part of the problem specifications but where the integral above cannot be evaluated analytically. *Expectation Maximization* (EM) techniques provide a fairly general approach for maximizing $p(\mathbf{y}; \boldsymbol{\theta}, \boldsymbol{\vartheta})$ on the basis of the joint likelihood $p(\mathbf{y}, \mathbf{x}; \boldsymbol{\theta}, \boldsymbol{\vartheta}) = p(\mathbf{y} | \mathbf{x}; \boldsymbol{\vartheta}) p(\mathbf{x}; \boldsymbol{\theta})$, but without explicit evaluation of the integral present in equation (8.33). We should stress that bringing in the quantities $p(\mathbf{y} | \mathbf{x}; \boldsymbol{\vartheta})$ and $p(\mathbf{x}; \boldsymbol{\theta})$ will lead to the difficulties mentioned in section 8.2. We now present the EM algorithm before examining its application to hyperparameter estimation.

8.3.2. EM algorithm

Let $\boldsymbol{\theta}$ be a vector parameter that we intend to estimate by maximizing the likelihood $p(\mathbf{y}; \boldsymbol{\theta})$, where \mathbf{y} is a vector of observed data. The EM algorithm is an iterative procedure that produces an increase in the likelihood at each iteration and uses an auxiliary variable \mathbf{X} whose role is, roughly, to make the *extended likelihood* $p(\mathbf{y} | \mathbf{x}; \boldsymbol{\theta}) p(\mathbf{x}; \boldsymbol{\theta})$ easier to evaluate than the original likelihood $p(\mathbf{y}; \boldsymbol{\theta})$ [DEM 77]. Let $\boldsymbol{\theta}^0$ be the current value of the estimate and let us define the following quantities:

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^0; \mathbf{y}) \triangleq \int p(\mathbf{x} | \mathbf{y}; \boldsymbol{\theta}^0) \log p(\mathbf{y}, \mathbf{x}; \boldsymbol{\theta}) d\mathbf{x}, \quad (8.34)$$

$$D(\boldsymbol{\theta} || \boldsymbol{\theta}^0) \triangleq \int p(\mathbf{x} | \mathbf{y}; \boldsymbol{\theta}^0) \log \frac{p(\mathbf{x} | \mathbf{y}; \boldsymbol{\theta}^0)}{p(\mathbf{x} | \mathbf{y}; \boldsymbol{\theta})} d\mathbf{x}. \quad (8.35)$$

We observe that $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^0; \mathbf{y})$ can be interpreted as the following mathematical expectation:

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^0; \mathbf{y}) = \mathbb{E}(\log p(\mathbf{y}, \mathbf{x}; \boldsymbol{\theta}) | \mathbf{y}; \boldsymbol{\theta}^0) \quad (8.36)$$

and that $D(\boldsymbol{\theta} || \boldsymbol{\theta}^0)$ is the Kullback pseudo-distance between the probability densities $p(\mathbf{x} | \mathbf{y}; \boldsymbol{\theta})$ and $p(\mathbf{x} | \mathbf{y}; \boldsymbol{\theta}^0)$, which is always positive or zero [DAC 86]. In addition, it is easy to obtain the relation:

$$\log p(\mathbf{y}; \boldsymbol{\theta}) - \log p(\mathbf{y}; \boldsymbol{\theta}^0) = Q(\boldsymbol{\theta}, \boldsymbol{\theta}^0; \mathbf{y}) - Q(\boldsymbol{\theta}^0, \boldsymbol{\theta}^0; \mathbf{y}) + D(\boldsymbol{\theta} || \boldsymbol{\theta}^0). \quad (8.37)$$

By the positivity of the function $D(\boldsymbol{\theta} || \boldsymbol{\theta}^0)$, any value of $\boldsymbol{\theta}$ such that $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^0; \mathbf{y}) > Q(\boldsymbol{\theta}^0, \boldsymbol{\theta}^0; \mathbf{y})$ produces an increase in the likelihood. The general idea of the EM algorithm is to increase the likelihood at each iteration by choosing the

value of θ that maximizes $Q(\theta, \theta^0; \mathbf{y})$. One iteration of the algorithm comprises the following two steps:

$$\text{Expectation (E): } \text{calculation of } Q(\theta, \theta^k; \mathbf{y}), \quad (8.38)$$

$$\text{Maximization (M): } \theta^{k+1} = \arg \max_{\theta} Q(\theta, \theta^k; \mathbf{y}). \quad (8.39)$$

It can be shown that this algorithm has a rate of convergence sufficient to ensure convergence towards a local maximum of the likelihood [DEM 77]. The EM algorithm will only be of interest if the calculation and then the maximization of $Q(\theta, \theta^k; \mathbf{y})$ are simpler than the maximization of the likelihood $p(\mathbf{y}; \theta)$.

8.3.3. Application to estimation of the parameters of a GMRF

Using the EM approach to maximize the likelihood $p(\mathbf{y}; \theta, \vartheta)$ first requires us to choose the auxiliary variable \mathbf{X} . Here, the obvious choice is to take the indirectly observed field as the auxiliary variable. It remains to be seen whether, with such a choice, it is possible to evaluate the quantity Q defined in equation (8.34), and then to maximize it with respect to the hyperparameters. We recall that $p(\mathbf{y} | \mathbf{x})$ and $p(\mathbf{x})$ take the form given by (8.3)-(8.4). By putting these expressions into (8.36), we obtain [ZHA 94]:

$$Q(\theta, \vartheta) =$$

$$\mathbb{E}(-\Phi_{\theta}(\mathbf{x}) - \log Z(\theta) | \mathbf{y}; \theta^0, \vartheta^0) + \mathbb{E}(-\Psi_{\vartheta}(\mathbf{y} - \mathbf{H}\mathbf{x}) - \log Z(\vartheta) | \mathbf{y}; \theta^0, \vartheta^0)$$

where the dependence of Q on the parameters θ^0 and ϑ^0 and on \mathbf{y} has been omitted in order to lighten the notation. The expectations are taken relative to the probability density $p(\mathbf{x} | \mathbf{y}; \theta^0, \vartheta^0)$ and it is clear that evaluating them analytically, which would require specific calculation of integrals similar to that of equation (8.34), is not possible in general. However, because of the Markov nature of \mathbf{X} , the MCMC techniques described in the previous chapter are well suited to the numerical evaluation of such expectations. Using the stochastic sampling techniques presented in section 7.4.2, it is possible to draw N values $\{\mathbf{x}_n, 1 \leq n \leq N\}$ at random, distributed according to $p(\mathbf{x} | \mathbf{y}; \theta^0, \vartheta^0)$. The mathematical expectation of a quantity $g(\mathbf{X})$ thus verifies the relation:

$$\mathbb{E}(g(\mathbf{X}) | \mathbf{y}; \theta^0, \vartheta^0) = \lim_{N \rightarrow +\infty} \frac{1}{N} \sum_{n=1}^N g(\mathbf{x}_n)$$

and, for “sufficiently” large N :

$$\mathbb{E}(g(\mathbf{X}) | \mathbf{y}; \theta^0, \vartheta^0) \simeq \frac{1}{N} \sum_{n=1}^N g(\mathbf{x}_n).$$

Thus, the hyperparameters of a GMRF can be estimated by an EM algorithm according to the following procedure:

- 1) determine initial values $\boldsymbol{\theta}^0$ and $\boldsymbol{\vartheta}^0$;
- 2) at iteration k (current value of hyperparameters: $\boldsymbol{\theta}^k$ and $\boldsymbol{\vartheta}^k$):
 - draw N values $\{\mathbf{x}_n, 1 \leq n \leq N\}$ according to $p(\mathbf{x} | \mathbf{y}; \boldsymbol{\theta}^k, \boldsymbol{\vartheta}^k)$,
 - approximate $E(-\Phi_{\boldsymbol{\theta}}(\mathbf{x}) - \log Z(\boldsymbol{\theta}) | \mathbf{y}; \boldsymbol{\theta}^k, \boldsymbol{\vartheta}^k)$ by:

$$\tilde{Q}_{\boldsymbol{\theta}}^k \triangleq \frac{1}{N} \sum_{n=1}^N (-\Phi_{\boldsymbol{\theta}}(\mathbf{x}_n) - \log Z(\boldsymbol{\theta}))$$

and $E(-\Psi_{\boldsymbol{\vartheta}}(\mathbf{y} - \mathbf{H}\mathbf{x}) - \log Z(\boldsymbol{\vartheta}) | \mathbf{y}; \boldsymbol{\theta}^k, \boldsymbol{\vartheta}^k)$ by:

$$\tilde{Q}_{\boldsymbol{\vartheta}}^k \triangleq \frac{1}{N} \sum_{n=1}^N (-\Psi_{\boldsymbol{\vartheta}}(\mathbf{y} - \mathbf{H}\mathbf{x}_n) - \log Z(\boldsymbol{\vartheta}));$$

$$-\boldsymbol{\theta}^{k+1} = \arg \max_{\boldsymbol{\theta}} \tilde{Q}_{\boldsymbol{\theta}}^k \text{ et } \boldsymbol{\vartheta}^{k+1} = \arg \max_{\boldsymbol{\vartheta}} \tilde{Q}_{\boldsymbol{\vartheta}}^k;$$

- 3) stop iterations by a convergence criterion typically based upon the difference between successive estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\vartheta}$.

We must obviously ask ourselves about the difficulties involved in evaluating then maximizing the quantities $\tilde{Q}_{\boldsymbol{\theta}}$ and $\tilde{Q}_{\boldsymbol{\vartheta}}$. It should be stressed that, as far as $\tilde{Q}_{\boldsymbol{\theta}}$ is concerned, we find ourselves in a situation resembling that studied in the first part of this chapter, except that the direct observation of a realization of \mathbf{X} is replaced by the indirect observation of N realizations. This change has few practical consequences in as much as most of the difficulties come from the normalization term $Z(\boldsymbol{\theta})$, which is independent of the number of realizations. We will therefore need to use the techniques presented in section 8.2: evaluate $\tilde{Q}_{\boldsymbol{\theta}}$, then maximize it with respect to $\boldsymbol{\theta}$, *at each iteration, k , of the EM algorithm*. Thus, the resulting procedure is numerically very heavy. As for the parameter $\boldsymbol{\vartheta}$, the difficulty depends on the nature of the noise \mathbf{b} . If we keep to simple noise models, the evaluation of $\tilde{Q}_{\boldsymbol{\vartheta}}$ and its maximization with respect to $\boldsymbol{\vartheta}$ can be carried out relatively easily. For an identically distributed, independent, zero-mean, Gaussian noise, it is possible to obtain a closed-form expression for $Z(\boldsymbol{\vartheta})$, then $\boldsymbol{\vartheta}^{k+1}$ [BRU 00, RED 84].

8.3.4. EM algorithm and gradient

The EM algorithm is not the only procedure that enables recursive maximization of the likelihood. Often, when it is possible to use an EM algorithm, it is also possible to maximize the log-likelihood by a gradient technique. If we return to the framework defined in section 8.3.2; the gradient of the log-likelihood is defined by:

$$\nabla_{\boldsymbol{\theta}} \log p(\mathbf{y}; \boldsymbol{\theta}) = \frac{1}{p(\mathbf{y}; \boldsymbol{\theta})} \nabla_{\boldsymbol{\theta}} p(\mathbf{y}; \boldsymbol{\theta}). \quad (8.40)$$

Under weak conditions of regularity for the extended likelihood $p(\mathbf{y}, \mathbf{x}; \boldsymbol{\theta})$, we can differentiate under the integral sign and write:

$$\begin{aligned}
 \nabla_{\boldsymbol{\theta}} \log p(\mathbf{y}; \boldsymbol{\theta}) &= \frac{1}{p(\mathbf{y}; \boldsymbol{\theta})} \int \nabla_{\boldsymbol{\theta}} p(\mathbf{y}, \mathbf{x}; \boldsymbol{\theta}) d\mathbf{x} \\
 &= \frac{1}{p(\mathbf{y}; \boldsymbol{\theta})} \int \frac{\nabla_{\boldsymbol{\theta}} p(\mathbf{y}, \mathbf{x}; \boldsymbol{\theta})}{p(\mathbf{x} | \mathbf{y}; \boldsymbol{\theta})} p(\mathbf{x} | \mathbf{y}; \boldsymbol{\theta}) d\mathbf{x} \\
 &= \int \frac{\nabla_{\boldsymbol{\theta}} p(\mathbf{y}, \mathbf{x}; \boldsymbol{\theta})}{p(\mathbf{y}, \mathbf{x}; \boldsymbol{\theta})} p(\mathbf{x} | \mathbf{y}; \boldsymbol{\theta}) d\mathbf{x} \\
 &= \int \nabla_{\boldsymbol{\theta}} \log p(\mathbf{y}, \mathbf{x}; \boldsymbol{\theta}) p(\mathbf{x} | \mathbf{y}; \boldsymbol{\theta}) d\mathbf{x} \quad (8.41)
 \end{aligned}$$

$$= \mathbb{E}(\nabla_{\boldsymbol{\theta}} \log p(\mathbf{y}, \mathbf{x}; \boldsymbol{\theta}) | \mathbf{y}; \boldsymbol{\theta}). \quad (8.42)$$

Equations (8.41-8.42) show that the gradient of the log-likelihood takes an expression analogous to that for quantity Q , the term $\log p(\mathbf{y}, \mathbf{x}; \boldsymbol{\theta})$ being replaced by its gradient with respect to $\boldsymbol{\theta}$.

The above equations suggest that it is possible to estimate the parameters of a GMRF by a gradient method according to a scheme very close to that of an EM technique. If we go back to the notation of section 8.3.3, we see that the gradient of $\log p(\mathbf{y}, \mathbf{x}; \boldsymbol{\theta}, \boldsymbol{\vartheta})$ with respect to the hyperparameters can be broken down into two terms that can be expressed:

$$\nabla_{\boldsymbol{\theta}} \log p(\mathbf{y}, \mathbf{x}; \boldsymbol{\theta}, \boldsymbol{\vartheta}) = \nabla_{\boldsymbol{\theta}} \log p(\mathbf{x}; \boldsymbol{\theta}) = -\nabla_{\boldsymbol{\theta}} \Phi_{\boldsymbol{\theta}}(\mathbf{x}) - \frac{\nabla_{\boldsymbol{\theta}} Z(\boldsymbol{\theta})}{Z(\boldsymbol{\theta})}$$

$$\nabla_{\boldsymbol{\vartheta}} \log p(\mathbf{y}, \mathbf{x}; \boldsymbol{\theta}, \boldsymbol{\vartheta}) = \nabla_{\boldsymbol{\vartheta}} \log p(\mathbf{y} | \mathbf{x}; \boldsymbol{\vartheta}) = -\nabla_{\boldsymbol{\vartheta}} \Psi_{\boldsymbol{\vartheta}}(\mathbf{y} - \mathbf{H}\mathbf{x}) - \frac{\nabla_{\boldsymbol{\vartheta}} Z(\boldsymbol{\vartheta})}{Z(\boldsymbol{\vartheta})}.$$

The estimation of $\boldsymbol{\theta}$ and $\boldsymbol{\vartheta}$ by a gradient method can thus be implemented according to the following procedure:

- 1) determine initial values $\boldsymbol{\theta}^0$ and $\boldsymbol{\vartheta}^0$;
- 2) at iteration k (current values of hyperparameters $\boldsymbol{\theta}^k$ and $\boldsymbol{\vartheta}^k$):
 - draw N values $\{\mathbf{x}_n, 1 \leq n \leq N\}$ according to $p(\mathbf{X} | \mathbf{y}; \boldsymbol{\theta}^k, \boldsymbol{\vartheta}^k)$,
 - approximate $\mathbb{E}(\nabla_{\boldsymbol{\theta}} \log p(\mathbf{y}, \mathbf{x}; \boldsymbol{\theta}^k, \boldsymbol{\vartheta}^k) | \mathbf{y}; \boldsymbol{\theta}^k, \boldsymbol{\vartheta}^k)$ by:

$$G_{\boldsymbol{\theta}}^k \triangleq \frac{1}{N} \sum_{n=1}^N \left(-\nabla_{\boldsymbol{\theta}} \Phi_{\boldsymbol{\theta}^k}(\mathbf{x}_n) - \nabla_{\boldsymbol{\theta}} Z(\boldsymbol{\theta}^k) / Z(\boldsymbol{\theta}^k) \right)$$

and $\mathbb{E}(\nabla_{\boldsymbol{\vartheta}} \log p(\mathbf{y}, \mathbf{x}; \boldsymbol{\theta}^k, \boldsymbol{\vartheta}^k) | \mathbf{y}; \boldsymbol{\theta}^k, \boldsymbol{\vartheta}^k)$ by:

$$G_{\boldsymbol{\vartheta}}^k \triangleq \frac{1}{N} \sum_{n=1}^N \left(-\nabla_{\boldsymbol{\vartheta}} \Psi_{\boldsymbol{\vartheta}^k}(\mathbf{y} - \mathbf{H}\mathbf{x}_n) - \nabla_{\boldsymbol{\vartheta}} Z(\boldsymbol{\vartheta}^k) / Z(\boldsymbol{\vartheta}^k) \right),$$

$-\boldsymbol{\theta}^{k+1} = \boldsymbol{\theta}^k + \mu^k G_{\boldsymbol{\theta}}^k$ and $\boldsymbol{\vartheta}^{k+1} = \boldsymbol{\vartheta}^k + \mu^k G_{\boldsymbol{\vartheta}}^k$, where μ^k is the step of the gradient algorithm;

3) stop iterations by a convergence criterion typically based upon the difference between successive estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\vartheta}$.

If we compare the EM and gradient approaches, we see that, in both cases, the difficulty of evaluating the likelihood $p(\mathbf{y}; \boldsymbol{\theta})$ is avoided by making use of the extended likelihood $p(\mathbf{y}, \mathbf{x}; \boldsymbol{\theta})$. Also in both cases, the mathematical expectation is evaluated numerically by resorting to stochastic sampling techniques (see Chapter 7). As far as differences are concerned, the EM approach includes a maximization step while the gradient approach requires the evaluation of derivatives with respect to the hyperparameters. The difficulty involved in one or the other approach depends largely on the form of $p(\mathbf{y}, \mathbf{x}; \boldsymbol{\theta})$ and the choice between them is thus very dependent on the problem to be tackled.

Concerning questions of convergence, both approaches are local iterative techniques that can only be guaranteed to converge towards a local maximum of the likelihood, which, in general, is not convex. Both approaches have a linear convergence rate but the effective speed of convergence depends strongly on the conditioning of the estimation problem and the distance between the current value of the parameters and the solution. Here again, the choice of an approach must be adapted to the characteristics of the problem at hand. It is worth mentioning that, when both approaches are possible, we can envisage using one of them for the first iterations and the other for the rest, or even alternating between the two so as to maximize the convergence speed. An example of this type of technique is to be found in [RID 97].

8.3.5. Linear GMRF relative to hyperparameters

To illustrate what we have just seen, we give a few details below on the use of EM and gradient techniques when \mathbf{X} is a field whose energy is linear with respect to the parameters. $p(\mathbf{x})$ thus takes the form given in equation (8.6). We also assume that matrix \mathbf{H} is known and that the additive noise \mathbf{b} is Gaussian, zero-mean, independent, identically distributed with variance $\boldsymbol{\vartheta}$. We thus have:

$$p(\mathbf{b}) = (2\pi\boldsymbol{\vartheta})^{-P/2} \exp \left\{ -\|\mathbf{b}\|^2 / 2\boldsymbol{\vartheta} \right\} \quad (8.43)$$

where P is the size of \mathbf{b} .

Whatever the technique used, each iteration requires N realizations of \mathbf{X} to be drawn according to $p(\mathbf{x} | \mathbf{y}; \boldsymbol{\theta}^k, \boldsymbol{\vartheta}^k)$. This is a traditional operation, amply described in Chapter 7. Its computational volume essentially depends on the extent of the support of \mathbf{H} and on the complexity of the neighborhood system of field \mathbf{X} .

If we use an EM technique, we next have to evaluate then maximize the quantities \tilde{Q}_{θ}^k and \tilde{Q}_{ϑ}^k relative to θ and ϑ respectively. Note first of all that the term \tilde{Q}_{ϑ}^k does not present any difficulty since, according to equation (8.43), the normalization term $Z(\vartheta)$ is known explicitly and is equal to $(2\pi\vartheta)^{P/2}$. From this, we simply deduce that the equation for updating ϑ is:

$$\vartheta^{k+1} = \frac{1}{NP} \sum_{n=1}^N \|\mathbf{y} - \mathbf{H}\mathbf{x}_n\|^2. \quad (8.44)$$

The main difficulty occurs at the level of \tilde{Q}_{θ}^k and, more precisely, the normalization term $Z(\theta)$. As mentioned in section 8.3.3, the expression for \tilde{Q}_{θ}^k is very close to that for the log-likelihood of a directly observed field given in equation (8.8) and all the techniques for determining θ for a directly observed field presented in section 8.2 can, in principle, be used for the step of maximization with respect to θ of the EM algorithm. However, in practice, it is not possible to envisage using a cumbersome, iterative, optimization method *at each iteration* of an EM algorithm that is itself quite heavy. If the current values of the parameters θ^k and ϑ^k are not too far from the solution, we can consider using the importance sampling technique, taking θ^k as the reference parameter, or even keeping the same reference parameter over several iterations of the EM algorithm. However, to avoid being faced with an insurmountable volume of calculations, it is more often than not better to use the approximations described in section 8.2.3.

Now let us look at the gradient approach. Once the N realizations of \mathbf{X} have been drawn according to $p(\mathbf{x} | \mathbf{y}; \theta^k, \vartheta^k)$, it is necessary to evaluate quantities G_{θ}^k and G_{ϑ}^k . Here again, the term in ϑ causes little difficulty; it is easy to show that it can be calculated explicitly and has the expression:

$$G_{\vartheta}^k = \frac{1}{N} \sum_{n=1}^N \frac{\|\mathbf{y} - \mathbf{H}\mathbf{x}_n\|^2}{2\vartheta^2} - \frac{P}{2\vartheta} \quad (8.45)$$

As far as the evaluation of G_{θ}^k is concerned, the calculations are similar to those leading to equation (8.9) and each component of G_{θ}^k can be written:

$$(G_{\theta}^k)_i = -\frac{1}{N} \sum_{n=1}^N N_i(\mathbf{x}_n) + E_{\theta^k}(N_i(\mathbf{x})). \quad (8.46)$$

The evaluation of $E_{\theta^k}(N_i(\mathbf{x}))$ thus makes it necessary to sample \mathbf{X} at each iteration, this time according to the law $p(\mathbf{x}; \theta^k)$ ². To avoid this operation, we can also approach $p(\mathbf{x}; \theta)$ here as indicated in section 8.2.3, then calculate the gradient of the approximation obtained.

2. Note that the resulting procedure has the stochastic approximation algorithm proposed in [YOU 91] as a special case, for which the author gives the convergence conditions.

The elements presented above point out the difficulty of unsupervised problems and the complexity of the methods for solving them. In the example above, whatever the approach chosen, at least one, and possibly two, stochastic sampling steps are necessary at each iteration. The approximations that may lighten the implementation of the approach cannot avoid reducing its robustness and reducing the quality of the estimator. Great caution is therefore advised when employing such techniques.

8.3.6. Extensions and approximations

8.3.6.1. Generalized maximum likelihood

As mentioned above, estimating the hyperparameters of an indirectly observed quantity generally necessitates a great volume of calculations, leading to the interest of any approach intended to reduce this numerical complexity. *Generalized maximum likelihood* (GML) methods were developed precisely with a view to simplifying the calculations.

These techniques fall within the Bayesian framework used up to now. They consist of simultaneously estimating the quantity of interest \mathbf{X} and the hyperparameters $(\boldsymbol{\theta}, \boldsymbol{\vartheta})$ by maximizing the *generalized likelihood* $p(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}, \boldsymbol{\vartheta})$. We have:

$$(\mathbf{x}^*, \boldsymbol{\theta}^*, \boldsymbol{\vartheta}^*) = \arg \max_{\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\vartheta}} p(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}, \boldsymbol{\vartheta}) \quad (8.47)$$

where \mathbf{x}^* , $\boldsymbol{\theta}^*$ and $\boldsymbol{\vartheta}^*$ are the estimates of \mathbf{x} , $\boldsymbol{\theta}$ and $\boldsymbol{\vartheta}$ respectively in the GML sense.

Maximizing the generalized likelihood jointly with respect to all the variables is still difficult. For this reason, we generally use a sub-optimal iterative *block-maximization* procedure according to the following scheme:

$$\mathbf{x}^{k+1} = \arg \max_{\mathbf{x}} p(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}^k, \boldsymbol{\vartheta}^k) \quad (8.48)$$

$$(\boldsymbol{\theta}^{k+1}, \boldsymbol{\vartheta}^{k+1}) = \arg \max_{\boldsymbol{\theta}, \boldsymbol{\vartheta}} p(\mathbf{x}^{k+1}, \mathbf{y}; \boldsymbol{\theta}, \boldsymbol{\vartheta}) \quad (8.49)$$

where k is the iteration index. Decomposing the joint law in the form $p(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}, \boldsymbol{\vartheta}) = p(\mathbf{y} | \mathbf{x}; \boldsymbol{\vartheta})p(\mathbf{x}; \boldsymbol{\theta})$ decouples the estimation of the hyperparameters as follows:

$$\boldsymbol{\theta}^{k+1} = \arg \max_{\boldsymbol{\theta}} p(\mathbf{x}^{k+1}; \boldsymbol{\theta}), \quad (8.50)$$

$$\boldsymbol{\vartheta}^{k+1} = \arg \max_{\boldsymbol{\vartheta}} p(\mathbf{y} | \mathbf{x}^{k+1}; \boldsymbol{\vartheta}). \quad (8.51)$$

This approach is attractive in that the quantity of interest, \mathbf{x} , and the hyperparameters are estimated in a single framework by means of a single criterion: the generalized likelihood. Furthermore, the determination of \mathbf{x} by equation (8.48) corresponds

exactly to the estimation of \mathbf{x} with a MAP criterion, the hyperparameters taking the value $(\boldsymbol{\theta}^k, \boldsymbol{\vartheta}^k)$. In a similar way, the determination of the hyperparameters using equations (8.50-8.51) coincides exactly with their estimation by ML with the data observed directly, \mathbf{x} taking the value \mathbf{x}^{k+1} . At each iteration, we just need to solve a problem of the same type as the one covered in section 8.2, which is a remarkable simplification with respect to the EM or gradient methods described in sections 8.3.2 and 8.3.4.

Nevertheless, the use of this approach leads to difficulties essentially linked with the characteristics of the GML estimator and the relevance of the generalized likelihood criterion. Unlike the estimators of the maximum simple or posterior likelihood [DAC 86], the characteristics of the GML estimator are not well known and it is difficult to link the estimates obtained with the true values of the parameters, even in an asymptotic framework. Furthermore, it sometimes happens that the generalized likelihood has no upper bound in the natural domain of the parameters $(\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\vartheta})$, which can cause the iterative procedure defined by equations (8.48-8.49) to diverge [GAS 92].

From a practical point of view, several authors have reported obtaining interesting results with the GML approach, both in the processing of images modeled as GM-RFs [KHO 98, LAK 89] and in the processing of one-dimensional signals [CHA 96]. It is clear that the GML approach deserves attention because of its simplicity of implementation but the theoretical difficulties highlighted above should put us very much on our guard when interpreting the results obtained.

8.3.6.2. Full Bayesian approach

The general idea underlying what are known as *full Bayesian* approaches is to probabilize the hyperparameters of the problem by introducing *second rank priors*. Staying within the framework of the hypotheses adopted in section 8.3.5, this means that we probabilize quantities $\boldsymbol{\theta}$ and $\boldsymbol{\vartheta}$ for which the corresponding prior probabilities will be noted $p(\boldsymbol{\theta})$ and $p(\boldsymbol{\vartheta})$ respectively. The quantity of interest \mathbf{x} and hyperparameters $\boldsymbol{\theta}$ and $\boldsymbol{\vartheta}$ are estimated from the *full posterior* likelihood $p(\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\vartheta} | \mathbf{y})$. Note that this makes sense since $\boldsymbol{\theta}$ and $\boldsymbol{\vartheta}$ are now probabilized.

The question then arises as to how to use the full posterior likelihood. The most immediate approach would be to maximize it jointly with respect to \mathbf{x} , $\boldsymbol{\theta}$ and $\boldsymbol{\vartheta}$. However, this maximization can prove tricky, depending on the form given to the second ranking priors. Above all though, the characteristics of the corresponding estimator are as poorly known as those of the GML estimator described in the section 8.3.6.1. To convince ourselves of this, we just need to note that when $p(\boldsymbol{\theta})$ and $p(\boldsymbol{\vartheta})$ are densities that are uniform over the domain of the parameters, the full likelihood becomes identical to the generalized likelihood and maximizing the full likelihood then has the same limitations as those of the GML approach. This is why full Bayesian approaches are based on sampling quantities \mathbf{X} , $\boldsymbol{\theta}$ and $\boldsymbol{\vartheta}$ according to probability $p(\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\vartheta} | \mathbf{y})$, the sampling being performed in practice using the stochastic sampling methods presented in section 7.4.2. We thus construct a Markov chain $\{(\mathbf{X}^k, \boldsymbol{\theta}^k, \boldsymbol{\vartheta}^k)\}$ the elements

of which are chosen at random by means of the following *complete conditionals*:

$$\mathbf{x}^{k+1} : p(\mathbf{x} | \mathbf{y}, \boldsymbol{\theta}^k, \boldsymbol{\vartheta}^k) \propto p(\mathbf{y} | \mathbf{x}, \boldsymbol{\vartheta}^k) p(\mathbf{x} | \boldsymbol{\theta}^k), \quad (8.52)$$

$$\boldsymbol{\theta}^{k+1} : p(\boldsymbol{\theta} | \mathbf{y}, \mathbf{x}^{k+1}, \boldsymbol{\vartheta}^k) \propto p(\mathbf{x}^{k+1} | \boldsymbol{\theta}) p(\boldsymbol{\theta}), \quad (8.53)$$

$$\boldsymbol{\vartheta}^{k+1} : p(\boldsymbol{\vartheta} | \mathbf{y}, \mathbf{x}^{k+1}, \boldsymbol{\theta}^{k+1}) \propto p(\mathbf{y} | \mathbf{x}^{k+1}, \boldsymbol{\vartheta}) p(\boldsymbol{\vartheta}). \quad (8.54)$$

Sampling \mathbf{X} according to equation (8.52) has already been mentioned several times in this chapter and poses no particular problems. At this stage, two main choices remain to be made: (i) the choice of $p(\boldsymbol{\theta})$ and $p(\boldsymbol{\vartheta})$ that both makes the second rank priors pertinent and gives the possibility of evaluating full marginals, or sampling according to the latter; (ii) the choice of the estimator which will use the quantities produced by the sampler of equations (8.52) to (8.54).

Concerning point (ii), as we saw in section 8.3.3, the sampling methods allow the mathematical expectation to be calculated empirically by simple averaging once the Markov chain has reached its equilibrium state. This mathematical expectation is taken with respect to $p(\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\vartheta} | \mathbf{y})$, which, for $\boldsymbol{\theta}$ and $\boldsymbol{\vartheta}$, gives the estimators according to the posterior mean $\hat{\boldsymbol{\theta}} = E(\boldsymbol{\theta} | \mathbf{y})$ and $\hat{\boldsymbol{\vartheta}} = E(\boldsymbol{\vartheta} | \mathbf{y})$. Averaging the values of $\boldsymbol{\theta}$ and $\boldsymbol{\vartheta}$ provided by the sampler thus approaches the above estimates. We also observe that, although the stochastic sampler uses full conditional probabilities, the estimate obtained is marginalized with respect to all the quantities other than the observations \mathbf{y} .

For point (i), the way to proceed largely depends on the availability of precise prior information on the hyperparameters. If this information exists, the problem is close to that of traditional Bayesian estimation, the difficulty being to formalize the prior information in the form of a probability law and adjust the trade-off between the accuracy of the prior model and the complexity of the resulting estimator. However, in general we have only very little prior information on the hyperparameters, except perhaps support information. A simple, prudent solution is then to take a function that is uniform over the definition domain of the parameter as the second rank prior. A conjugate prior can also be chosen, i.e., such that the second rank prior and the corresponding full conditional belong to the same parameterized family of probabilities. This technique, which is fairly widespread in the literature [CHE 96, DUN 97, MCM 96], seems to correspond to esthetic rather than practical concerns as introducing a conjugate prior rarely simplifies the main problem, that of sampling the hyperparameters according to the full conditional. To do this, all the techniques conventionally used in statistics can be employed [PRE 92].

Once the hyperparameters have been estimated according to the above procedure, we can go on to estimate \mathbf{X} in a supervised framework by using either one of the methods presented in Chapter 7 or the same technique and the same estimator as for the hyperparameters [MCM 96]. By averaging the values of \mathbf{x} obtained during the iterations defined by equations (8.52) to (8.54), we obtain an approximate value of

$E(\mathbf{X} | \mathbf{y})$, so it is not necessary to use a different procedure to estimate \mathbf{X} . This type of approach is adopted in [CHE 96] for a one-dimensional problem and in [DUN 97], where a simplified image model involving a *Markov network* is also used. The choice of one approach or the other for estimating \mathbf{X} is essentially a question of finding an acceptable compromise between simplicity of use and quality of results. We should also stress that, by adopting the full Bayesian approach, it is possible to include linear degradation \mathbf{H} among the quantities to be estimated. This possibility has been successfully used in a one-dimensional framework [CHE 96] but the same does not appear to be true in 2D, perhaps because of the volume of the calculations and the convergence of the stochastic sampler.

8.4. Conclusion

In this chapter, we have presented the main tools for estimating the hyperparameters when solving multidimensional inverse problems with Markov regularization. One of the major characteristics of the techniques presented is their complexity or heaviness, from both the methodological and algorithmic points of view. It is thus legitimate to wonder if it is really useful or desirable to resort to such methods.

Of course, it is always possible to set the hyperparameters empirically. For a given application with a well specified type of data, the parameters can be fixed after a calibration phase. However, as a general rule, the results of the estimation of \mathbf{X} vary greatly depending on the hyperparameters and the determination of an acceptable range of values can be difficult, particularly for an operator who is not a specialist in signal or image processing. Furthermore, for a given problem, the range of acceptable values can vary according to the experimental conditions. The development of unsupervised methods is thus of real interest.

Here, we have only presented methods based on the use of a likelihood. There are other approaches for estimating hyperparameters, such as cross-validation [GOL 79] or the “L-curve” [HAN 92], but their application to the problem treated here seems neither very realistic from an algorithmic point of view nor well founded from a methodological one.

As already pointed out, the methods we have described are relatively complex, which may be seen as an indication of the difficulty of the problem of hyperparameter estimation. We can wonder if the quality of the results is proportional to the efforts needed to obtain them. From this point of view, the situation is contrasted. For 1D signals, synthetic or real, the likelihood-based techniques generally produce acceptable results. In 2D, in a simulation framework where the image to be processed is a realization of the GMRF that regularizes the inversion, the results are very encouraging (see, for example, [FOR 93]). On the other hand, for real data, the same approach shows a significant lack of robustness with respect to the image to be processed. The

parameters that lead to the best results are not always those obtained by maximum likelihood [DES 99]. This seems to suggest that GMRFs, although useful for regularizing the estimation of \mathbf{X} , do not model images coming from the real world very faithfully. This is why the search for more effective estimation methods and also for models that are more robust and more faithful to the solution sought is an interesting path to follow in the aim of improving the methods for estimating hyperparameters. It is also why, as things stand at the moment, it is necessary to be meticulous in making validations when such methods are employed.

8.5. Bibliography

- [BES 74] BESAG J. E., "Spatial interaction and the statistical analysis of lattice systems (with discussion)", *J. R. Statist. Soc. B*, vol. 36, num. 2, p. 192-236, 1974.
- [BRU 00] BRUZZONE L., FERNANDEZ PRIETO D., "Automatic analysis of the difference image for unsupervised change detection", *IEEE Trans. Geosci. Remote Sensing*, vol. 38, p. 1171-1182, 2000.
- [CEL 03] CELEUX G., FORBES F., PEYRARD N., "EM procedures using mean field-like approximations for Markov model-based image segmentation", *Pattern Recognition*, vol. 36, num. 1, p. 131-144, 2003.
- [CHA 96] CHAMPAGNAT F., GOUSSARD Y., IDIER J., "Unsupervised deconvolution of sparse spike trains using stochastic approximation", *IEEE Trans. Signal Processing*, vol. 44, num. 12, p. 2988-2998, Dec. 1996.
- [CHE 96] CHENG Q., CHEN R., LI T.-H., "Simultaneous wavelet estimation and deconvolution of reflection seismic signals", *IEEE Trans. Geosci. Remote Sensing*, vol. 34, p. 377-384, Mar. 1996.
- [DAC 86] DACUNHA-CASTELLE D., DUFLO M., *Probability and Statistics*, vol. 1, Springer Verlag, New York, NY, 1986.
- [DEM 77] DEMPSTER A. P., LAIRD N. M., RUBIN D. B., "Maximum likelihood from incomplete data via the EM algorithm", *J. R. Statist. Soc. B*, vol. 39, p. 1-38, 1977.
- [DES 99] DESCOMBES X., MORRIS R., ZERUBIA J., BERTHOD M., "Estimation of Markov random field prior parameters using Markov chain Monte Carlo maximum likelihood", *IEEE Trans. Image Processing*, vol. 8, p. 954-963, 1999.
- [DUN 97] DUNMUR A. P., TITTERINGTON D. M., "Computational Bayesian analysis of hidden Markov mesh models", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-19, num. 11, p. 1296-1300, Nov. 1997.
- [FOR 93] FORTIER N., DEMOMENT G., GOUSSARD Y., "GCV and ML methods of determining parameters in image restoration by regularization: fast computation in the spatial domain and experimental comparison", *J. Visual Comm. Image Repres.*, vol. 4, num. 2, p. 157-170, June 1993.
- [GAS 92] GASSIAT E., MONFRONT F., GOUSSARD Y., "On simultaneous signal estimation and parameter identification using a generalized likelihood approach", *IEEE Trans. Inf. Theory*, vol. 38, p. 157-162, Jan. 1992.

- [GEI 91] GEIGER D., GIROSI F., "Parallel and deterministic algorithms from MRF's: Surface reconstruction", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, num. 5, p. 401-412, May 1991.
- [GEY 92] GEYER C. J., THOMPSON E. A., "Constrained Monte Carlo maximum likelihood for dependent data", *J. R. Statist. Soc. B*, vol. 54, p. 657-699, 1992.
- [GOL 79] GOLUB G. H., HEATH M., WAHBA G., "Generalized cross-validation as a method for choosing a good ridge parameter", *Technometrics*, vol. 21, num. 2, p. 215-223, May 1979.
- [HAN 92] HANSEN P., "Analysis of discrete ill-posed problems by means of the L-curve", *SIAM Rev.*, vol. 34, p. 561-580, 1992.
- [KHO 98] KHOUMRI M., BLANC-FERAUD L., ZERUBIA J., "Unsupervised deconvolution of satellite images", in *Proc. IEEE ICIP*, vol. 2, Chicago, IL, p. 84-87, 1998.
- [LAK 89] LAKSHMANAN S., DERIN H., "Simultaneous parameter estimation and segmentation of Gibbs random fields using simulated annealing", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-11, num. 8, p. 799-813, Aug. 1989.
- [MCM 96] MCMILLAN N. J., BERLINER L. M., "Hierarchical image reconstruction using Markov random fields", in *Bayesian Statistics 5*, Spain, Fifth Valencia Int. Meeting on Bayesian Statistics, June 1996.
- [PRE 92] PRESS W. H., TEUKOLSKY S. A., VETTERLING W. T., FLANNERY B. P., *Numerical Recipes in C, the Art of Scientific Computing*, Cambridge University Press, New York, 2nd edition, 1992.
- [RED 84] REDNER R. A., WALKER H. F., "Mixture densities, maximum likelihood and the EM algorithm", *SIAM Rev.*, vol. 26, num. 2, p. 195-239, Apr. 1984.
- [RID 97] RIDOLFI A., Maximum Likelihood Estimation of Hidden Markov Model Parameters, with Application to Medical Image Segmentation, Thesis, Politecnico di Milano, Facoltà di Ingegneria, Milan, Italy, 1997.
- [YOU 88] YOUNES L., "Estimation and annealing for Gibbsian fields", *Ann. Inst. Henri Poincaré*, vol. 24, num. 2, p. 269-294, Feb. 1988.
- [YOU 91] YOUNES L., "Maximum likelihood estimation for Gibbs fields", in POSSOLO A. (Ed.), *Spatial Statistics and Imaging: Proceedings of an AMS-IMS-SIAM Joint Conference*, Lecture Notes – Monograph Series, Hayward, Institute of Mathematical Statistics, 1991.
- [ZER 93] ZERUBIA J., CHELLAPPA R., "Mean field annealing using compound GMRF for edge detection and image estimation", *IEEE Transactions on Neural Networks*, vol. TNN-4, p. 703-709, 1993.
- [ZHA 93] ZHANG J., "The mean field theory in EM procedures for blind Markov random field image restoration", *IEEE Trans. Image Processing*, vol. 2, num. 1, p. 27-40, Jan. 1993.
- [ZHA 94] ZHANG J., MODESTINO J. W., LANGAN D. A., "Maximum-likelihood parameter estimation for unsupervised stochastic model-based image segmentation", *IEEE Trans. Image Processing*, vol. 3, p. 404-420, 1994.

PART IV

Some Applications

Chapter 9

Deconvolution Applied to Ultrasonic Non-destructive Evaluation

9.1. Introduction

In non-destructive evaluation (NDE) using ultrasound, a wave is emitted at the outer surface of the part under inspection and propagates through the material. At each discontinuity the incident beam meets, a wave is reflected and the receiver located on the surface of the part thus receives a succession of reflected waves. Thus, the received signal results from the interaction of the wave with the material and gives information on the discontinuities in the medium. The operator's goal is to find out, from the received signal, whether the part has defects and, if so, to determine their characteristics. In concrete terms, a discontinuity shows up in the measurements as an echo having a duration equivalent to the period of the incident wave. However, the wavelength is of the same order as or even longer than the distance between two discontinuities and the presence of two defects close together gives a single common echo. Deconvolution helps with the interpretation of the signals by getting around the lack of resolution of the measurements.

We will start by presenting the difficulties involved in interpreting the measurements by means of an example of a weld evaluation using ultrasound. Then, the direct model will be introduced in the form of a convolution: the data is taken to be the result of the convolution between the discontinuities and a convolution kernel formed from the transmitted wave. When the waveform is known, it is possible to infer the discontinuities by applying the approaches already proposed in Chapter 5 concerning

spike train deconvolution. In reality, the incident wave is very often unknown and estimating the discontinuities is similar to solving a *blind deconvolution* problem for which the discontinuities and the kernel are unknown. The heart of this chapter is therefore devoted to the blind deconvolution techniques put forward for various types of applications. Finally, in the case of NDE, the contributions made by deconvolution are illustrated on the weld evaluation example.

9.2. Example of evaluation and difficulties of interpretation

Checking a weld was chosen as the example because the data concerned is notoriously difficult to interpret. The characteristics of welding and the evaluation principle will be recalled first. Then the conclusions drawn simply from the exploitation of the raw measurements will be presented. Finally, we will define the objectives fixed for the processing to help with the interpretation of the measurements.

The data used in this chapter were provided by EDF-R&D, the R&D division of the French power utility EDF. The expert NDE knowledge required to write the chapter was generously contributed by Daniel Villard, EDF-R&D, to whom we are most grateful.

9.2.1. Description of the part to be inspected

The part to be inspected is a stainless steel/stainless steel weld with a notch 3 mm high and 0.2 mm thick made by electro-erosion at the level of the welding cord, parallel to the weld axis, see Figure 9.1. This notch is representative of a crack; it constitutes a planar defect, so called by opposition to “volume” defects, which are thicker.

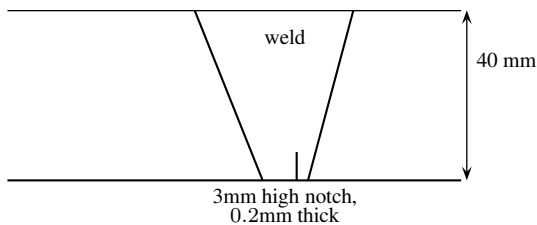


Figure 9.1. Cross-section of the part to be inspected

9.2.2. Evaluation principle

The evaluation takes place at the upper surface of the block (Figure 9.2). Both the emitter and the receiver are embedded in the sensor, which scans the surface of the

part, perpendicularly to the welding axis, emitting 2 MHz longitudinal waves inclined at 60° to the normal of the surface of the block.

Two sorts of echoes are generated in response to the incident wave: a corner echo backscattered by reflection at the bottom of the notch and a diffraction echo created at the top of the notch and scattered in all directions (see Figure 9.2). The latter is only generated when a diffracting point is present and would be absent in the case of a volume defect. It is thus characteristic of a planar defect. For some positions of the sensor, these two echoes may appear simultaneously because of the width of the beam. The evaluation should thus (i) detect a planar defect characterized by the presence of a diffraction echo in addition to the corner echo and (ii) determine the height of the defect from the time difference between the two echoes.

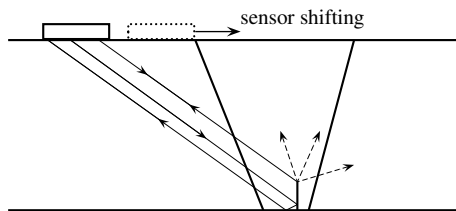


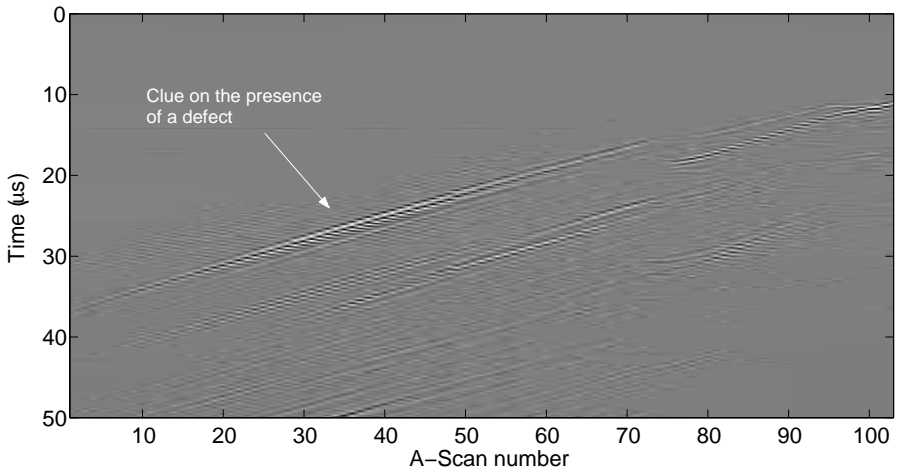
Figure 9.2. Corner and diffraction echoes

9.2.3. Evaluation results and interpretation

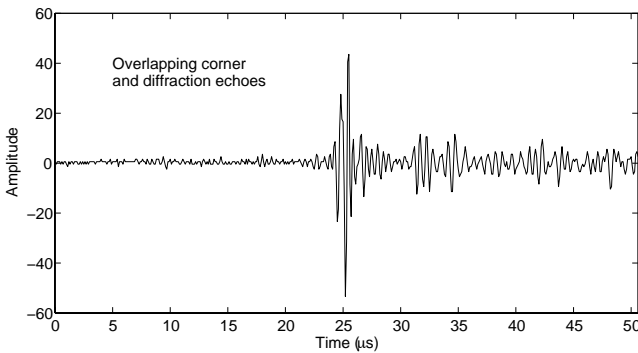
The signal recorded for a given sensor position is called an *A-scan*. The image produced by expressing the A-scans obtained for successive sensor positions as grey levels is called a *B-scan*. The B-scan corresponding to the evaluation described above is shown in Figure 9.3a. The presence of a high amplitude echo can be distinguished over a large part of it. This echo appears to be slanted, which signifies that the time to cover the path decreased as the sensor moved.

Using this echo, the operator can detect the presence of a defect but cannot find its characteristics. It is very difficult to distinguish between corner and diffraction echoes. In A-scan number 40, which is given as an example (see Figure 9.3b), the diffraction and corner echoes merge into one another: the top and bottom of the notch are very close together and result in a common echo.

In the evaluation case, the raw data are very difficult to use; it is impossible to state that the defect is planar and, *a fortiori*, to determine its height. Interpreting the data is thus a problem.



(a) B-scan



(b) A-scan # 40

Figure 9.3. Raw data corresponding to the system of Figure 9.2

9.2.4. Help with interpretation by restoration of discontinuities

It can be seen from the example above that one of the main difficulties in interpreting ultrasonic test data stems from the lack of resolution of the measurements. The presence of several reflectors very close together results in a common echo. Thus, processing that could bring out the discontinuities masked in the data would be of great help for the interpretation. In the above mentioned example, this would enable the two reflectors associated with the angle and the top of the notch to be separated and the height of the defect to be deduced. In what follows, we will therefore study processing methods for restoring discontinuities.

9.3. Definition of direct convolution model

As the aim of the processing is to restore the discontinuities, the part is represented by a function that is characteristic of what we are looking for. The reflectivity is chosen for this purpose as this function is related to the spatial derivative of the acoustic impedance: it is non-zero at a discontinuity (the impedance changes) and is zero everywhere else (the impedance is constant). Starting with this representation, the reflectivity is connected to the measurements by a convolution model: each A-scan is interpreted as the result of convolution between the reflectivity function and a kernel representing the wavelet that propagates in the object. A rigorous justification of this model would necessitate a series of restrictive hypotheses on the properties of the material under study: the velocity, C , of the wave in the medium must be constant, so the material must have only small inhomogeneities. Finally, the propagation is assumed to be one-dimensional and the measurements are considered A-scan by A-scan. We will use the notations $r(d)$ for the reflectivity at depth d , $h(t)$ for the convolution kernel, and $y(t)$ for the measurement. The observation equation can thus be written:

$$y(t) = \int h(s) r(C(t - s)) ds.$$

The numerical framework of the processing performed on the data leads us to choose a discrete representation of the model. Thus, a sampled echogram \mathbf{y} is assumed to be the result of a one-dimensional, discrete convolution between wavelet \mathbf{h} emitted by the sensor and the sequence of reflectivities \mathbf{r} situated along the path followed by the beam. To take account of measurement and model errors, this convolution result is perturbed by additive noise \mathbf{b} , assumed to be independent of \mathbf{r} , white and Gaussian. The direct model for an echogram can thus be written in the form:

$$\mathbf{y} = \mathbf{H}\mathbf{r} + \mathbf{b} \quad (9.1)$$

where \mathbf{H} is a Toeplitz matrix composed of elements of \mathbf{h} .

This model is rather rough. In particular, it does not take account of the attenuation and deformation of the wave during propagation. Also, the width of the ultrasound beam is not modeled: this would make it necessary to use a two- or even three-dimensional kernel [SAI 85]. Finally, reflections connected with the structure of the material are not modeled although they may be non-negligible for some types of steel. Rigorous modeling of ultrasonic wave propagation in a material is a very complex task and use of the simplified model in the form of a convolution is rarely seen in the NDE community. However, the direct model is chosen for processing and, from this point of view, the convolution model provides a good compromise between representativeness of the physical phenomena and the possibilities for efficient exploitation of the processing.

9.4. Blind deconvolution

In as far as the insonified medium is made up of homogeneous zones, the reflectivities can be modeled by a succession of spikes where each spike corresponds to a discontinuity. In light of this prior information on the reflectivity and the direct convolution model (9.1), estimating the reflectivities is similar to solving a spike train deconvolution problem for which the input and the kernel are the reflectivity and the wavelet respectively. In the case where the wave is known, the spike train deconvolution techniques presented in Chapter 5 should enable the discontinuities to be recovered. In fact, the incident wave is often unknown and it is necessary to solve a blind deconvolution problem where both the input and the convolution kernel are unknown.

This section will first give an overview of the various approaches to be found in the literature for solving blind deconvolution problems. An extension of the L2Hy and BG spike deconvolution methods will then be presented. The resulting methods make it possible to take account of phase deformations occurring in the wavelet as it propagates and corresponding to an enrichment of the direct model, introduced via bivariate reflectivity. They are christened DL2Hy and DBG, the “D” referring to the double reflectivity. Initially put forward to take phase rotations into account in ultrasound imaging, these methods also palliate possible phase imperfections in the wavelet introduced for the deconvolution. Finally, an original blind deconvolution approach will be developed using sequential exploitation of a sub-optimal wavelet estimation technique and the DL2Hy and DBG deconvolution methods.

9.4.1. Overview of approaches for blind deconvolution

This subsection presents the principal approaches suggested for solving a blind deconvolution problem. The presentation centers around the methodology and thus concerns a broader field of application than ultrasound imaging. We recall that, in the specific case of deconvolution in NDE, the input and kernel are the reflectivity and the wavelet respectively.

The most widespread approach to be found in the literature gets around the lack of knowledge of the kernel by finding an estimator of the input that does not depend explicitly on the kernel. This first path is followed by the predictive deconvolution, minimum entropy deconvolution, and “multipulse” techniques. A second approach estimates the kernel then applies a spike deconvolution technique. Finally, some approaches set out to estimate the kernel and input jointly.

9.4.1.1. Predictive deconvolution

The first work on blind deconvolution was carried out for geophysical applications. Among these works, the 1950s studies by Robinson on predictive deconvolution [ROB 67] have conditioned much research activity up to the present day. In this

approach, the convolution model is implicitly noise-free; the measurement is the result of input filtering using a recursive filter according to:

$$y_n = \sum_{\ell=1}^q a_{\ell} y_{n-\ell} + r_n, \quad (9.2)$$

which comes down to considering the kernel as the IR of a stable, causal, autoregressive filter. If the input is also assumed to be white, predictive deconvolution first estimates the prediction coefficients, and then identifies the input to the prediction error. The estimator of the prediction vector $\mathbf{a} = [a_1, \dots, a_q]^T$ is obtained by minimizing a least squares criterion formed from equation (9.2), for an ad hoc boundary hypothesis such that $y_n = 0, \forall n < 1$:

$$\hat{\mathbf{a}} = \arg \min_{\mathbf{a}} \sum_n \left(y_n - \sum_{\ell=1}^q a_{\ell} y_{n-\ell} \right)^2. \quad (9.3)$$

The input is then given by the prediction error of the minimum norm as:

$$\hat{r}_n = y_n - \sum_{\ell=1}^q \hat{a}_{\ell} y_{n-\ell}, \quad (9.4)$$

and any reference to the coefficients of the IR has thus disappeared from the definition of the estimated input.

This approach only makes use of the second order characteristics of the measurements. In addition, the estimation of vector \mathbf{a} according to (9.3) is equivalent to solving an AR spectral analysis problem: it gives a recursive filter having an amplitude spectrum which matches that of the measurements. For some boundary hypotheses, the filter obtained is causal and stable. In the literature, this filter is said to be *minimum phase* [ORF 85]. Other stable (but not causal) AR filters with the same amplitude spectrum could be identified by taking the poles outside the unit circle.

This way of looking at deconvolution where, on the one hand, only the spectral characteristics of the kernel are used and, on the other, the reflectivity is identified with the error of estimation of the coefficients of a filter, is still widely shared today.

The restriction to only second order characteristics is often put forward to explain the limited performance of predictive deconvolution. Many studies have been conducted to exploit higher order (higher than two) statistics of the measurements and thus design estimators sensitive to the phase. However, these approaches require a large number of measurement points, which severely limits their field of application [LAZ 93].

Finally, the choice of a criterion for the quadratic prediction error comes down to implicitly supposing that the input is Gaussian. This shows us that, to the great

displeasure of many supporters of predictive deconvolution, this approach introduces prior information on the input. Thus, application of an L_p ($p < 2$) norm on the prediction error recovers spikier inputs than in the case of conventional predictive deconvolution without, however, reaching the performance of L2LP deconvolution. In fact, not taking observation noise and modeling errors into consideration constitutes the weak point of approaches connected with predictive deconvolution.

9.4.1.2. *Minimum entropy deconvolution*

In the seismic-reflection context, Wiggins [WIG 78] proposes finding the input by linear filtering of the measurements, in the form (9.4), but his method for obtaining the coefficients of the filter differs from predictive deconvolution. Wiggins defines the “varimax” norm to measure the disorder in a signal. Assuming the input to be made up of impulses, he finds the filter coefficients that minimize the disorder of the input. Associating the notion of disorder with entropy, the author calls his approach *minimum entropy deconvolution*. Here again, any reference to the kernel has disappeared; implicitly, the latter remains the IR of a causal recursive filter. The empirical approach in fact consists of finding the “inverse filter” of the convolution kernel by cleverly taking advantage of prior information on the input. Nevertheless, its performance remains limited by the implicit hypothesis of filter causality and the failure to take observation noise and modeling errors into account.

9.4.1.3. *Deconvolution by “multipulse” technique*

This approach was put forward in [ATA 82] as a speech encoding method of the LPC (*linear predictive coding*) type but its principle can be adapted to pulse train deconvolution [COO 90]. It is close to that of the minimum entropy method: model the input in AR form, as in predictive deconvolution, and introduce the pulse character of the input. Here, this character is imposed in an ad hoc way by locating the maximum prediction error peaks and iterating (or not) a procedure alternating detection of the peaks, re-estimation of the amplitudes of the previously detected peaks and re-estimation of the coefficients of the filter. Overall, this method suffers from the same restrictions as Wiggins’ method.

9.4.1.4. *Sequential estimation: estimation of the kernel, then the input*

Using a deconvolution method suited to the case of a known IR, the solution of a blind deconvolution problem can be carried out in two stages: estimation of the kernel, then estimation of the input with the help of available deconvolution techniques. Historically, this way of proceeding did not find favor because efficient deconvolution tools were not available. However, now that more efficient deconvolution methods have been developed, this approach seems natural to us. Relative to conventional deconvolution, the extra difficulty lies in estimating the kernel from the measurements when the input is unknown. In fact, little work has been done in this direction. The main method for estimating the kernel presented here was initially proposed by Vivet [VIV 89] for NDE, but it can be extended to other applications.

With the hypothesis of the input being white and there being no noise, the amplitude spectrum of the measurements corresponds to that of the kernel. The proposed method thus consists of obtaining a wavelet having an amplitude spectrum identical to that of the measurements. As recalled in section 9.4.1.1, the estimator (9.3) provides precisely the coefficients of such a filter in recursive form. In the absence of knowledge on the phase of the wavelet, the kernel selected is the truncated IR of the minimum phase filter. This IR is calculated by recursively applying

$$h_n = \sum_{\ell=1}^q \hat{a}_\ell h_{n-\ell} + \delta_n, \quad (9.5)$$

with $h_n = 0, \forall n < 0$ as initialization. With reference to the name of the filter, this kernel is also called “minimum phase”.

The main limitation of this approach is, of course, the arbitrary choice of the phase of the kernel. This limitation can be compensated by the extensions of the deconvolution techniques that will be presented below.

9.4.1.5. Joint estimation of kernel and input

In the framework of a Bayesian approach, the way of working proposed for spike train deconvolution can be extended perfectly naturally by using a joint density that is a function of the kernel and the input (see Chapter 3). For this purpose, a prior model is defined for the kernel in the same way as for the prior model of the input. Typically, in the case of geophysical or NDE applications, the prior model of the kernel can be a correlated Gaussian process corresponding to smoothness prior information. Various paths can then be explored to exploit the joint density.

We will look at the maximum *a posteriori* estimator first, using Bernoulli-Gaussian (BG) prior models for the input [GOU 89] at the beginning, then convex L_p models [GAU 96, GAU 97]. The calculation requires the minimization of criteria whose properties are poorly known¹. It is carried out by alternately minimizing the criterion according to the kernel and the input. In both cases, the results of the processing on simulated data appear convincing. However, in the case of NDE applications [GAU 96], this approach has proved to be less efficient than the one proposed in the subsections that follow.

Finally, more recently, the estimator of the posterior mean [CHE 96] has been proposed for a BG-type input. To this end, a series of samples corresponding to realizations according to the posterior joint density are generated using a Gibbs sampler (see Table 7.1). Then the estimator of the posterior mean is approached by empirical means on the available samples.

1. Here we can again voice the reserves mentioned in Chapter 3 concerning the generalized maximum likelihood — see section 3.5.

9.4.2. DL2Hy/DBG deconvolution

In the direct model proposed in section 9.3, the wave is assumed not to deform as it propagates. However, this hypothesis is not very realistic in ultrasound imaging. In particular, the form and/or the polarity of the echoes is reputed to vary with the type of defect encountered. These phenomena probably correspond to changes in the phase of the wave but remain poorly understood. The direct model can be improved by taking into account possible wave deformations, modeled by phase “rotations” of the wavelet at the level of each reflectivity. Modeling the rotations leads to the definition of a double reflectivity, with which some prior information is associated. This approach was first put forward in Bernoulli-Gaussian deconvolution [CHA 93], then extended to convex prior models [GAU 97]. It was initially developed for ultrasonic NDE but would be suitable for other applications, in particular when the wave undergoes phase deformations.

9.4.2.1. Improved direct model

Let h be a function of a real variable. g , its Hilbert transform (HT), is defined as the function that has the same amplitude spectrum as h but whose phase is shifted by $\pi/2$. Function h_θ , such that $h_\theta = h \cos \theta + g \sin \theta$, has a phase that is shifted by θ with respect to that of h ; but, h_θ and h have the same amplitude spectrum. Thus, a signal with a phase that has been shifted by a constant can be expressed as a linear combination of the initial signal and its HT. By making use of the linearity of the convolution, the model of the phase rotations for each reflectivity can be transferred on to a split reflectivity sequence (r, s) , where r and s are convolved by the wavelet and its HT respectively. The direct model can thus be expressed as $\mathbf{y} = \mathbf{H}\mathbf{r} + \mathbf{G}\mathbf{s} + \mathbf{b}$, where \mathbf{y} is the observed trace, \mathbf{H} the convolution matrix corresponding to the known wavelet, \mathbf{G} the convolution matrix associated with the HT of the wavelet, and \mathbf{b} white, zero-mean, Gaussian noise independent of \mathbf{r} and \mathbf{s} . This model is purely an enrichment of the direct model with respect to the simple convolution model: the double reflectivity is simply introduced to help model the wavelet phase rotation phenomena.

9.4.2.2. Prior information on double reflectivity

The deconvolution process inverts this direct model to estimate an $(\hat{\mathbf{r}}, \hat{\mathbf{s}})$ pair. As before, the inversion requires prior information to be brought in. The reflectivity is taken to be spiky: the components of \mathbf{r} and \mathbf{s} are, *a priori*, zero except for where discontinuities occur. Similarly, the reflectivity at one point is not, *a priori*, connected with the reflectivity at other points along the path: for $i \neq j$, the pairs (r_i, s_i) and (r_j, s_j) are independent.

9.4.2.3. Double Bernoulli-Gaussian (DBG) deconvolution

In order to exploit these prior hypotheses, [CHA 93] introduces the triplet $(\mathbf{q}, \mathbf{r}, \mathbf{s})$ composed of independent variables q_i, r_i, s_i where:

- \mathbf{q} is a Bernoulli process indicating the presence/absence of a discontinuity;
- conditionally on q_i , the pair (r_i, s_i) is Gaussian, zero-mean, and has a covariance matrix $r_x q_i \mathbf{I}$, with r_x being the prior reflectivity variance.

Then, as in “simple” Bernoulli-Gaussian deconvolution (see Chapter 5), [CHA 93] proposes a sequential estimation approach based first on determining vector $\hat{\mathbf{q}}$ from the measurements, then jointly estimating $(\hat{\mathbf{r}}, \hat{\mathbf{s}})$ conditionally on $\hat{\mathbf{q}}$. The estimators can be obtained by tweaking the algorithms described in Chapter 5. The deconvolution method thus obtained is called “double” Bernoulli-Gaussian (DBG) deconvolution.

9.4.2.4. Double hyperbolic (DL2Hy) deconvolution

[GAU 97] has adapted the principle of “double” deconvolution to convex prior models, in the following form:

$$(\hat{\mathbf{r}}, \hat{\mathbf{s}}) = \arg \min_{\mathbf{r}, \mathbf{s}} \left(\|\mathbf{y} - \mathbf{H}\mathbf{r} + \mathbf{G}\mathbf{s}\|^2 + \lambda \sum_i \rho(r_i, s_i) \right), \quad \lambda \geq 0,$$

where λ is a regularization parameter and the bivariate function ρ allows the prior information on the double reflectivity sequence to be included. The desire to obtain a strictly convex, differentiable criterion naturally leads to a choice of a function ρ that is also strictly convex and differentiable. Moreover, considering the prior information, the monovariate restriction of this bivariate function must behave in a way very similar to that seen in the simple deconvolution case. The transposition of the hyperbolic case to the bivariate case leads us to look for a function $\rho(u, v)$ that is quadratic for small values of u and v and conic for large values of one or other of the variables. The “bivariate hyperbolic” function $\rho(u, v) = \sqrt{T^2 + u^2 + v^2}$ has these properties (see Figure 9.4). The parameter T allows the quadratic zone to be dilated or reduced, which introduces more or less strong interdependence between u and v : the smaller T is, the more u and v are linked (at the extreme when T tends to infinity, u and v are independent).

The associated estimator is finally given by:

$$(\hat{\mathbf{r}}, \hat{\mathbf{s}}) = \arg \min_{\mathbf{r}, \mathbf{s}} \left(\|\mathbf{y} - \mathbf{H}\mathbf{r} + \mathbf{G}\mathbf{s}\|^2 + \lambda \sum_i \sqrt{T^2 + r_i^2 + s_i^2} \right).$$

As the criterion to be minimized is jointly convex in \mathbf{r} and \mathbf{s} , the solution can be obtained by a descent algorithm. Considering the regularization function chosen, the method associated with this estimator is called double hyperbolic (DL2Hy) deconvolution.

9.4.2.5. Behavior of DL2Hy/DBG deconvolution methods

The behavior of these methods is mainly the result of the modeling of the prior information. Thus, the discussion of the behavior of L2Hy/BG methods can be extended to DL2Hy/DBG cases. Since the DBG deconvolution includes a decision step,

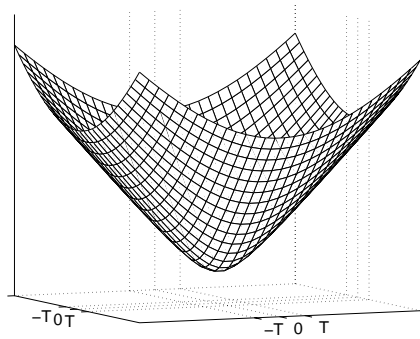


Figure 9.4. Graphic representation of a bivariate hyperbolic function

it gives more marked peaks than DL2Hy deconvolution. From another point of view, the absence of a decision process in the DL2Hy case makes DL2Hy deconvolution more robust with respect to variations in the data and the hyperparameters.

9.4.3. Blind DL2Hy/DBG deconvolution

On the basis of the elements above, a sequential approach to blind deconvolution can be proposed using the two following steps [GAU 97]. First of all, the convolution kernel is estimated by AR estimation, as in section 9.4.1. Then the input is estimated by conventional DL2Hy or DBG deconvolution. In this case, the phase of the kernel is chosen arbitrarily, but the DL2Hy/DBG deconvolution makes it possible to adapt to phase rotations of the kernel and thus, in particular, to compensate for a phase shift in the wave introduced into the deconvolution process.

From now on, this approach will be referred to simply as “blind DL2Hy/DBG deconvolution”. Its application should not be restricted to cases where a phase rotation is foreseen; the DL2Hy/DBG deconvolution no doubt gives the method a degree of adaptability sufficient to compensate for kernel deformations more general than just phase rotations. Thus, using DL2Hy/DBG deconvolution allows us to take account of any phase changes of the kernel and also to make up for imperfections of the kernel estimation technique.

9.5. Processing real data

The performance of the deconvolution will be illustrated using the weld evaluation example of section 9.2. In this data, the top and bottom of the notch appear in a

common echo. The aim of the processing is to bring out the reflectors associated with the extremities of the notch so as to characterize the presence of a planar defect and estimate the height of the notch.

For this work, the wave emitted by the ultrasonic sensor was measured experimentally, but such measurements are not always available. To carry out blind deconvolution of the raw B-scan of Figure 9.3a, a “minimum phase” wavelet of 50 samples was estimated for each A-scan from eight AR coefficients. Figure 9.5 shows the estimated wavelet for A-scan number 40 and the measured wave. Comparing them reveals the limits of the wave estimation technique: in particular, the energy of the wavelet is concentrated in the first instants (whence the term “minimum phase”), which is not the case for the measured wave.

All the processing performed is one-dimensional: a deconvolved B-scan is simply the result of the juxtaposition of the separately processed A-scans. First, the realistic case of blind deconvolution will be tackled. Then, the reflectivities restored using the measured wave for the deconvolution will be presented. Finally, elements for comparing the DL2Hy and DBG approaches will be provided.

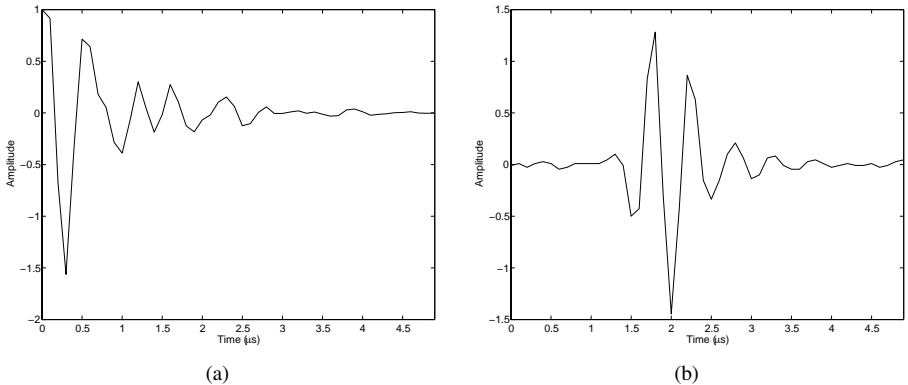


Figure 9.5. (a) *Estimated wavelet for A-scan # 40, using AR modeling, $q = 8$; (b) measured wavelet*

9.5.1. Processing by blind deconvolution

The result of the predictive deconvolution of Figure 9.6 illustrates the failure of this approach, which does not bring out the two distinct reflectivities corresponding to the top and bottom of the notch at all. In the rest of this section, the discussion will concern the behavior of the L2Hy/BG and DL2Hy/DBG approaches implemented by using an estimated wave for each A-scan.

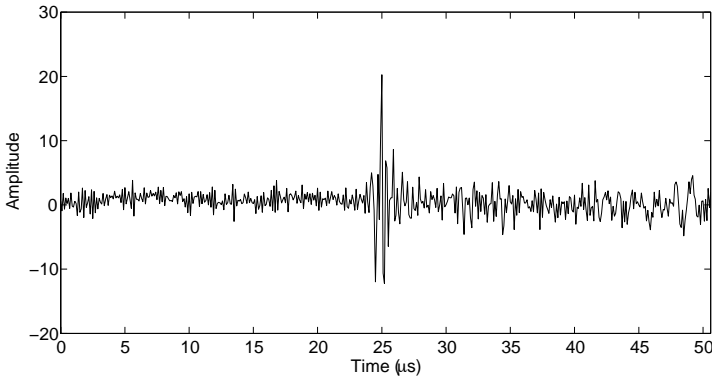


Figure 9.6. Predictive deconvolution of the A-scan of Figure 9.3b for $q = 8$

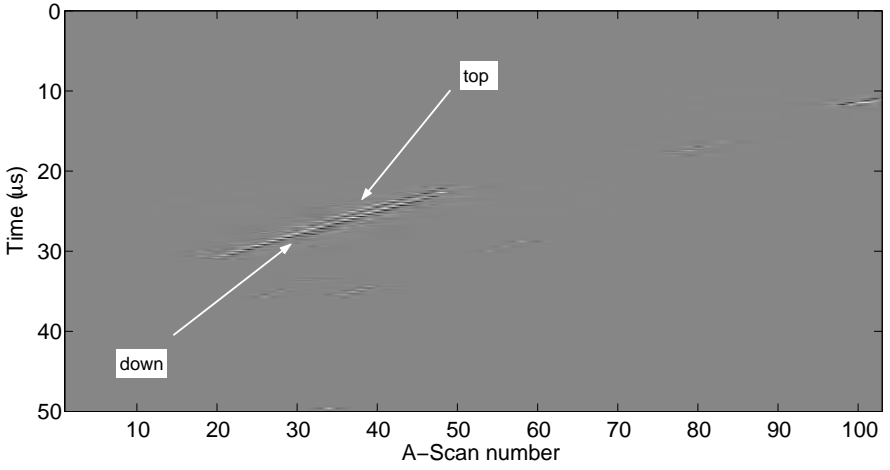
The L2Hy deconvolution brings out peaks connected with the top and the bottom of the notch, expressing the presence of this planar defect in the weld, Figures 9.7a and b. The result of BG deconvolution is comparable (see Figure 9.7c). However, the assistance provided by these methods is perturbed by the splitting of the reflectivities, probably due to poor estimation of the phase of the wavelet.

The use of DL2Hy deconvolution allows us to place ourselves directly in the more favorable framework described in section 9.4.3 and, particularly, to improve the previous results: the two reflectivities connected with the extremities of the notch appear even more clearly than in the L2Hy deconvolution case (compare Figures 9.8a and 9.7a); the splitting of the reflectivity has disappeared (Figure 9.7b relative to Figure 9.8b). Finally, DBG deconvolution gives a comparable result here too (Figure 9.7c relative to Figure 9.8c).

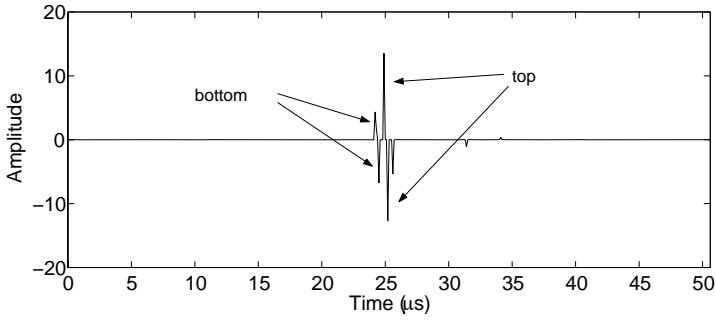
Blind deconvolution considerably improves the resolution of ultrasonic evaluation: the reflectivities initially drowned in a common echo appear separately after deconvolution. Exploiting the time difference between the reflectivities would enable the depth of the notch to be estimated easily. The DL2Hy/DBG deconvolutions give better results than the L2Hy/BG methods and effectively compensate for the poor estimation of the wavelet phase.

9.5.2. Deconvolution with a measured wave

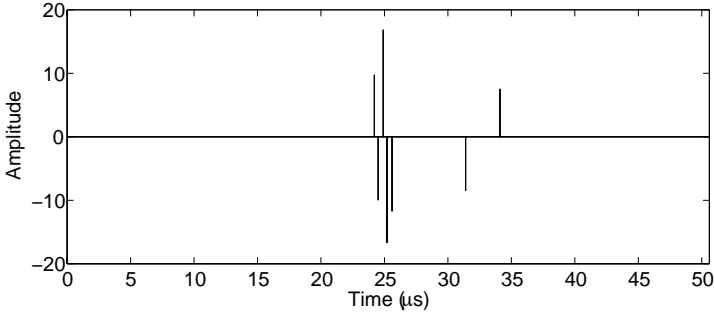
Using the measured wave in L2Hy deconvolution brings an improvement over the blind L2Hy version (see Figures 9.7a and 9.9a) but does not reach the performance levels of blind DL2Hy deconvolution (Figures 9.8a and 9.9a). The measured wave



(a) L2Hy deconvolution of B-scan, $(\lambda, T) = (100, 0.005)$

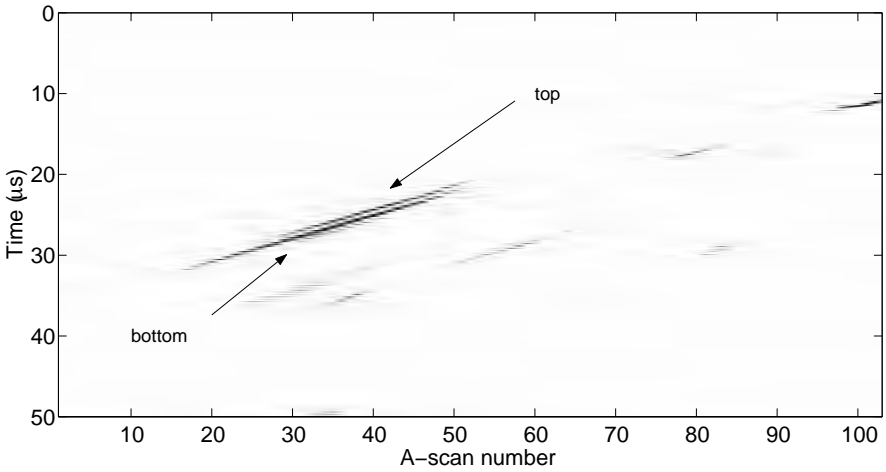


(b) L2Hy deconvolution of A-scan # 40, $(\lambda, T) = (100, 0.005)$

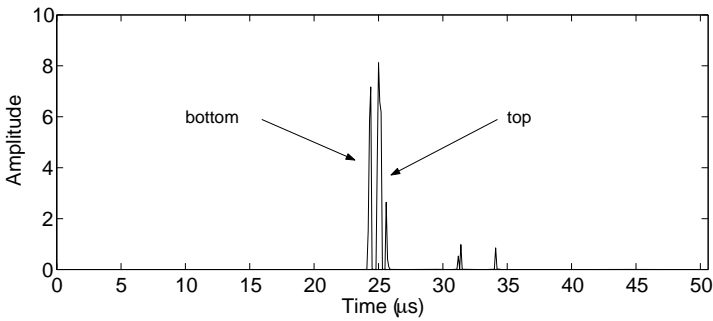


(c) BG deconvolution of A-scan # 40, $(\lambda, r_x, r_n) = (0.005, 50, 30)$

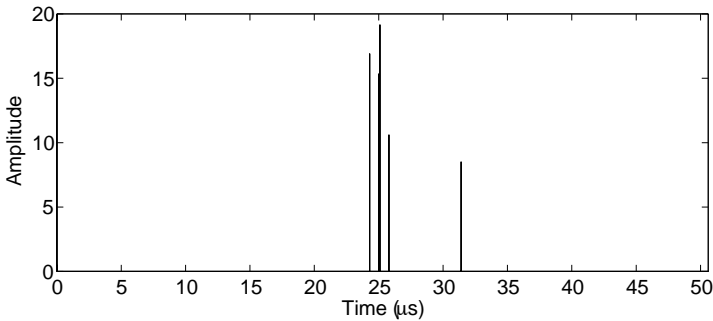
Figure 9.7. Blind L2Hy/BG deconvolution of the data of Figure 9.3. The wavelet was estimated beforehand, A-scan by A-scan, by AR estimation, $q = 8$



(a) blind DL2Hy deconvolution of B-scan, $(\lambda, T) = (100, 0.005)$



(b) DL2Hy deconvolution of A-scan # 40, $(\lambda, T) = (100, 0.005)$



(c) DBG deconvolution of A-scan # 40, $(\lambda, r_x, r_n) = (0.005, 50, 30)$

Figure 9.8. Blind DL2Hy/DBG deconvolution of the data of Figure 9.3. The wavelet was estimated beforehand, A-scan by A-scan, by AR estimation, $q = 8$

doubtless does not precisely correspond to the wave actually transmitted and moreover, unlike L2Hy deconvolution, DL2Hy deconvolution allows deformations of the wave as it propagates to be taken into account to some degree.

DL2Hy deconvolution gives a slightly “cleaner” deconvolved B-scan than when the estimated wavelet is used (see Figures 9.8a and 9.9b). This better result can no doubt be explained by the better fidelity to reality of the spectral content of the measured wave.

Thus, blind DL2Hy/DBG deconvolution gives results which, although degraded, are very close to those obtained using the measured wave. When we consider the more realistic situation in which the measured wave is difficult to obtain, the approach proposed for blind DL2Hy/DBG deconvolution appears particularly efficient.

9.5.3. *Comparison between DL2Hy and DBG*

As mentioned in the presentation of the results of blind deconvolution, the DL2Hy and DBG deconvolutions produce similar results. Thus, the comparison made here concerns the detailed behavior of the methods. The study is centered on restoring the reflectivities that are the most interesting from an application point of view with respect to those of the reflectors for the top and bottom of the notch. It was carried out for the case where the measured wave was used in order to study, in particular, the stability of the results for the various A-scans.

Close-ups around the interesting reflectors restored by the DL2Hy and DBG deconvolutions are shown in Figures 9.10a and 9.10b. From these B-scans, the results do indeed look very similar. However, the comparison can be made finer by studying the amplitudes of the two reflectors and the time difference between them.

Theoretically, the corner echo travels in a single direction while the diffraction scatters in all directions. Therefore, except for the limit case when the bottom of the notch is at the edge of the ultrasound beam, the amplitude of the reflector associated with the bottom of the notch should be higher than that of the second reflector. This amplitude difference is indeed found in the case of DL2Hy deconvolution (see Figure 9.11a). However, in DBG deconvolution, the reflector connected with the top of the notch sometimes has an amplitude comparable to that restored for the bottom (see Figure 9.11b). Furthermore, if we compare Figures 9.11a and b, it appears that the variation of the amplitudes of the reflectors is more stable in the case of DL2Hy deconvolution than for DBG.

The time difference between the two reflectors allows the height of the defect to be estimated from the speed of propagation and the angle of incidence of the beam. Theoretically, this difference should be stable from one sensor position to another. The

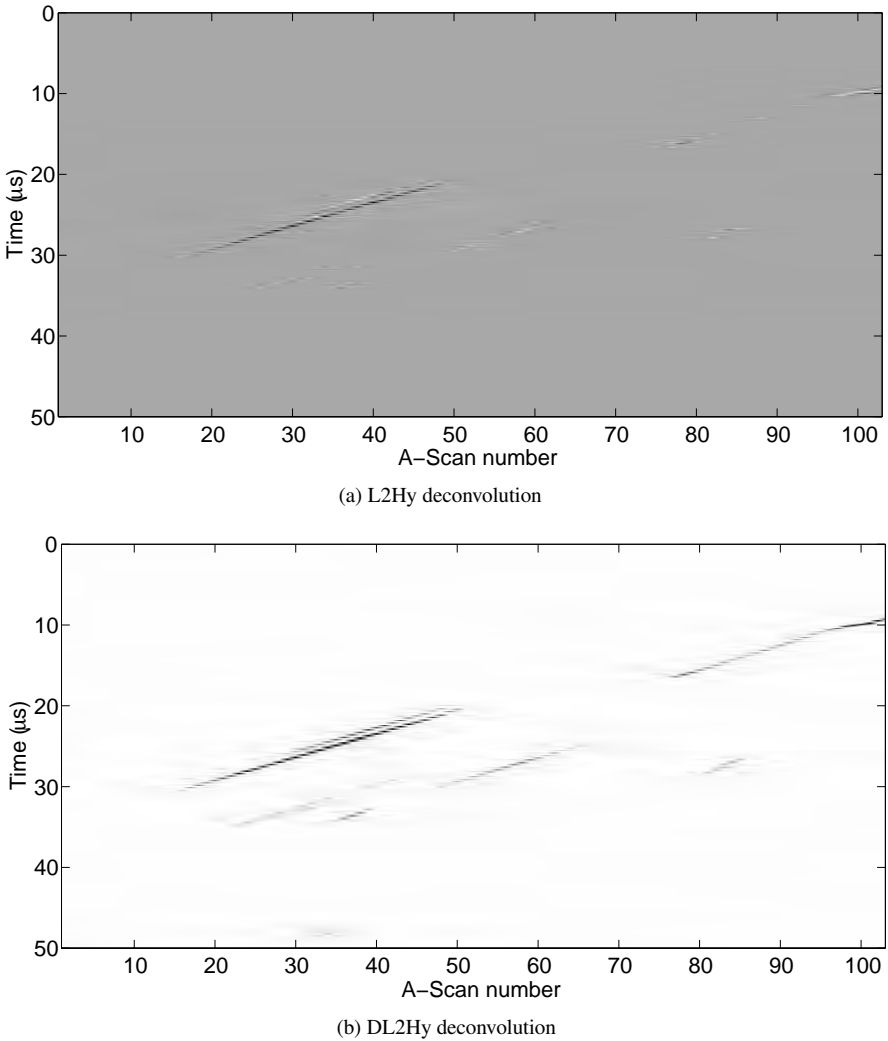
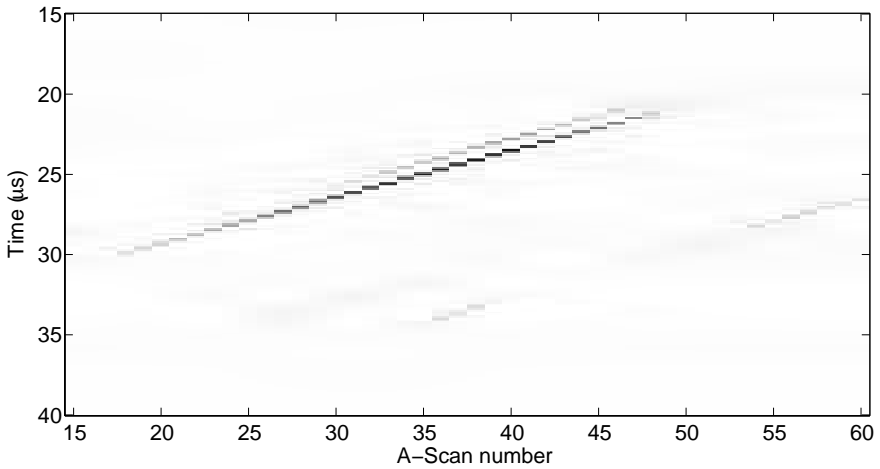
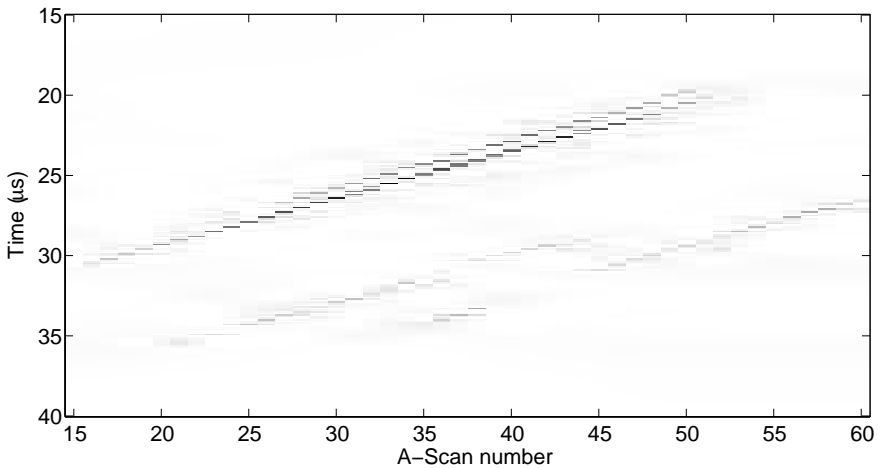


Figure 9.9. Comparison between L2Hy and DL2Hy deconvolution of B-scan of Figure 9.3a by measured wavelet; $(\lambda, T) = (100, 0.005)$ in both cases

processing results for the number of samples between the two reflectors is reported in Figure 9.11c for each sensor position. Overall, the difference is smaller for DBG than DL2Hy. The difference is not discriminating, however, as the uncertainty on the height measurement depends on other parameters such as errors of estimation of the velocity at which the wave propagates in the material. In contrast, the dispersion of



(a) zoom on DL2Hy deconvolution, $(\lambda, T) = (100, 0.005)$



(b) zoom on DBG deconvolution, $(\lambda, r_x, r_n) = (0.005, 50, 25)$

Figure 9.10. Comparison between DL2Hy and DBG deconvolution of B-scan of Figure 9.3a by measured wavelet (close-up of zone of interest)

the difference is greater in the DBG case: between five and eight samples, whereas it is between seven and eight in the DL2Hy case. Here again, DL2Hy is more stable than DBG.

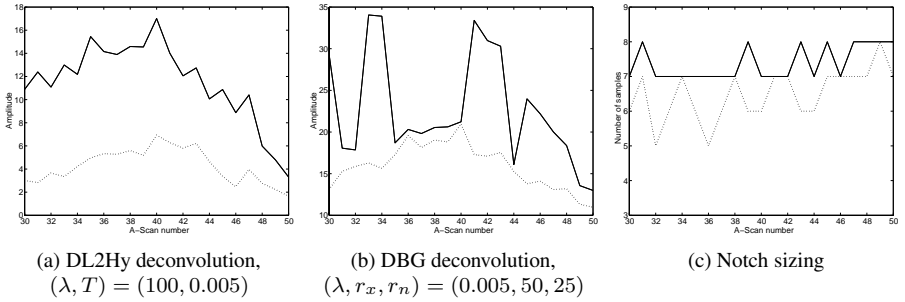


Figure 9.11. Amplitude of reflectors connected with the top (—) and bottom of the notch (---), restored by deconvolution by measured wavelet: (a) DL2Hy method; (b) DBG method; (c) number of samples separating the reflectors connected with the top and bottom of the notch for DL2Hy (—) and DBG (---) deconvolution

9.5.4. Summary

This processing of real data illustrates how NDE can gain from deconvolution. In particular, blind DL2Hy/DBG deconvolution provides a great improvement in resolution and enables an accurate inter-reflector distance measurement. This approach improves testing performance and opens up new perspectives. NDE experts sometimes find these results disconcerting as they give access to information that was masked in the measurements. The application of these new techniques by non-destructive evaluation specialists should therefore be preceded by the decisive step of an introduction to deconvolution. The results provided by DL2Hy and DBG deconvolution are qualitatively similar. However, the stability study on the amplitude of the reflectivities and the distance between reflectors shows that DL2Hy deconvolution is the more robust of the two. The interest of robustness does not appear clearly in the case of a notch but can prove decisive in the case of a crack.

9.6. Conclusion

Its lack of time resolution is a limitation of ultrasonic inspection: two reflectors that are close together are merged into a single echo. It is possible to model measurement acquisition as the result of one-dimensional convolution between a sequence of reflectors representative of the object under inspection and a kernel associated with the incident wave. In this framework, improving the resolution of ultrasonic evaluation has been treated here as a blind pulse train deconvolution problem.

We have given an overview of blind deconvolution techniques going beyond the NDE framework, then presented an original approach for blind pulse train

deconvolution based on imperfect estimation of the kernel followed by DL2Hy/DBG pulse deconvolution. The second step compensates for both the imperfections in the wave estimation technique and possible phase deformations of the kernel during propagation.

Applied to ultrasonic NDE, this approach pushes back the limits of inspection by providing a marked improvement in the time-resolution and appreciable help for the interpretation of measurements. These new results are surprising for NDE experts and must be accompanied by an effort at explanation and training from signal processing specialists. From this point of view, presenting deconvolution as a problem of minimizing a penalized criterion has the advantage of simplicity.

The approach put forward for blind deconvolution could be applied to other types of imaging by propagation of mechanical waves: biomedical ultrasound scans, seismic reflection in geophysics, etc. Concerning the use of DL2Hy/DBG deconvolutions only, they would be suitable for numerous pulse deconvolution problems, particularly when the kernel is poorly known or tends to deform with time.

From a methodological point of view, DL2Hy/DBG deconvolution can be adapted to the case of a two-dimensional kernel. Applied in ultrasonic NDE, it has been successfully achieved by Labat *et al.* [LAB 05] in the aim of improving the resolution in the lateral direction. In parallel or independently, the introduction of a prior two-dimensional model for the reflectivity would enable the lateral continuity of the reflectors to be taken into account [IDI 93]. Finally, a way of taking correlated structures in the noise into consideration could be easily introduced by modifying the data fitting term and would be advantageous for dealing with measurements containing strong structural noise.

9.7. Bibliography

- [ATA 82] ATAL B. S., REMDE J. R., "A new method of LPC excitation for producing natural sounding speech at low bit rates", in *Proc. IEEE ICASSP*, vol. 1, Paris, France, p. 614-617, May 1982.
- [CHA 93] CHAMPAGNAT F., IDIER J., DEMOMENT G., "Deconvolution of sparse spike trains accounting for wavelet phase shifts and colored noise", in *Proc. IEEE ICASSP*, Minneapolis, MN, p. 452-455, 1993.
- [CHE 96] CHENG Q., CHEN R., LI T.-H., "Simultaneous wavelet estimation and deconvolution of reflection seismic signals", *IEEE Trans. Geosci. Remote Sensing*, vol. 34, p. 377-384, Mar. 1996.
- [COO 90] COOKEY M., TRUSSELL H. J., WON I. J., "Seismic deconvolution by multipulse coding", *IEEE Trans. Acoust. Speech, Signal Processing*, vol. 38, num. 1, p. 156-160, Jan. 1990.

- [GAU 96] GAUTIER S., Fusion de données gammagraphiques et ultrasonores. Application au contrôle non destructif, PhD thesis, University of Paris XI, France, Dec. 1996.
- [GAU 97] GAUTIER S., IDIER J., MOHAMMAD-DJAFARI A., LAVAYSSIÈRE B., “Traitement d’échogrammes ultrasonores par déconvolution aveugle”, in *Actes 16^e coll. GRETSI*, Grenoble, France, p. 1431-1434, Sep. 1997.
- [GOU 89] GOUSSARD Y., DEMOMENT G., “Recursive deconvolution of Bernoulli-Gaussian processes using a MA representation”, *IEEE Trans. Geosci. Remote Sensing*, vol. GE-27, p. 384-394, 1989.
- [IDI 93] IDIER J., GOUSSARD Y., “Multichannel seismic deconvolution”, *IEEE Trans. Geosci. Remote Sensing*, vol. 31, num. 5, p. 961-979, Sep. 1993.
- [LAB 05] LABAT C., IDIER J., RICHARD B., CHATELLIER L., “Ultrasonic nondestructive testing based on 2D deconvolution”, in *PSIP’2005 : Physics in Signal and Image Processing*, Toulouse, France, Jan. 2005.
- [LAZ 93] LAZEAR G. D., “Mixed-phase wavelet estimation using fourth-order cumumants”, *Geophysics*, vol. 58, p. 1042-1049, 1993.
- [ORF 85] ORFANIDIS S. J., *Optimal Signal Processing – An Introduction*, Macmillan, New York, NY, 1985.
- [ROB 67] ROBINSON E. A., “Predictive decomposition of time series with application to seismic exploration”, *Geophysics*, vol. 32, p. 418-484, 1967.
- [SAI 85] SAINT-FELIX D., HERMENT A., DU X.-C., “Fast deconvolution: application to acoustical imaging”, in J.M. THIJSEN, V. MASSEO (Eds.), *Ultrasonic Tissue Characterization and Echographic Imaging*, Nijmegen, The Netherlands, Faculty of Medicine Printing Office, p. 161-172, 1985.
- [VIV 89] VIVET L., Amélioration de la résolution des méthodes d’échographie ultrasonore en contrôle non destructif par déconvolution adaptative, PhD thesis, University of Paris XI, France, Sep. 1989.
- [WIG 78] WIGGINS R. A., “Minimum entropy deconvolution”, *Geoexploration*, vol. 16, p. 21-35, 1978.

Chapter 10

Inversion in Optical Imaging through Atmospheric Turbulence

10.1. Optical imaging through turbulence

10.1.1. *Introduction*

The theoretical resolving power of a telescope is limited by its diameter. In real instruments, this theoretical limit, called the diffraction-limit resolution, often cannot be reached because of the presence of optical aberrations. These aberrations may come from the telescope itself or from the light wave propagation medium. In the case of ground-based astronomy, aberrations are mostly due to atmospheric turbulence. Several techniques have been developed to improve the resolution of observation instruments and avoid the degradation caused by turbulence. In this section, we recall some essential ideas in optical imaging, in particular on the optical effects of turbulence, then review the various techniques of high-resolution imaging through turbulence.

Section 10.2 gives a brief presentation of the inversion approach and the regularization criteria used in this chapter. Section 10.3 is an introduction to wavefront sensors (WFSs) and the processing problems that arise from their use. WFSs are devices

Chapter written by Laurent MUGNIER, Guy LE BESNERAIS and Serge MEIMON.

Laurent Mugnier and Serge Meimon are grateful to their colleagues of the High Angular Resolution team and particularly to its successive leaders Marc Séchaud and Vincent Michau, who created and maintained a stimulating team spirit made up of intellectual curiosity and a desire to share knowledge. With special thanks from Laurent Mugnier for the discussions with Jean-Marc Conan that introduced him to this field and from Serge Meimon for his talks with Frédéric Cassaing.

that measure optical aberrations and are essential components of many high resolution optical imaging instruments today.

Three imaging techniques are illustrated by the inverse problems associated with them. These inverse problems are: image restoration for deconvolution from wave-front sensing and for imaging using adaptive optics, discussed in section 10.4, and image reconstruction for optical interferometry (section 10.5).

10.1.2. Image formation

10.1.2.1. Diffraction

Image formation is well described by the scalar theory of diffraction, presented in detail in reference works such as [GOO 68, BOR 93]. A modern introductory overview can be found in [MAR 89]. Image formation can be modeled by a convolution, at least within the instrument's so-called isoplanatic patch. At visible wavelengths, this patch is typically of the order of a degree when only aberrations due to the telescope itself are considered and a few arcseconds ($1 \text{ arcsec} = 1/3600^\circ$) for a telescope observing space through turbulence.

The instantaneous point-spread function (PSF) of a telescope or “telescope + atmosphere” system is equal to the square modulus of the Fourier transform (FT) of the complex amplitude of the field $\psi = P \exp(j\varphi)$ present in the aperture of the instrument when the object observed is a point source:

$$h(\boldsymbol{\xi}) = \left| \text{FT}^{-1} \left(P(\lambda \mathbf{u}) e^{j\varphi(\lambda \mathbf{u})} \right) \right|^2 (\boldsymbol{\xi}) \quad (10.1)$$

where λ is the imaging wavelength and imaging is assumed quasi-monochromatic. This PSF is conventionally normalized to a unit integral. In expression (10.1), the FT is the field transformation performed by the telescope between the pupil plane and the focal plane, and the square modulus is due to the detection being quadratic i.e., in intensity. Vector $\boldsymbol{\xi} = [\xi, \zeta]^T$ is composed of angles in the sky, in radians. For a perfect telescope in the absence of turbulence, P is constant in the aperture and φ is zero¹. For a real telescope, the variations of the field $P \exp(j\varphi)$ are due both to aberrations belonging to the telescope and to those introduced by the turbulence.

In what follows, we assume that P is simply the aperture indicatrix, i.e., that the variations of intensity in the input pupil are negligible. This hypothesis is generally valid in astronomical imaging and is called the near-field approximation.

1. The corresponding PSF is called the Airy pattern.

Equation (10.1) indicates that the optical transfer function, or OTF, is the autocorrelation of $\psi = P e^{j\varphi}$ dilated by the inverse of the wavelength, which is written

$$\hat{h}(\mathbf{u}) = P e^{j\varphi} \otimes P e^{j\varphi}(\lambda \mathbf{u}). \quad (10.2)$$

In the absence of aberrations, i.e., when the phase φ is zero, the OTF is the autocorrelation of the aperture P . It has a spatial cut-off frequency equal to $D/\lambda \text{ rad}^{-1}$, where D is the diameter of the aperture, and is strictly zero beyond it. The ultimate resolution of a telescope (sometimes called a monolithic telescope in contrast to the interferometers described below) is thus limited by its diameter D . Today's technology limits diameters to ten meters or so for ground-based telescopes and a few meters for telescopes on board satellites because of size and mass constraints. Optical interferometry (OI) is a technique that allows us to go beyond the resulting resolution limitation.

10.1.2.2. Principle of optical interferometry

This technique consists of taking the electromagnetic fields received at each of the apertures of an array of apertures (basic telescopes or mirror segments) and making them interfere. For each pair (k, ℓ) of apertures, the data contains high-resolution information at (or around) the angular spatial frequency $\mathbf{B}_{k,\ell}/\lambda$, where $\mathbf{B}_{k,\ell}$ is the vector separating the apertures, or *baseline*. This spatial frequency can be much larger than the cut-off frequency D/λ of the individual apertures.

Depending on the type of interferometer and beam combination, it is possible either to form and measure an image of the object directly (the interferometer is then called an imaging interferometer) or to measure a discrete set of spatial frequencies of the object of interest (the interferometer can then be called a “correlation interferometer” as it measures the correlation of the electromagnetic fields between apertures [CAS 97]). The reader interested in a more precise description of the different types of optical interferometers is invited to consult [ROU 01].

For a monolithic telescope, as for an interferometer, the transfer function is the autocorrelation of the input pupil (see equation (10.2)) provided that, if the interferometer is of the correlation type, the apertures are assimilated to points. For a long-baseline interferometer, i.e., when the baselines are large relative to the diameter of the individual apertures – which is generally the case for correlation interferometers – the difference between imaging and correlation interferometers becomes negligible as far as the information recorded in the data is concerned. The transfer functions of a monolithic telescope, an imaging interferometer and a correlation interferometer are illustrated in Figure 10.1. For an imaging interferometer, the processing required is, to a good approximation, a deconvolution, with a PSF still given by equation (10.1) but more irregular than with a monolithic telescope because of the shape of the aperture.

For a correlation interferometer, the nature of the data processing problem changes: here, the aim is to reconstruct an object from Fourier coefficients, a problem called Fourier synthesis. This is the problem that will be tackled in section 10.5.

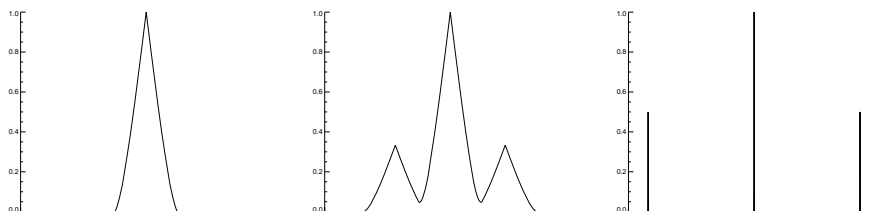


Figure 10.1. Cross-sections of transfer functions of a monolithic telescope (left), a three-telescope imaging interferometer (center) and a two-telescope correlation interferometer (right)

An intuitive way of representing data formation in a long-baseline interferometer is Young's double hole experiment, in which the aperture of each telescope is a (small) hole letting through the light coming from an object located at a great distance. Each pair (k, ℓ) of telescopes then gives a fringe pattern with a spatial frequency of $B_{k,\ell}/\lambda$, where $B_{k,\ell}$ is the vector linking telescopes k and ℓ . The contrasts and positions of these fringes can be measured and grouped together naturally in a number called the "complex visibility", which, in an ideal situation and in the absence of turbulence, gives the value of $\hat{x}(B_{k,\ell}/\lambda)/\hat{x}(0)$ (Van Cittert-Zernike theorem [GOO 85, MAR 89]).

10.1.3. Effect of turbulence on image formation

10.1.3.1. Turbulence and phase

The inhomogeneities in the air temperature of the atmosphere generate inhomogeneities in the refractive index of the air, which perturb the propagation of light waves through the atmosphere. These perturbations lead to space and time variations of the pupil phase φ , which can be modeled by a random process. In this section we recall a few results that enable the turbulent aperture phase to be modeled up to the second order. We will use the assumption, which is generally well verified, at least for scales of less than about ten meters, that the random variations of the refractive index of the air obey Kolmogorov's law: they follow a Gaussian probability law with zero mean and power spectral density (PSD) proportional to $|\nu|^{-11/3}$, where ν is the 3D spatial frequency [ROD 81].

By integrating the phase along the optical path and in the framework of the near-field approximation, the spatial statistics of the phase in the telescope aperture can be deduced for a plane wave entering the atmosphere. The phase in the aperture is Gaussian since it is the result of the sum of all the index perturbations from the upper

atmosphere down to the ground [ROD 81]. The PSD of this phase is [NOL 76]:

$$S_\varphi(\mathbf{u}) = 0.023 r_0^{-5/3} u^{-11/3} \quad (10.3)$$

where \mathbf{u} is the 2D spatial frequency in the aperture, u is its modulus, and r_0 is the key parameter quantifying the strength of the turbulence, called Fried's diameter [FRI 65]. The smaller r_0 , the stronger the turbulence. Typically, its value is about 10 cm in the visible range at a relatively good site.

The typical variation time τ of the turbulent phase in the aperture is given by the ratio of characteristic scale r_0 of this phase to mean wind speed Δv (which, more accurately, is a standard deviation of the distribution of the moduli of wind velocities [ROD 82]):

$$\tau = r_0 / \Delta v. \quad (10.4)$$

For $r_0 \simeq 10$ cm and $\Delta v \simeq 10 \text{ m.s}^{-1}$, we obtain $\tau \simeq 10^{-2}$ s. Thus, we talk about long exposures for images corresponding to an integration markedly longer than this time and short exposures for images with shorter integration times. For a full treatment of the time statistics of the turbulent phase, see [CON 95].

10.1.3.2. Long-exposure imaging

The turbulent long-exposure OTF is the product of the so-called static OTF, \widehat{h}^s , of the telescope without the atmosphere and an atmospheric transfer function, \widehat{h}^a , which has a cut-off frequency r_0/λ [ROD 81]:

$$\widehat{h}(\mathbf{u}) \triangleq \langle \widehat{h}_t(\mathbf{u}) \rangle = \widehat{h}^s(\mathbf{u}) \widehat{h}^a(\mathbf{u}) \text{ with } \widehat{h}^a(\mathbf{u}) = \exp\{-3.44 (\lambda u / r_0)^{5/3}\}, \quad (10.5)$$

where the angle brackets $\langle \cdot \rangle$ denote a temporal mean over an arbitrarily long time. Thus we see that, for a telescope with a large diameter $D \gg r_0$, the long-exposure imaging resolution is limited by the turbulence and is no better than for a telescope of diameter r_0 .

10.1.3.3. Short-exposure imaging

As noted by Labeyrie [LAB 70], when the exposure time is short enough to freeze the turbulence (typically less than 10 ms, see equation (10.4)), the images retain the high-frequency information in the form of *speckles*, having a typical size λ/D and random positions. This is illustrated in Figure 10.2, which shows the simulated image of a star viewed through turbulence ($D/r_0 = 10$) using short (left) and long (right) exposures.

It is possible to quantify the high-frequency information present in short-exposure images by evaluating the speckle transfer function (STF), defined as the second order

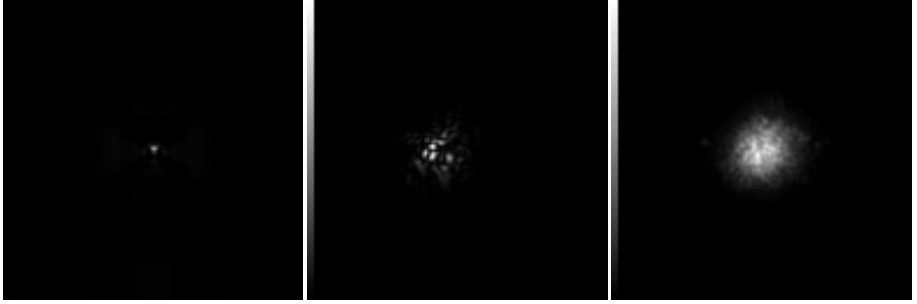


Figure 10.2. Images of a star simulated without atmospheric turbulence (left) and through turbulence (short exposure in center and long exposure on right). The strength of the turbulence is $D/r_0 = 10$. Image sampling respects Shannon's condition

moment of the instantaneous transfer function, $\langle |\widehat{h}_t(\mathbf{u})|^2 \rangle$. For a large-diameter telescope ($D \gg r_0$), if we take an approximation on the turbulence statistics, we can find an approximate expression for the STF [ROD 81]:

$$\langle |\widehat{h}_t(\mathbf{u})|^2 \rangle \simeq \langle \widehat{h}_t(\mathbf{u}) \rangle^2 + 0.435 (r_0/D)^2 \widehat{h}_0^s(\mathbf{u}) \quad (10.6)$$

where \widehat{h}_0^s is the transfer function of a perfect telescope (i.e., with no aberration) of diameter D .

This expression allows us to describe the STF as the sum of the square of the long-exposure transfer function, which is low-frequency (LF), and a high-frequency (HF) component that extends up to the cut-off frequency of the telescope with an attenuation proportional to $(D/r_0)^2$. Thus, if we process a set of short-exposure images using a more judicious method than a simple average, it is possible to recover a high-resolution image of the observed object.

10.1.3.4. Case of a long-baseline interferometer

Equation (10.5) applies whatever the shape of the instrument's aperture and thus, in particular, applies to an interferometer. In a long exposure, the contrast of the fringes measured for a baseline $\mathbf{B}_{k,\ell}/\lambda$ is therefore multiplied by $\widehat{h}^a(\mathbf{B}_{k,\ell}/\lambda)$ and so strongly attenuated as to make the measurement of $\widehat{x}(\mathbf{B}_{k,\ell}/\lambda)$ unusable.

In a short exposure, for an interferometer where each aperture has a diameter smaller than the Fried's diameter, r_0 , or where turbulence is corrected using adaptive optics (see section 10.1.4.3), the impact of turbulence on the interferometer measurements is easy to model: in the Young's holes analogy mentioned above, each hole k adds a phase shift (or piston) $\varphi_k(t)$ to the wave going through it, because of aberrations introduced by the turbulence in front of this aperture. The interference between

two apertures k and ℓ are thus out of phase by the “differential piston” $\varphi_\ell(t) - \varphi_k(t)$, which, in a short exposure, is expressed by a random displacement of the fringes without attenuation of the contrast. The contrast attenuation in long exposures results from the averaging of these random displacements. Section 10.5 will present averaging techniques that get around the differential pistons. The short-exposure transfer function, at frequency $\mathbf{B}_{k,\ell}/\lambda$, can be written:

$$\hat{h}_t(\mathbf{B}_{k,\ell}/\lambda) = \eta_{k,\ell}(t) e^{j(\varphi_\ell(t) - \varphi_k(t))} \quad (10.7)$$

where $\eta_{k,\ell}(t)$ is a number that is often called the “instrumental visibility”. In the absence of the many potential sources of visibility loss (residual perturbations of the wavefront at each telescope, differential tilts between telescopes, differential polarization effects, non-zero spectral width, etc.), the value of $\eta_{k,\ell}(t)$ is the inverse of the number of apertures interfering simultaneously (equation (10.2) considering that P is a sum of Dirac delta functions). In practice, this instrumental visibility is calibrated on a star known to be unresolved by the interferometer. Taking this calibration into account, we can thus replace $\eta_{k,\ell}(t)$ by 1 in equation (10.7).

Note that the measurement baseline $\mathbf{B}_{k,\ell}$ between apertures k and ℓ depends on time: the aperture configuration as seen from the object changes as the Earth rotates. This is used in “super-synthesis”, a technique that consists, when the source emission does not vary in time, of repeating the measurements in the course of a night of observation to increase the frequency coverage of the interferometer.

10.1.4. Imaging techniques

The aim of high-resolution imaging through turbulence is to restore the HFs beyond the cut-off frequency r_0/λ of the long-exposure imaging. This is made possible by various experimental techniques that avoid the time-integration of phase defects introduced by the turbulence. A measure of the quality of the technique is thus the resulting signal-to-noise ratio (SNR) at high spatial frequencies.

10.1.4.1. Speckle techniques

The first high-resolution techniques were based on the acquisition of a series of short-exposure images and the calculation of empirical moments. Speckle interferometry² [LAB 70] uses the quadratic mean of the FTs of the images, which allows the autocorrelation of the observed object to be estimated. Knox and Thomson [KNO 74], then Weigelt [WEI 77] put forward processing methods using the cross-spectrum and

2. The term interferometry could mislead the reader into thinking that the instrument used here is an interferometer. This is by no means the case; the interferences in question arise from the aperture of a monolithic telescope.

the bi-spectrum respectively of the short-exposure images so as to estimate the object and not only its autocorrelation. These methods require the averages to be taken over a large number of images, even for simple objects, both to make the estimation of the statistical quantities valid and to improve the SNR.

10.1.4.2. *Deconvolution from wavefront sensing (DWFS)*

A notable enhancement of short-exposure imaging through turbulence was thus brought about, not by improving the processing of measurements but by changing the experimental technique itself. In 1985, Fontanella [FON 85] proposed a new imaging technique: deconvolution from wavefront sensing. This technique, based on the use of a device called a wavefront sensor (WFS), was experimentally validated shortly afterwards [PRI 88, PRI 90].

The aim of WFSs, which had so far only been used for controlling the surface quality of telescope mirrors, is to measure the aberrations of optical systems (the phase φ of equation (10.1)). Some of them, such as the Hartmann-Shack sensor used in deconvolution from wavefront sensing, work even if the object of interest is extended (rather than being a point source).

The technique of deconvolution from wavefront sensing consists of simultaneously recording a series of short-exposure images and Hartmann-Shack wavefront measurements. In practice, at least ten or so short-exposure images are typically needed to give correct spatial frequency coverage up to the telescope cut-off (equation (10.6)). The number of images required is greater if the observed object is not very bright.

Deconvolution from wavefront sensing is a considerable improvement on the other short-exposure techniques mentioned above. First of all, like the Knox-Thomson or bi-spectral techniques, it enables us to recover not the autocorrelation of the object but the object itself. Then, unlike the previous short-exposure techniques, this one does not need images of a reference star to be recorded, and it is called *self-referenced speckle interferometry* for this reason. Finally, its measurements are efficient in terms of photons collected: as the short-exposure images must be quasi-monochromatic to keep the speckles unblurred, all the remaining photons can be diverted towards the WFS without any loss of signal on the image channel. This technique thus makes it possible to record more information than the previous short-exposure techniques and, unlike those techniques, has a SNR that is not limited by the speckle noise at high flux [ROD 88b], because of its self-referenced nature. This explains why speckle interferometry has fallen into disuse nowadays.

Section 10.4.2 gives more details on the data processing in this technique, which is a double inverse problem (estimation of wavefronts from WFS measurements, which allows the instantaneous PSF corresponding to each image to be calculated, and estimation of the object from images and WFS measurements).

10.1.4.3. *Adaptive optics*

The imaging technique with the best performance in terms of SNR is adaptive optics (AO), which provides a real-time compensation for the aberrations introduced by atmospheric turbulence, generally by use of a mirror whose surface is deformed at all instants via a servo-loop, according to the measurements made by a WFS.

This technique thus enables long-exposure images (typically exposed for between a few seconds and several tens of minutes) to be recorded while retaining the HF of the observed object up to the cut-off frequency of the telescope. The HF is nevertheless attenuated as the correction is only partial [CON 94] and deconvolution is necessary. This deconvolution, for which the PSF is often imperfectly known, is presented in section 10.4.3.

The most commonly used WFS is the Hartmann-Shack sensor (see section 10.3.2). The associated deformable mirror has actuators made of, e.g., stacked piezoelectric material controlled by high voltages. The AO technique was proposed by Babcock as early as 1953, and developed from the 1970s for defence purposes, first in the USA, then in France, but it was not until the late 1980s that the first AO system for astronomy came into being [ROU 90]. Any reader interested in a detailed account of AO should consult a reference work such as [ROD 99].

10.1.4.4. *Optical interferometry*

This section describes some of the major steps in the development of ground-based stellar interferometry and takes its inspiration partly from [MON 03].

10.1.4.4.1. The first measurements of stars

The use of interferometry for observing stars was first suggested by Fizeau in 1868, the aim being simply to measure the size of celestial bodies. However, it was not until 1890 that the technique was implemented experimentally by Michelson [MIC 91], who measured the diameters of Jupiter's moons by masking a telescope with two fine slits four inches apart. In 1920-21, he and Pease measured the diameter of the star Betelgeuse using a 20 foot (6 meter) interferometer [MIC 21].

Pease's unsuccessful attempts to reach a baseline of 50 feet marked the start of a difficult period for optical interferometry. At the same time, the advances made in radar during the Second World War led to the development of interferometry at radio wavelengths. Resolutions smaller than a milliarcsecond were reached in radio interferometry while its optical counterpart rather fell into neglect because of the many technical difficulties involved in coherently combining the beams from two telescopes. In optics, it is impossible to record the phase, and the beams therefore have to be combined in real time. Another handicap for optics is that the effects of turbulence evolve much faster than for radio.

10.1.4.4.2. Renewed interest in optical interferometry

The first coherent combination of optical beams emitted by a star using a *long baseline* interferometer was achieved by Labeyrie in 1974 [LAB 75], with a baseline of 12 meters on a 2-telescope interferometer called I2T. This was followed by a more ambitious version composed of telescopes 1.5 meter in diameter with a baseline of up to 65 meters, which was named the Grand Interféromètre à 2 Télésopes (GI2T) [MOU 94]. Up to that time, interferometry had been used with two apertures to measure the spatial spectrum of an astronomical scene from which a few parameters were extracted to validate or reject an astrophysical model. In particular, only the modulus of the spatial spectrum could be used. In addition, without the phase, it is generally impossible to determine the geometry of the observed scene. In 1987, by masking a monolithic telescope, Hannif *et al.* showed that it was possible to obtain interferometric arrays [HAN 87], i.e., to form interference fringes simultaneously for each pair of telescopes of the array. This technique, in addition to providing several measurements at once (15 frequencies per exposure for a 6-telescope interferometer), gave access to the phase of the spatial spectrum of the object [BAL 86], thus making interferometric imaging possible for scenes more complex than a uniform disk or a binary system. The remarkable potential of this method encouraged several teams to build such instruments. In 1995, the COAST interferometer made the first simultaneous combination with three telescopes [BAL 96], and was followed a few months later by NPOI [BEN 97] then IOTA [IOT] (now decommissioned). Since these instruments are evolving quickly, the interested reader is advised to visit <http://olbin.jpl.nasa.gov/> for up-to-date information.

10.1.4.4.3. Future of interferometry

The technology needed to build optical interferometric arrays has now come of age and draws upon the sister fields of integrated optics, adaptive optics and fiber optics:

- *integrated optics* has been successfully used in multi-telescope simultaneous combination for several years, in particular on the IOTA interferometer [BER 03] (the experimental data processed at the end of this chapter were obtained with this system);
- *adaptive optics* on large telescopes such as those of the Very Large Telescope Interferometer makes it possible to observe objects of low luminosity;
- *fiber optics* provides monomode fibers that will allow telescopes to be connected interferometrically over very large distances. The OHANA project plans to combine the seven largest telescopes on the summit of Mauna Kea in Hawaii to form an interferometer. The array thus formed will have a maximum baseline of 800 m [PER 06].

In parallel with the development of these large correlation interferometers, the technology for making *imaging* interferometers is now available. These instruments should eventually lead to considerable savings in volume and mass relative to an equivalent monolithic telescope, which would make them ideal candidates for space

missions. On the ground, they would be an alternative to giant monolithic telescopes (of several tens of meters). The first of them is the Large Binocular Telescope LBT [HIL 04], which will combine two eight-meter telescopes corrected by adaptive optics.

Segmented telescopes, such as those of the Keck Observatory [KEC], have been in use for several years and are at the boundary between imaging interferometers and telescopes. Their primary mirrors are composed of joined petals that are easier to manufacture than a monolithic mirror and this technique has been chosen for the future European giant telescope E-ELT (for *European Extremely Large Telescope*) and the US TMT (for *Thirty Meter Telescope*).

In addition to the correlation and imaging interferometers described in this section, there are other sorts of interferometers. P. Lawson [LAW 97] has collected together a selection of reference publications in this field as a whole. In this chapter, we shall only deal with the problem of processing the data collected by means of a correlation interferometer observing space from the ground through turbulence.

10.2. Inversion approach and regularization criteria used

Inversion in optical imaging through turbulence is generally an ill-posed problem in the case of a monolithic telescope and an under-determined problem in the case of an interferometer.

In deconvolution from wavefront sensing and in adaptive optics, we need to solve an image restoration problem for which the Bayesian approach already described in this book can be used directly. In the case of so-called *conventional* deconvolution, where the PSF is taken to be perfectly known, the estimated object is defined as the minimizer of a compound criterion containing a data fidelity term J_y and a prior fidelity term J_x . In OI the image is to be reconstructed from heterogeneous data and with a knowledge of the transfer function which is very incomplete because of the turbulence. There are several possible approaches for handling this type of data; the details are given below.

In all cases it is necessary to regularize the inversion to reach acceptable solutions. This will be done here by using a regularization term J_x in the minimized criterion to obtain the solution. The regularization criteria used in this chapter are taken from those described below and are all convex.

Quadratic criteria are the most widely used. We will use a criterion of this type in DWFS and OI with a parametric model of the object spectrum such as the one proposed for adaptive optics in [CON 98b]. An advantage of these criteria is that it is possible to estimate the parameters of the model easily, by maximum likelihood for

example. See [BLA 03] for the identification of the spectrum model of [CON 98b] with simultaneous estimation of the aberrations and [GRA 06] for an application to adaptive optics with known PSF. This model can also be identified from the data in OI [MEI 05a].

For objects with sharp edges such as artificial satellites, asteroids or planets, a quadratic criterion tends to oversmooth the edges and introduce spurious oscillations, or *ringing*, in their neighborhood. A solution is thus to use an edge-preserving criterion such as the so-called quadratic-linear, or L_2L_1 , criteria, which are quadratic for weak gradients of the object and linear for the stronger ones. The quadratic part ensures good noise smoothing and the linear part cancels out edge penalization (see Chapter 6). Here, for DWFS (section 10.4.2) and AO (section 10.4.3), we will use an isotropic version [MUG 01] of the criterion proposed by Rey [REY 83] in the context of robust estimation and used by Brette and Idier in image restoration [BRE 96]:

$$J_x(\mathbf{x}) = \mu\delta^2 \sum_r (\Lambda\mathbf{x}(\ell, m)/\delta - \log(1 + \Lambda\mathbf{x}(\ell, m)/\delta)) \quad (10.8)$$

where $\Lambda\mathbf{x}(\ell, m) = \sqrt{\nabla_\xi \mathbf{x}(\ell, m)^2 + \nabla_\zeta \mathbf{x}(\ell, m)^2}$, with $\nabla_\xi \mathbf{x}$ and $\nabla_\zeta \mathbf{x}$ as the gradient approximations by finite differences in the two spatial directions.

For objects composed of bright points on a fairly smooth background, such as are often found in astronomy, we can consider an L_2L_1 prior that is white, i.e., where pixels are independent. Such a prior is obtained by using the regularization of equation (10.8) but with $\Lambda\mathbf{x} = \mathbf{x}$. This is what we will do for all the interferometric data of section 10.5.4. Unlike for the case of quadratic regularization with an object spectrum model, the tuning of the hyperparameters has to be supervised here.

10.3. Measurement of aberrations

10.3.1. Introduction

The WFS is a key element of modern high-resolution imaging instruments as it allows the instrument aberrations and the atmospheric turbulence to be measured so that they can be compensated for, either in real time (AO) or by post-processing.

Many WFSs are currently available, and a thorough review is given in [ROU 99]. They can be divided into two families: focal-plane sensors and pupil-plane sensors.

Present-day AO systems use either a Hartmann-Shack sensor [SHA 71], which is well described in [FON 85], or a curvature sensor [ROD 88a]. Both belong to the pupil-plane family and use a fraction of the incident light, which is diverted by a dichroic beam-splitter. For AO, both have the appealing properties that they work with a broad spectral band (because they can be well described by geometrical optics)

and that the relationship between the unknown aberrations and the data is linear, so the inversion can be performed in real time. The next subsection presents the principle of the Hartmann-Shack sensor. This sensor will be seen later in the DWFS technique and is the most widely used in AO.

The focal-plane family of sensors was born from the very natural idea that an image of a given object contains information not only about the object, but also about the wavefront. A focal-plane sensor thus requires little or no optics other than the imaging sensor; it is also the only way to be sensitive to all aberrations down to the focal plane.

Section 10.3.3 briefly presents the focal-plane wavefront sensing technique known as phase diversity [GON 82]. This technique is simple in its hardware requirements and, like the Hartmann-Shack, works on very extended objects. Finally, it should be noted that there are special WFSs called co-phasing sensors that can measure the differential piston between apertures, which are aberrations specific to interferometers. Phase diversity can be used both as a WFS and as a co-phasing sensor. Differential pistons are not yet corrected on the interferometers in operation at present.

10.3.2. *Hartmann-Shack sensor*

The principle of this sensor is illustrated in Figure 10.3: an array of $N_{\text{ml}} \times N_{\text{ml}}$ micro-lenses is placed in a pupil plane (image of the telescope entrance pupil). It samples or, in other words, cuts up the incident wavefront. At the focus of the array, a set of detectors (CCD camera, for example) records the N_{ml}^2 sub-images, each of which is the image of the object observed through the part of the pupil cut out by the corresponding micro-lens. When the wavefront is perturbed by aberrations, each micro-lens sees approximately a tilted plane wavefront and the corresponding sub-image is thus shifted relative to its reference position by an amount proportional to the mean slope of the wavefront. In the case of aberrations due to atmospheric turbulence, N_{ml} should be chosen so that the size of each micro-lens relative to the entrance pupil of the instrument is of the order of the Fried's diameter r_0 . The position of the center of gravity of each sub-image is measured, thus giving a map of the mean slopes of the wavefront on a $N_{\text{ml}} \times N_{\text{ml}}$ grid³.

3. It is possible to envisage taking the measurements to be not this map of local slopes but directly the set of raw sub-images. In practice, these sub-images generate a large data flow and are therefore not generally stored on a disk: in imaging through turbulence, the wavefront has to be sampled at several tens, or even hundreds, of Hertz.

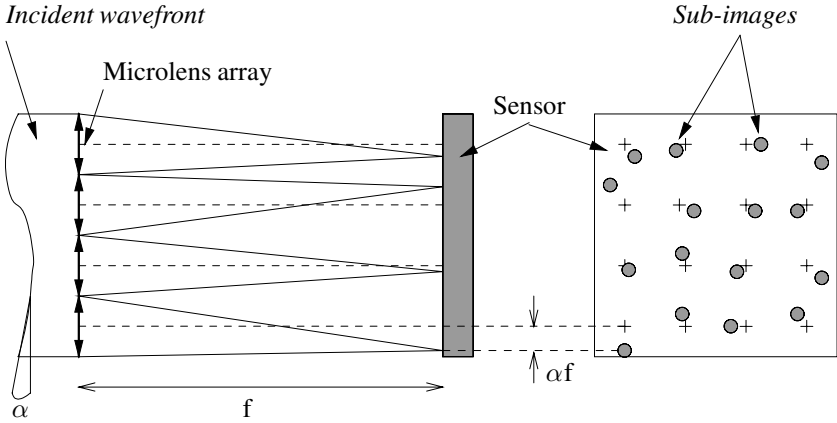


Figure 10.3. Principle of the Hartmann-Shack sensor

The unknown phase at instant t , denoted φ_t , is generally expanded into Zernike polynomials [NOL 76] and the coefficients of this expansion are denoted ϕ_t^q :

$$\varphi_t(\mathbf{r}) = \sum_q \phi_t^q Z_q(\mathbf{r}) \quad (10.9)$$

where \mathbf{r} is the current point in the pupil. The direct problem can then be put in the form:

$$\mathbf{s}_t = \mathbf{D}\phi_t + \mathbf{b}'_t$$

where \mathbf{s}_t is the vector concatenating the $2N_{\text{ml}}^2$ slope measurements (x and y), ϕ_t is the vector of the coordinates of the unknown phase and \mathbf{D} is essentially a sampled derivation operator called the “interaction matrix”.

The noise is generally assumed to be iid Gaussian. The independence among the measurements of the various sub-pupils is natural and the Gaussian character is justified because it results from an estimation of the center of gravity over a “large” number of pixels (typically a few tens).

The solution traditionally used for estimating the phase, in particular under real-time constraints (AO), is the least squares estimation. Matrix $\mathbf{D}^T \mathbf{D}$ is not invertible *a priori*, because the number of measurements is finite ($2N_{\text{ml}}^2$), whereas dimension K of vector ϕ is, in theory, infinite. In practice, even when K is chosen to be slightly smaller than $2N_{\text{ml}}^2$, $\mathbf{D}^T \mathbf{D}$ is ill-conditioned. The usual remedy is to reduce dimension K of the space of unknowns ϕ and to filter a few modes that are not seen or are poorly seen by the sensor. These correspond to the few zero or very small eigenvalues of $\mathbf{D}^T \mathbf{D}$. It is common to take $K \simeq N_{\text{ml}}^2$.

This remedy works correctly because the Zernike polynomial basis is well suited to atmospheric turbulence. Firstly, these polynomials are in an order corresponding to higher and higher spatial frequencies and the turbulence has a PSD that decreases quite fast (see equation (10.3)), so the diagonal of the turbulent phase covariance matrix on the basis of the Zernikes is decreasing. Secondly, this matrix is concentrated around its diagonal. In other words, the Zernike polynomials Z_i are quite close to the eigenmodes (Karhunen-Loève) of the turbulence. In consequence, truncation of the basis $\{Z_i\}$ at a Z_K selects a space containing the most energetic modes. Choosing the best K is quite problematic, as it depends on both the strength of the turbulence r_0 and the noise level on the WFS.

As our statistical knowledge of the turbulence is quite good (see references of section 10.1.3), a Bayesian approach is more appropriate and gives better results.

Since the problem of reconstructing the phase is linear and Gaussian, it leads to an analytical MMSE/MAP estimator, in *covariance* form in [WAL 83] and in *information* form in [BAK 94, SAS 85] (see Chapter 3). MAP estimation of each of the phases corresponds to minimizing the mixed criterion $J_{\text{MAP}} = J_s + J_\phi$, with:

$$J_s = \frac{1}{2}(\mathbf{s}_t - \mathbf{D}\phi_t)^T \mathbf{C}_{b'}^{-1}(\mathbf{s}_t - \mathbf{D}\phi_t) \quad (10.10)$$

and

$$J_\phi = \frac{1}{2}\phi_t^T \mathbf{C}_\phi^{-1}\phi_t \quad (10.11)$$

where $\mathbf{C}_{b'}$ is the covariance matrix of the slope measurement noise (diagonal, with a practically constant diagonal) and \mathbf{C}_ϕ is the covariance matrix of the turbulent phase in the Zernike basis, which is deduced from equation (10.3) [NOL 76] depending only on r_0 . The well known solution is:

$$\hat{\phi}_t = (\mathbf{D}^T \mathbf{C}_{b'}^{-1} \mathbf{D} + \mathbf{C}_\phi^{-1})^{-1} \mathbf{D}^T \mathbf{C}_{b'}^{-1} \mathbf{s}_t. \quad (10.12)$$

This solution takes advantage of our knowledge of the spatial statistics of the turbulence. For use in AO, where the sampling frequency is generally well above $1/\tau_0$, it is judicious to opt for a natural extension of this MMSE estimator that also uses prior knowledge on the time statistics of the turbulence. This extension is the optimal estimator of Kalman filtering [LER 04, PET 05, KUL 06].

10.3.3. Phase retrieval and phase diversity

Phase retrieval consists of estimating the aberrations seen by an instrument from the image of a point source. This comes down to inverting equation (10.1), i.e., estimating its phase φ from a measurement of h . This technique, first used in electron

microscopy [GER 72] then rediscovered in optics [GON 76], has two main limitations: (i) it only works with a point object and (ii) the solution obtained suffers from sign ambiguity and is generally not unique.

Gonsalves [GON 82] has shown that, by using a second image containing a known variation in the aberrations with respect to the first, e.g. a slight defocus, it is possible to estimate the aberrations even if the object is spatially extended and unknown. Moreover, this second image lifts the indetermination mentioned above and the estimated aberrations are unique, in practice, for small aberrations. This technique is called phase diversity by analogy with a technique used in wireless telecommunications.

Phase diversity is used in two different contexts. We may wish to obtain an image of a remote object, e.g. in solar astronomy, or we may wish to measure the aberrations seen by an instrument in order to correct them in real time or off-line. These two problems are connected but nevertheless distinct. In both cases, the basis of the inversion is to estimate the aberrations and object that are the most compatible with the measured images. The conventional approach is a joint estimation of the object and the phase [GON 82] possibly with a regularization for both unknowns. Although this type of joint estimation usually has poor statistical properties, in the specific case of phase diversity, it has been shown that joint estimation leads to a consistent estimator for aberrations [IDI 05]. In addition, a so-called marginal approach that integrates the object out of the problem so as to only estimate the phase has been proposed recently and leads to better robustness with respect to noise [BLA 03].

The interested reader will find a more complete history and a review of the applications of this WFS in [MUG 06], which also contains a detailed study of the two estimators mentioned above.

10.4. Myopic restoration in imaging

10.4.1. *Motivation and noise statistic*

In imaging through turbulence with a monolithic telescope, the data processing needed is essentially a deconvolution. Nevertheless, the estimation or the measurement of the PSF is often imperfect and the best deconvolution results are generally obtained by specifically taking the partial lack of knowledge of the instrument's response into account. This is what we will call myopic deconvolution, which can take different forms depending on whether the turbulence is corrected off-line, by DWFS (section 10.4.2) or in real time by AO (section 10.4.3).

The most used data-fidelity term is the ordinary least squares criterion. In a probabilistic interpretation, this criterion corresponds to the assumption that the noise is

white, Gaussian and stationary (see Chapter 3):

$$J_y(\mathbf{x}) = \frac{1}{2\sigma_b^2} \|\mathbf{H}\mathbf{x} - \mathbf{y}\|^2 \quad (10.13)$$

where \mathbf{x} is the observed object, \mathbf{y} the recorded image, \mathbf{H} the imaging operator and σ_b the standard deviation of the noise. In astronomical imaging, this interpretation is generally a crude approximation, except for a large bright object, as the predominant noise is generally photonic and thus follows Poisson statistics leading to the following data-fidelity term:

$$J_y(\mathbf{x}) = \sum_{\ell, m} (\mathbf{H}\mathbf{x} - \mathbf{y} \log \mathbf{H}\mathbf{x})(\ell, m) \quad (10.14)$$

This non-quadratic criterion can cause practical difficulties for the minimization when gradient-based numerical methods are used. What is more, in very dark parts of the image, the electronic noises of the sensor become non-negligible relative to the photonic noise and fine modeling of the noise must take the simultaneous presence of noise from the sensor (typically a CCD device) and photonic noise into account. These specific difficulties will be examined more closely in Chapter 14.

A good compromise between fine modeling of the noise and efficient minimization can be obtained as follows. A quadratic approximation of (10.14) is deduced first, which corresponds to purely photonic noise. For an image that is not too dark (in practice, ten or so photons per pixel can suffice), the approximation $\mathbf{H}\mathbf{x} - \mathbf{y} \ll \mathbf{y}$ can be taken and (10.14) expanded to the second order. The result corresponds to white, non-stationary, Gaussian noise with variance equal to the image at each point. Then, by simply summing the variances, a data-fidelity criterion is obtained that models the simultaneous presence of sensor and photonic noise [MUG 04]:

$$J_y(\mathbf{x}) = \sum_{\ell, m} \frac{1}{2(\sigma_{\text{ph}}^2(\ell, m) + \sigma_{\text{det}}^2)} |(\mathbf{H}\mathbf{x})(\ell, m) - \mathbf{y}(\ell, m)|^2 \quad (10.15)$$

where $\sigma_{\text{ph}}^2(\ell, m) = \max\{\mathbf{y}(\ell, m), 0\}$ is an estimator of the variance of the photonic noise at each pixel and σ_{det}^2 the variance of the sensor noise, estimated beforehand.

10.4.2. Data processing in deconvolution from wavefront sensing

10.4.2.1. Conventional processing of short-exposure images

In this subsection, we describe non-myopic multiframe deconvolution; in other words, we consider that the PSFs deduced from the WFS measurements are true. In the next subsection, we show how a myopic deconvolution, i.e., the joint processing of WFS data and images, can improve the estimation of the observed object.

We have a series of N_{im} short-exposure images of an object that is smaller than the isoplanatic patch. The equation of the discretized direct problem can be written:

$$\mathbf{y}_t = \mathbf{h}_t \star \mathbf{x} + \mathbf{b}_t = \mathbf{H}_t \mathbf{x} + \mathbf{b}_t, \quad 1 \leq t \leq N_{im} \quad (10.16)$$

where \mathbf{x} and \mathbf{y}_t are the discretized object and image respectively at time t and where the PSF \mathbf{h}_t is related to the phase ϕ_t in the pupil at the same instant by equation (10.1). We also have wavefront measurements, in this case Hartmann-Shack slope measurements \mathbf{s}_t associated with each image.

The conventional DWFS data processing is sequential: first we estimate the phases ϕ_t via (10.12), then deduce the PSFs \mathbf{h}_t via (10.1) and finally estimate the object by multiframe deconvolution. The details of this sequential processing are given below.

The image processing used in the early days of the DWFS technique was a simple multiframe least squares method [PRI 90]; the solution was thus the multiframe inverse filter, which in practice had to be regularized by adding a small constant to the denominator in the Fourier domain. A better approach is to explicitly regularize the criterion to be minimized. For objects with clearly marked edges such as artificial satellites, the regularization criterion to be used is the L_2L_1 of equation (10.8).

Using the Bayesian framework presented in Chapter 3, we estimate the object in the MAP sense. Two considerations allow the likelihood of the set of images to be simplified: firstly, the noise is independent between images and, secondly, the delay between successive acquisitions is generally longer than the typical turbulence evolution time. The likelihood can thus be rewritten as the product of the likelihoods of the individual images, each being conditioned simply by the object and the phase at the same instant. The estimate of the object is then the one that minimizes

$$J_{\text{MAP}}^x(\mathbf{x}) = \sum_{t=1}^{N_{im}} J_y(\mathbf{x}; \phi_t, \mathbf{y}_t) + J_x(\mathbf{x}), \quad (10.17)$$

where $J_y(\mathbf{x}; \phi_t, \mathbf{y}_t) = -\log p(\mathbf{y}_t | \mathbf{x}, \phi_t)$. In practice, for both simulations (see section 10.4.2.3) and experimental data (see section 10.4.2.4), the data-fidelity term used for J_y will be the least squares term of equation (10.13). The minimization is performed numerically on the object variables and the presence of the ϕ_t in the J_{MAP}^x criterion above is simply a reminder of the dependence of the criterion on the phase.

10.4.2.2. Myopic deconvolution of short-exposure images

In conventional DWFS data processing, the information concerning the wavefronts is extracted from the WFS data only, not from the images. And yet there is exploitable information on the PSF in the short-exposure images, as proved by the results some authors [SCH 93, THI 95] have obtained by blind deconvolution (i.e., without WFS, but using the models (10.1) and (10.9)).

However, the criteria to be minimized in blind deconvolution generally have local minima and parametrizing the PSF by the pupil phase is not sufficient to ensure that the solution is unique. This is why WFS data should certainly not be ignored but should instead be used in conjunction with the images.

Myopic deconvolution consists of searching for the most probable object \mathbf{x} and turbulent phases ϕ_t jointly, given the images \mathbf{y}_t , the WFS measurements \mathbf{s}_t and the prior information on \mathbf{x} and ϕ_t [MUG 01]. Using Bayes' rule and the same independence hypotheses as in section 10.4.2.1, it can be shown that the estimates $(\hat{\mathbf{x}}, \{\hat{\phi}_t\})$ in the joint MAP sense are those that minimize:

$$J_{\text{MAP}}(\mathbf{x}, \{\phi_t\}) = \sum_{t=1}^{N_{\text{im}}} J_y(\mathbf{x}, \phi_t; \mathbf{y}_t) + J_x(\mathbf{x}) + \sum_{t=1}^{N_{\text{im}}} J_s(\phi_t; \mathbf{s}_t) + \sum_{t=1}^{N_{\text{im}}} J_\phi(\phi_t)$$

where:

- the J_y are the image-fidelity terms; $J_y(\mathbf{x}, \phi_t; \mathbf{y}_t)$ is the anti-log-likelihood of the t th image; it is now a function of the object *and* the phases;
- $J_x(\mathbf{x})$ is the object prior, which, in what follows, will be the L_2L_1 model of equation (10.8);
- the J_s are the fidelity to WFS data terms; with the hypotheses used, they are quadratic and given by equation (10.10);
- the J_ϕ are the *a priori* terms on the phases given by equation (10.11).

The criterion is minimized by a method based on conjugate gradients, alternating minimizations on the object \mathbf{x} (for the current phase estimate) and on the set of phases ϕ_t (for the current object estimate).

To speed up the minimization and also to avoid, in practice, the local minima often present in joint criteria, the initial object and phases are taken to be the MAP estimates obtained in the sequential processing described in section 10.4.2.1.

10.4.2.3. Simulations

A set of 100 images were simulated with the associated WFS measurements. The 100 wavefronts were obtained by a modal method [ROD 90] in which each phase is expanded on a basis of Zernike polynomials (see equation (10.9)) and follows Kolmogorov statistics (see equation (10.3)). The turbulence strength corresponds to a ratio $D/r_0 = 10$. Each of the turbulent wavefronts is used to calculate a short-exposure image of dimensions 128×128 , sampled at the Shannon frequency using equations (10.1) and (10.16). The noise added to the images is white, Gaussian and stationary with a variance equal to the mean flux to be simulated, i.e., $10^4/128^2 = 0.61$ photon/pixel. Figure 10.4 shows the object, which is a numerical model of the SPOT satellite, and one of the 100 simulated images. The corresponding PSF is the

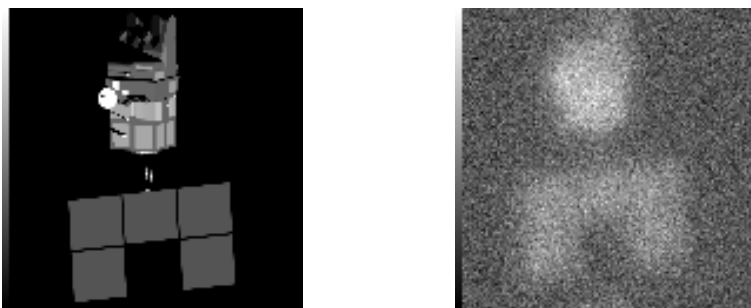


Figure 10.4. *Original object (SPOT satellite, left) and one of the 100 short-exposure images ($D/r_0 = 10$, right)*

image on the left of Figure 10.2. The simulated WFS is a Hartmann-Shack having 20×20 sub-apertures, without central obscuration. White Gaussian noise is added to the local slopes of the wavefront so that the SNR of the slopes measured, defined as the variance of the slopes over the variance of the noise, is 1.

Figure 10.5 compares the results of the sequential and myopic estimations for the same L_2L_1 prior on the object (equation (10.8)), associated with a positivity constraint⁴. On the left, the non-myopic restoration is the MAP estimation of the wavefronts followed by a restoration with PSFs deduced from the estimated wavefronts and gives an MSE with the actual object of 0.45 photon (per pixel). On the right, the joint estimation gives an MSE of 0.39 photon.

In addition, the myopic estimation also allows the quality of the reconstructed wavefronts to be improved [MUG 01].

10.4.2.4. *Experimental results*

The processing methods described above were applied to ten experimental images of the double star Capella recorded on 8th November 1990 with the DWFS system of ONERA, installed on the 4.20 m diameter William Herschel telescope (La Palma, Canary Islands). The experimental conditions were the following: a flux of 67,500 photons per image, an exposure time of 5 ms, a D/r_0 of 13 and a SNR of 5 on the WFS. The WFS was a Hartmann-Shack with 29×29 sub-apertures, 560 of which were used.

Figure 10.6, taken from [MUG 01], shows the results of the deconvolution. On the left, the sequential processing consisted of an estimation of the wavefronts by MAP, then a quadratic image restoration. The binary nature of Capella is visible, but is

4. Our thanks to Clélia Robert for processing the DWFS data.

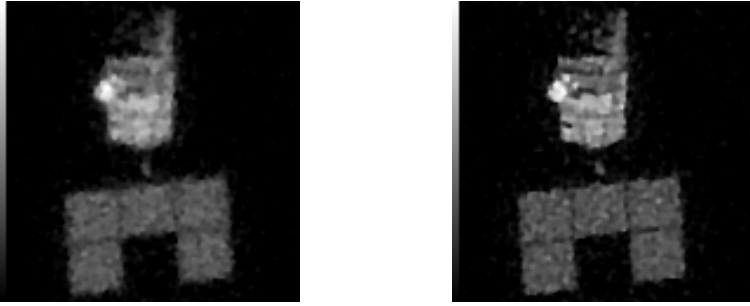


Figure 10.5. Object restored by non-myopic (left) and myopic (right) estimation. In both cases, an L_2L_1 prior and an object positivity constraint were used. The MSE with the actual object are 0.45 and 0.39 photon respectively



Figure 10.6. Deconvolved experimental images of Capella: left, estimation of wavefronts by MAP then quadratic deconvolution; right, myopic deconvolution. In both cases, the prior used was Gaussian with a constant PSD deduced from the measured flux, with a positivity constraint

almost drowned in strong fluctuations. On the right, the myopic deconvolution has eliminated almost all the artefacts of the non-myopic deconvolution. In both cases, the same quadratic object regularization with a positivity constraint was used, with a constant PSD whose value was deduced from the measured flux.

10.4.3. Restoration of images corrected by adaptive optics

10.4.3.1. Myopic deconvolution of images corrected by adaptive optics

Long-exposure images corrected by AO must be deconvolved, since the correction is only partial [CON 94]. If we take the PSF as known, the object estimated in the MAP sense, denoted $\hat{\mathbf{x}}_{\text{MAP}}$, is the one that maximizes $p(\mathbf{x} | \mathbf{y}; \mathbf{h})$, and thus minimizes

$J_y(\mathbf{x}; \mathbf{h}, \mathbf{y}) + J_x(\mathbf{x})$. The most usual method for estimating the PSF is to record the image of a star just before or just after the image of the object of interest. This star image may be noticeably different from the PSF corresponding to the image we are interested in for a variety of reasons: first, the turbulence changes with time [CON 98a]; thus, the response of the AO may be different when going from a spatially extended object to a point object, even if the star is of the same magnitude as the object, since the wavefront sensing error increases with the extent of the object; and finally, there is noise on the star image itself. A method has been proposed and validated for estimating the turbulent part of the long-exposure transfer function corrected by the AO from measurements of the residual wavefront of the control loop [VER 97a, VER 97b]. Nevertheless, apart from the fact that the static or slowly varying aberrations of the telescope may not be properly known, the accuracy of this estimation of the transfer function is limited by the noise on the WFS. Thus, it is often necessary to consider that the PSF is imperfectly known.

Many authors have tackled the problem of deconvolving an image degraded by turbulence with unknown PSF. Ayers and Dainty [AYE 88] used a Gerchberg-Saxton-Papoulis algorithm [GER 72] and came up against problems of convergence with this type of algorithm. Others have used maximum likelihood methods, with an EM algorithm [HOL 92] or minimization of an explicit criterion [JEF 93, LAN 92, LAN 96, THI 95]. They generally recognize the need for regularization rather than just positivity (of the object and the PSF) and, in particular, have introduced a (legitimate) limited bandwidth constraint on the PSF through an ad hoc prior [HOL 92, JEF 93].

The Bayesian framework allows this joint estimation (called myopic estimation) of the object and the PSF to be made with a natural regularization for the PSF and without having to adjust any additional hyperparameters. The joint MAP estimator is given by:

$$\begin{aligned} (\hat{\mathbf{x}}, \hat{\mathbf{h}}) &= \arg \max_{\mathbf{x}, \mathbf{h}} p(\mathbf{x}, \mathbf{h} | \mathbf{y}) = \arg \max_{\mathbf{x}, \mathbf{h}} p(\mathbf{y} | \mathbf{x}, \mathbf{h}) \times p(\mathbf{x}) \times p(\mathbf{h}) \\ &= \arg \min_{\mathbf{x}, \mathbf{h}} (J_y(\mathbf{x}, \mathbf{h}; \mathbf{y}) + J_x(\mathbf{x}) + J_h(\mathbf{h})) \end{aligned}$$

The long-exposure PSF can be considered as the sum of a large number of independent short-exposure PSFs, and thus modeled by a Gaussian prior (truncated to positive values). We also assume that the difference between the PSF and the mean PSF is approximately stationary. The regularization of the PSF is thus a quadratic penalization of the transfer function, which is independent between frequencies [CON 98b, FUS 99, MUG 04]:

$$J_h(\mathbf{h}) = \frac{1}{2} \sum_f |\hat{\mathbf{h}}(\mathbf{u}) - \hat{\mathbf{h}}_m(\mathbf{u})|^2 / S_h(\mathbf{u})$$

where $\hat{\mathbf{h}}_m = E(\hat{\mathbf{h}})$ is the mean transfer function and $S_h = E(|\hat{\mathbf{h}}(\mathbf{u}) - \hat{\mathbf{h}}_m(\mathbf{u})|^2)$ the energy spectral density (ESD) of the PSF. Note that S_h is zero beyond the cut-off frequency of the telescope and that this regularization, in particular, forces \mathbf{h} to comply with the limited bandwidth constraint.

In practice, the mean transfer function and the ESD of the PSF are estimated by replacing the expectations in their definitions by empirical means on the various images of the star acquired before or after the object of interest. If only a single image of the star is available, the expectation can be replaced by a circular mean in the Fourier domain because of the isotropy of the quantities to be estimated.

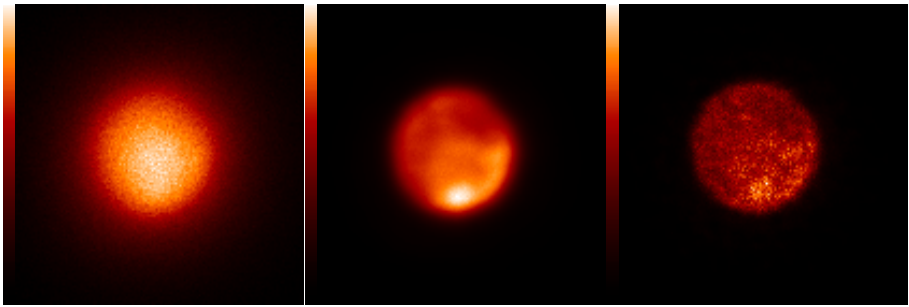
In order to be able to restore objects with a large dynamic range, which are frequent in astronomy, the data-fidelity term J_y must include fine modeling of the noise, such as the mixture of photonic and electronic noise of equation (10.15), rather than a simple least squares. The regularization criterion J_x used here is the L_2L_1 model of equation (10.8), which is well suited to objects with sharp edges such as planets and asteroids.

The restoration method known as MISTRAL [MUG 04] combines the myopic estimation of the object and PSF described earlier with the white inhomogeneous data-fidelity term and the L_2L_1 regularization just mentioned. This method was used to obtain the deconvolution results presented below. The criterion was minimized by the conjugate gradient method, jointly on the object and PSF variables. A positivity constraint was added on \mathbf{x} and on \mathbf{h} .

10.4.3.2. *Experimental results*

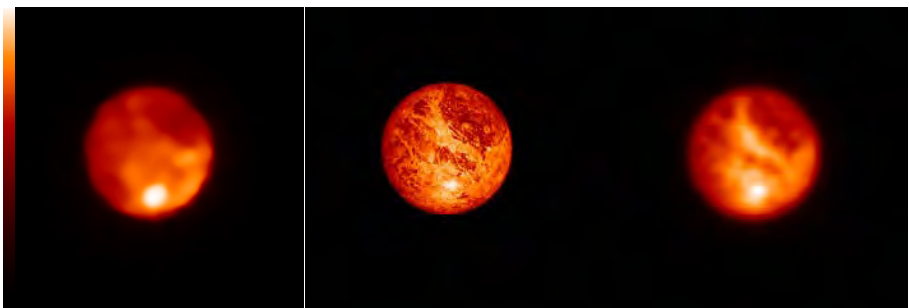
Figure 10.7a shows an AO-corrected long-exposure image of Ganymede, a satellite of Jupiter. This image was recorded on 28th September 1997 on the ONERA AO system installed on the 1.52 m telescope of the Haute-Provence observatory. This system has an 80 Hz passband; it comprises a Hartmann-Shack wavefront sensor with 9×9 sub-apertures (64 of which are active) and a deformable mirror with 10×10 piezo actuators, 88 of which are active. The imaging wavelength is $\lambda = 0.85 \mu\text{m}$ and the exposure time 100 s. The total estimated flux is $8 \cdot 10^7$ photons and the estimated D/r_x ratio is 23. The total field is 7.9 arcsec, only half of which is shown here. The mean PSF and its ESD were estimated from 50 recorded images of a nearby bright star. Figures 10.7b and c show the restorations obtained by the Richardson-Lucy algorithm (maximum likelihood for a Poisson noise), interrupted at 200 and 3,000 iterations respectively. In the first case, the restored image is quite blurred and shows ringing, and in the second case, the noise dominates the restoration. The image of Figure 10.8a illustrates myopic deconvolution [MUG 04] with an L_2L_1 prior⁵. Figure 10.8b shows a wide-band synthetic image obtained from photos taken by a NASA/JPL space probe (see

5. Our thanks to Thierry Fusco for processing the AO images.



(a) image corrected by AO (b) Richardson-Lucy, 200 it. (c) Richardson-Lucy, 3,000 it.

Figure 10.7. (a) *Observation of Ganymede with the ONERA AO system on 28th September 1997*; (b) *Richardson-Lucy restoration interrupted at 200 iterations*; (c) *at 3,000 iterations*



(a) L_2L_1 myopic deconvolution (b) JPL database (NASA/JPL/Caltech) (c) image (b) + PSF of perfect telescope

Figure 10.8. (a) L_2L_1 myopic deconvolution of the image of Ganymede of Figure 10.7; (b) for comparison, a wideband synthetic image obtained from the NASA/JPL database; (c) same synthetic image convolved by the perfect PSF of a 1.52 m-diameter telescope

<http://space.jpl.nasa.gov/>) as it passed near Ganymede. The comparison shows that many features of Ganymede have been correctly restored. A fairer comparison is to examine the myopic deconvolution performed by MISTRAL together with the image of Figure 10.8b convolved with the perfect PSF of a 1.52 m telescope, presented in Figure 10.8c.

Figure 10.9 shows three images of Neptune recorded at half-hour intervals on 6th July 1998 with the curvature-based adaptive optics system of the Institute for Astronomy of the University of Hawaii⁶ called Hokupa'a. This system, which was in operation until 2003, had 36 actuators and was installed on the Canada-France-Hawaii (CFH) 3.6 m telescope. It produced the first high-resolution infrared images of Neptune in November 1997 and July 1998 [ROD 98]. The imaging wavelength was $1.72\ \mu\text{m}$, which is situated in a methane absorption band. The exposure time was 10 minutes per image. The images restored by myopic deconvolution with an prior are shown in Figure 10.10 [CON 00]. The image of a star near Neptune was also recorded in order to estimate the mean PSF and the ESD of the PSF by the circular mean in the Fourier domain. Because the atmosphere of Neptune is very dark at the imaging wavelength, these images show the fine structures of the cloud bands in the upper atmosphere with good contrast. Note, in particular, that the fine structures of the cloud bands can be followed from image to image as the planet turns. This was the first time it had been possible to study the details of Neptune's atmospheric activity from the ground.



Figure 10.9. Images of Neptune obtained at 30-minute intervals on 6th July 1998 with the Hokupa'a adaptive optics system on the Canada-France-Hawaii telescope. The imaging wavelength was $1.72\ \mu\text{m}$ and the exposure time for each image was 10 minutes

10.4.4. Conclusion

The restoration of images degraded by turbulence and corresponding to a convolutive imaging model is now a well mastered technique. The observation systems currently being developed have more complex acquisition modes for which processing will no doubt be largely called upon. Representative examples are the wide-field

6. Our thanks to François and Claude Roddier for so kindly providing us with these images.

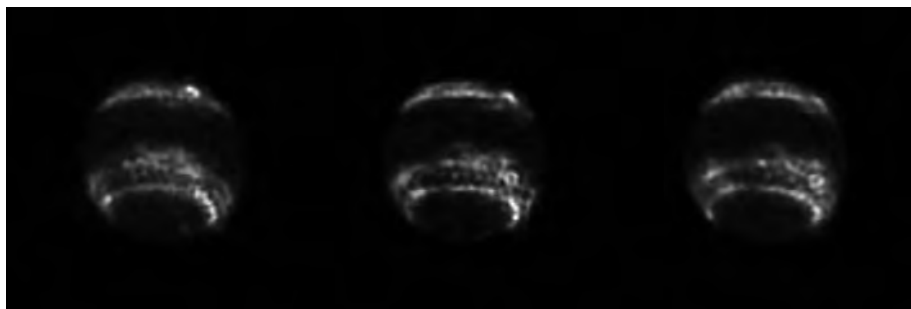


Figure 10.10. *Images of Figure 10.9 restored by myopic deconvolution with L_2L_1 prior*

systems with so-called multiconjugate AO [CON 05], for which the PSF cannot be considered as spatially invariant, and the systems such as SPHERE [DOH 06] or GPI that combine high-performance AO (known as extreme AO) with a coronagraph in the aim of detecting exoplanets. For such systems, imaging is fundamentally non-convolutive and specific processing has to be developed. AO has also found an application in retinal imaging in recent years and several teams are developing operational systems (see for example [GLA 02, GLA 04] and the references therein). In this context, the image measured and the object to be restored are three-dimensional [CHE 07].

10.5. Image reconstruction in optical interferometry (OI)

This section is devoted to the reconstruction of images from data coming from a correlation interferometer. Section 10.1.2.2 presented the measurement principle and the type of transfer function associated with these systems. The observation model is presented more precisely in section 10.5.1, then sections 10.5.2 and 10.5.3 describe the main avenues towards image reconstruction at present. Finally, results on synthetic and real data are the subject of section 10.5.4.

10.5.1. Observation model

Let us consider a two-telescope interferometer. The positions of the telescopes in a plane normal to the observation direction are \mathbf{r}_1 and \mathbf{r}_2 . Due to the Earth's rotation, the observation direction changes with time and baseline $\mathbf{r}_2 - \mathbf{r}_1$ thus varies, as does the spatial frequency corresponding to it:

$$\mathbf{u}_{12}(t) \triangleq (\mathbf{r}_2(t) - \mathbf{r}_1(t)) / \lambda.$$

When a *complete* interferometer array is used, i.e., one in which all the possible two-telescope baselines can be formed simultaneously, there are $N_b = N_t(N_t - 1)/2$ measurement frequencies given by

$$\mathbf{u}_{k\ell}(t) = (\mathbf{r}_\ell(t) - \mathbf{r}_k(t)) / \lambda, \quad 1 \leq k < \ell \leq N_t.$$

Each baseline (T_k, T_ℓ) produces interference fringes. The measurement of the contrast and position of these fringes defines the complex visibility $y_{k\ell}^{\text{data}}(t)$ and gives information on the modulus $a_{k\ell}(\mathbf{x}, t)$ and the phase $\phi_{k\ell}(\mathbf{x}, t)$ of the FT of object \mathbf{x} at spatial frequency $\mathbf{u}_{k\ell}$.

When the instrument is calibrated, generally by prior observation of an unresolved object, we no longer have to consider the possibly complex gains, that come into measurement (10.7) presented in section 10.1.2.2. On the other hand, the effects of turbulence, which vary rapidly, cannot be precalibrated. We can thus take it that the main perturbation affecting the *short-exposure* phase measurement is an additive term $\varphi_\ell(t) - \varphi_k(t)$ known as the differential piston term:

$$\phi_{k\ell}^{\text{data}}(t) = \phi_{k\ell}(\mathbf{x}, t) + \varphi_\ell(t) - \varphi_k(t) + \text{noise} [2\pi] \quad (10.18)$$

where $\phi_{k\ell}^{\text{data}}(t)$ is the phase of $y_{k\ell}^{\text{data}}(t)$. Thus, in matrix form, $\phi^{\text{data}}(t) = \phi(\mathbf{x}, t) + \mathbf{B}\varphi(t) + \text{noise} [2\pi]$, where the *baseline operator* \mathbf{B} has dimensions $N_b \times N_t$.

As mentioned in section 10.1.2.2, the differential piston is the result of the random differences introduced in the optical path between the apertures of the system by turbulence. For a long baseline (relative to the Fried diameter), the optical path difference may be very much greater than the observation wavelength and thus lead to random phase differences much larger than 2π . The aliased perturbation that affects the phase (10.18) is then practically uniformly distributed in $[0, 2\pi]$. In consequence, averaging the phases of short-exposure visibility (10.18) does not improve the signal-to-noise ratio. A solution is to carry out *phase closures* [JEN 58] before the averaging. For any set of three telescopes (T_k, T_ℓ, T_m) the short-exposure visibility phase data is

$$\begin{cases} \phi_{k\ell}^{\text{data}}(t) = \phi_{k\ell}(\mathbf{x}, t) + \varphi_\ell(t) - \varphi_k(t) + \text{noise} \\ \phi_{\ell m}^{\text{data}}(t) = \phi_{\ell m}(\mathbf{x}, t) + \varphi_m(t) - \varphi_\ell(t) + \text{noise} \\ \phi_{mk}^{\text{data}}(t) = \phi_{mk}(\mathbf{x}, t) + \varphi_k(t) - \varphi_m(t) + \text{noise} \end{cases} \quad (10.19)$$

and the turbulent pistons are canceled out in the closure phase defined by:

$$\begin{aligned} \beta_{k\ell m}^{\text{data}}(t) &\triangleq \phi_{k\ell}^{\text{data}}(t) + \phi_{\ell m}^{\text{data}}(t) + \phi_{mk}^{\text{data}}(t) + \text{noise} \\ &= \phi_{k\ell}(\mathbf{x}, t) - \phi_{\ell m}(\mathbf{x}, t) + \phi_{mk}(\mathbf{x}, t) + \text{noise} \\ &= \beta_{k\ell m}(\mathbf{x}, t) + \text{noise} \end{aligned} \quad (10.20)$$

To form this type of expression it is necessary to measure 3 visibility phases simultaneously, and thus to use an array of 3 telescopes or more. For a complete array made up of N_t telescopes, the set of closure phases that can be formed is generated by, for example, the $\beta_{1k\ell}^{\text{data}}(t)$, $1 < k < \ell \leq N_t$, i.e., the closure phases measured on the triangles of telescopes including T_1 . It is easy to see that there are $(N_t - 1)(N_t - 2)/2$ of these independent closure phases. In what follows, the vector grouping together these independent closure phases will be noted β^{data} and a closure operator C is defined such that

$$\beta^{\text{data}} \triangleq C\phi^{\text{data}} = C\phi(x, t) + \text{noise}.$$

The second equation is a matrix version of (10.20): the closure operator cancels the differential pistons, a property that can be written $CB = 0$. It can be shown that this equation implies that the closure operator has a kernel of dimension $N_t - 1$, given by

$$\text{Ker } C = \{\bar{B}\alpha, \alpha \in \mathbb{R}^{N_t-1}\} \quad (10.21)$$

where \bar{B} is obtained by removing the first column from B . The closure phase measurement thus does not allow all the phase information to be measured. This result can also be obtained by counting up the phase unknowns, i.e., $N_t(N_t - 1)/2$ object visibility phases minus the number of independent closures, $(N_t - 1)(N_t - 2)/2$, which gives $N_t - 1$ missing phase data. In other words, optical interferometry through turbulence comes under Fourier synthesis *with partial phase information*. Note that, the more apertures there are in the array, the smaller the proportion of missing information will be.

We are now in a position to define the long-exposure observables of a correlation interferometer:

– *mean square amplitudes* $s^{\text{data}}(t) = \langle (a^{\text{data}}(t + \tau))^2 \rangle_\tau$, in preference to mean moduli as they have an easy-to-calculate bias, which can be subtracted from the measurements;

– *bispectra* $V_{1k\ell}^{\text{data}}(t)$, $k < \ell$, defined by

$$V_{1k\ell}^{\text{data}}(t) = \langle y_{1k}^{\text{data}}(t + \tau) y_{k\ell}^{\text{data}}(t + \tau) y_{\ell 1}^{\text{data}}(t + \tau) \rangle_\tau.$$

The modulus of the bispectrum is redundant with the squares of the amplitudes and is thus not used in image reconstruction. The phases of the bispectra $\beta_{1k\ell}^{\text{data}}(t)$, $k < \ell$ constitute unbiased long-exposure closure phase estimators.

Notation τ expresses the averaging in a time interval around instant t , an interval that must be short enough for the spatial frequency to be considered constant during the integration despite the rotation of the Earth. The integration time also determines the standard deviations of the residual noises on the measurements.

The long-exposure observation model is finally:

$$\begin{cases} \mathbf{s}^{\text{data}}(t) = \mathbf{a}^2(\mathbf{x}, t) + \mathbf{s}^{\text{noise}}(t), & \mathbf{s}^{\text{noise}}(t) \sim \mathcal{N}(0, \mathbf{R}_{\mathbf{s}(t)}) \\ \beta^{\text{data}}(t) = \mathbf{C}\phi(\mathbf{x}, t) + \beta^{\text{noise}}(t), & \beta^{\text{noise}}(t) \sim \mathcal{N}(0, \mathbf{R}_{\beta(t)}) \end{cases} \quad (10.22)$$

Estimating an object from such Fourier data is called *Fourier synthesis*. Matrices $\mathbf{R}_{\mathbf{s}(t)}$ and $\mathbf{R}_{\beta(t)}$ are generally assumed to be diagonal. In terms of prior knowledge, the object we are looking for is positive. Moreover, as visibilities are flux-normalized quantities, it is convenient to work with the constraint of unit flux. The constraints on the object are thus

$$\sum_{k, \ell} x(k, \ell) = 1, \quad \forall k, \ell, \quad x(k, \ell) \geq 0. \quad (10.23)$$

10.5.2. Traditional Bayesian approach

This approach first forms the anti-log-likelihood according to model (10.22)

$$J^{\text{data}}(\mathbf{x}) = \sum_t J^{\text{data}}(\mathbf{x}, t) = \sum_t \chi_{\mathbf{s}(t)}^2(\mathbf{x}) + \chi_{\beta(t)}^2(\mathbf{x}) \quad (10.24)$$

with the notation

$$\chi_{\mathbf{m}(t)}^2(\mathbf{x}) \triangleq (\mathbf{m}^{\text{data}}(t) - \mathbf{m}(\mathbf{x}, t))^T \mathbf{R}_{\mathbf{m}(t)}^{-1} (\mathbf{m}^{\text{data}}(t) - \mathbf{m}(\mathbf{x}, t)),$$

then associates J^{data} with a regularization term such as those presented in section 10.2. The problem thus is to minimize the composite criterion

$$J(\mathbf{x}) = J^{\text{data}}(\mathbf{x}) + J_x(\mathbf{x}) \quad (10.25)$$

obtained under the constraints (10.23). Among the references that adopt this approach for processing optical interferometry data, [THI 03] is one of the most noteworthy.

Such works are based on the use of local descent methods. Unfortunately, criterion J is non-convex. To be more precise, the difficulty of the problem can be summed up in the following three points:

1) the small number of Fourier coefficients makes the problem under-determined: the regularization term can get around this under-determination, e.g. by limiting the high frequencies of the reconstructed object [LAN 98];

2) the turbulence implies phase indetermination. This type of indetermination makes the Fourier synthesis problem non-convex and adding a regularization term does not generally correct the problem;

3) finally, the fact that we have phase modulus measurements with Gaussian noise leads to a non-Gaussian likelihood in \mathbf{x} and a non-convex log-likelihood. This point, which has long been known in the field of radar was identified only very recently in optical interferometry [MEI 05c]. In other words, even if we had all the complex visibility phase measurements instead of just the closure phases, the data fidelity term would still be non-convex.

These characteristics imply that optimizing J by a local descent algorithm can only work if the initialization puts us in the “right” valley of the criterion. The use of a global optimization algorithm has never been proposed in optical interferometry as far as we know. It would no doubt be useful to explore this path as long as the number of variables remained reasonable, in particular in comparison with the very large dimension maps that are reconstructed in radio interferometry.

10.5.3. Myopic modeling

Another approach is to put the problem in terms of missing data; this is phase data that is eliminated by the use of a closure operator, i.e., elements of the kernel of \mathbf{C} (10.21). The *myopic* approach thus consists of finding object \mathbf{x} and missing phase data $\boldsymbol{\alpha}$ jointly. This technique is called *self-calibration* in radio-interferometry [COR 81] and has enabled reliable images to be reconstructed in situations of partial phase indetermination. The first myopic approaches put forward in optical interferometry were strongly influenced by this work [LAN 98]. Recent findings indicate that these transpositions were based on too great a simplification of the measuring procedure belonging to optical interferometry. This section outlines a precise myopic approach applied to OI.

The construction of a myopic model starts from a generalized inverse solution to the phase closure equation of (10.22), using the operator

$$\mathbf{C}^\dagger \triangleq \mathbf{C}^T (\mathbf{C}\mathbf{C}^T)^{-1}.$$

By applying \mathbf{C}^\dagger on the left to (10.22) and (10.21) we have

$$\exists \boldsymbol{\alpha}(t) \mid \mathbf{C}^\dagger \boldsymbol{\beta}^{\text{data}}(t) = \boldsymbol{\phi}(\mathbf{x}, t) + \bar{\mathbf{B}}\boldsymbol{\alpha}(t) + \mathbf{C}^\dagger \boldsymbol{\beta}^{\text{noise}}(t).$$

It is thus tempting to define a pseudo-equation of visibility phase measurement by identifying the last term of the latter equation with a measurement pseudo-noise:

$$\boldsymbol{\phi}^{\text{data}}(t) = \underbrace{\boldsymbol{\phi}(\mathbf{x}, t) + \mathbf{B}\boldsymbol{\alpha}(t)}_{\boldsymbol{\phi}(\mathbf{x}, \boldsymbol{\alpha}(t), t)} + \boldsymbol{\phi}^{\text{noise}}(t). \quad (10.26)$$

This approach is similar to that presented in reference [LAN 01]. Unfortunately, as matrix \mathbf{C}^\dagger is singular, this identification is not rigorously possible and we are led to associate an ad hoc covariance matrix \mathbf{R}_ϕ with the term $\boldsymbol{\phi}^{\text{noise}}(t)$ so as to approximately

fit the statistical behavior of the closures. These problems of covariance approximation are ignored in [LAN 01]. The more recent references [MEI 05a, MUG 07] discuss the possible choices for \mathbf{R}_ϕ and propose the use of the following diagonal matrix:

$$\mathbf{R}_\phi \propto \text{Diag} \{ \mathbf{C}^\dagger \mathbf{R}_\beta \mathbf{C}^{\dagger,T} \}$$

where the expression $\text{Diag} \{ \mathbf{M} \}$ designates the diagonal matrix formed with the diagonal of \mathbf{M} .

Finding a suitable approximation for the covariance of the amplitude measurements (10.22), see [MEI 05a, MUG 07], gives a myopic measurement model, i.e., one that depends on the unknowns \mathbf{x} and α :

$$\begin{cases} \mathbf{a}^{\text{data}}(t) = \mathbf{a}(\mathbf{x}, t) + \mathbf{a}^{\text{noise}}(t), & \mathbf{a}^{\text{noise}}(t) \sim \mathcal{N}(\bar{\mathbf{a}}(t), \mathbf{R}_{\mathbf{a}(t)}) \\ \phi^{\text{data}}(t) = \phi(\mathbf{x}, \alpha(t), t) + \phi^{\text{noise}}(t), & \phi^{\text{noise}}(t) \sim \mathcal{N}(\bar{\phi}(t), \mathbf{R}_{\phi(t)}) \end{cases} \quad (10.27)$$

We now have an explicit model of the phase indetermination noted in section 10.5.2. At this stage, it is possible to envisage using, for example, alternating descent algorithms that successively optimize a regularized criterion coming from (10.27), according to \mathbf{x} and α . However, it is still true that, as this model is given in modulus and phase, it always leads to a data-fidelity term that is non-convex in \mathbf{x} , for fixed α . Below, we briefly present a convex approximation of this model.

From the pseudo-measurements $\mathbf{a}^{\text{data}}(t)$ and $\phi^{\text{data}}(t)$, let us form complex pseudo-visibility

$$\mathbf{y}^{\text{data}}(t) \triangleq \mathbf{a}^{\text{data}}(t) e^{j\phi^{\text{data}}(t)}.$$

The data model is thus

$$\mathbf{y}^{\text{data}}(t) = (\mathbf{a}(\mathbf{x}, t) + \mathbf{a}^{\text{noise}}(t)) e^{j(\phi(\mathbf{x}, \alpha(t), t) + \phi^{\text{noise}}(t))}.$$

The noise on these measurements, although additive and Gaussian in modulus and phase separately, is not a complex additive Gaussian noise. In reference [MEI 05b], the authors show how this distribution can be best approximated by an additive Gaussian noise $\mathbf{y}^{\text{noise}}(t)$.

$$\mathbf{y}^{\text{data}}(t) = \mathbf{y}(\mathbf{x}, \alpha(t), t) + \mathbf{y}^{\text{noise}}(t) \quad (10.28)$$

with

$$\mathbf{y}(\mathbf{x}, \alpha(t), t) \triangleq \mathbf{a}(\mathbf{x}, t) e^{j\phi(\mathbf{x}, \alpha(t), t)} \quad (10.29)$$

In general, this approximation leads to a data fitting term J^{pseudo} that is quadratic in the real and imaginary parts of the residuals $y_{k\ell}^{\text{data}}(t) - y_{k\ell}(\mathbf{x}, \alpha(t), t)$. By associating this term with a convex regularization term, we obtain a composite criterion that is convex in \mathbf{x} at fixed α . The WISARD algorithm [MUG 07] makes use of this property by minimizing this composite criterion alternately in \mathbf{x} for the current α and in α for the current \mathbf{x} .

10.5.4. Results

This section presents some results of processing using the WISARD algorithm [MUG 07] based on the myopic approach described in 10.5.3.

10.5.4.1. Processing of synthetic data

The first example takes synthetic interferometric data that was used in the international Imaging Beauty Contest organized by P. Lawson for the International Astronomical Union (IAU) [LAW 04]. This data simulates the observation of the synthetic object shown in Figure 10.11 with the NPOI [NPO] 6-telescope interferometer. The corresponding frequency coverage, shown in Figure 10.11, has a circular structure typical of the *super-synthesis* technique. We recall that super-synthesis consists of repeating the measurements over several instants of measurement (possibly over several nights of observation) so that the same baselines access different spatial frequencies because of the Earth's rotation. In total, there are 195 square visibility modules and 130 closure phases, together with the associated variances.

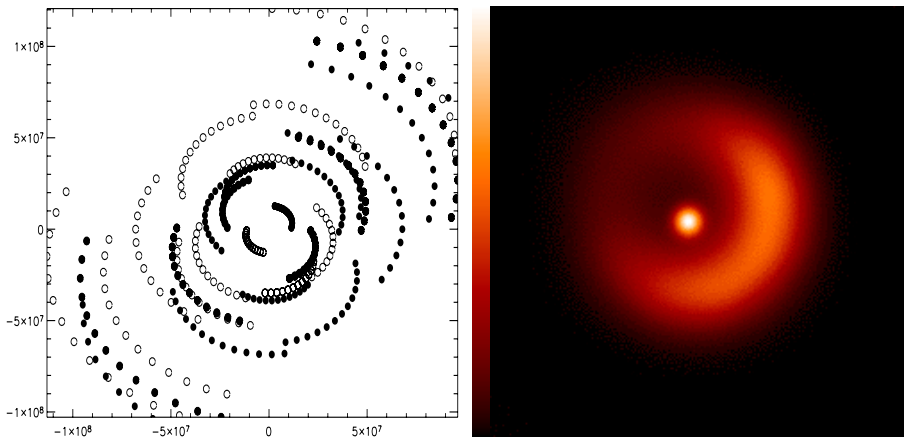


Figure 10.11. Synthetic object (right) and frequency coverage (left) from the Imaging Beauty Contest 2004

Three reconstructions obtained with WISARD are shown in Figure 10.12. On the left is a reconstruction using a quadratic regularization based on a PSD model in $1/|u|^3$ for a weak regularization parameter, in the center a reconstruction with a correct parameter. The latter gives a satisfactory level of smoothing but does not restore the peak in the center of the object. The peak is visible in the under-regularized reconstruction on the left but at the cost of too high a residual variance.

The reconstruction presented on the right is a good trade-off between smoothing and restoration of the central peak thanks to the use of the white prior term introduced

in section 10.2. The goodness of fit of the L_2L_1 reconstruction can be appreciated in Figure 10.13. The crosses show the reconstructed visibility moduli (i.e., of the FT of the reconstructed object at the measurement frequencies) and the squares the moduli of the measured visibilities. The difference between the two, weighted by 10 times the standard deviation of the moduli, is shown as the line. The mean value of this difference is 0.1, which shows a good fit, to one standard deviation.

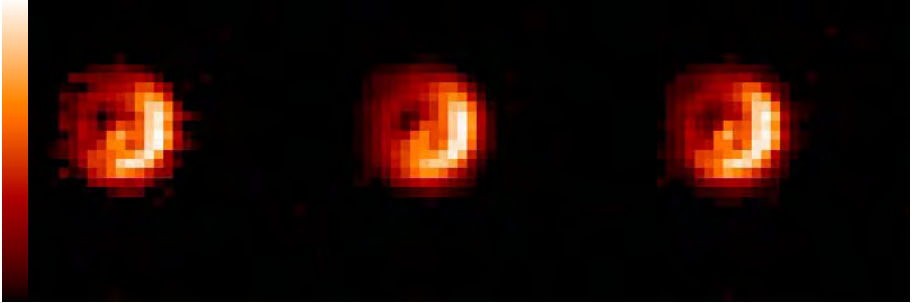


Figure 10.12. Reconstructions with WISARD. Left: under-regularized quadratic model; center: quadratic model with correct regularization parameter; right: white L_2L_1 model of equation (10.8)

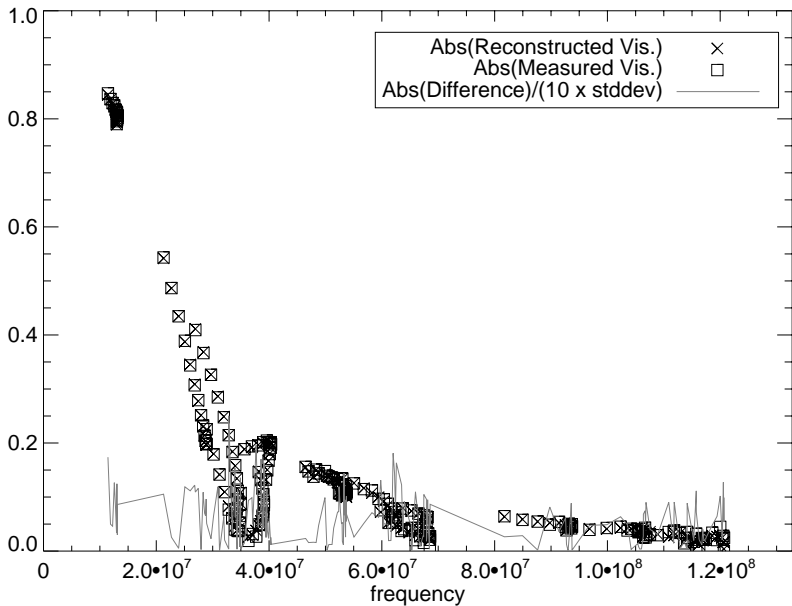


Figure 10.13. Goodness of fit at WISARD convergence

10.5.4.2. *Processing of experimental data*

Here, we present the reconstruction of the star χ Cygni from experimental data using the WISARD [MUG 07] algorithm. The data was obtained by S. Lacour and S. Meimon under the leadership of G. Perrin during a measuring campaign on the IOTA interferometer [IOT] in May 2005. As already mentioned, each measurement has to be calibrated by observation of an object that acts as a point source at the instrument's resolving power. The calibrators chosen were HD 180450 and HD 176670.

χ Cygni is a Mira-type star, Mira itself being an example of such stars. Perrin *et al.* [PER 04] propose a laminar model of Mira-type stars, composed of a photosphere, an empty layer, and a fine molecular layer. The aim of the mission was to obtain images of χ Cygni in the H band ($1.65 \mu\text{m} \pm 175 \text{ nm}$) and, in particular, to highlight possible dissymmetries in the structure of the molecular layer.

Figure 10.14 shows, on the left, the $u - v$ coverage obtained, i.e., the set of spatial frequencies measured, multiplied by the observation wavelength. As the sky is habitually represented with the west on the right, the coordinates used are, in fact, $-u, v$. The domain of the accessible $u - v$ plane is constrained by the geometry of the interferometer and the position of the star in the sky. The “hour-glass” shape is characteristic of the IOTA interferometer, and entails non-uniform resolution that affects the image reconstruction, shown on the right. The reconstructed angular field has sides of 30 milliarcseconds. In addition to the positivity constraint, the regularization used is the white L_2L_1 criterion described in section 10.2. The interested reader will find an astrophysical interpretation of this result in [LAC 07].

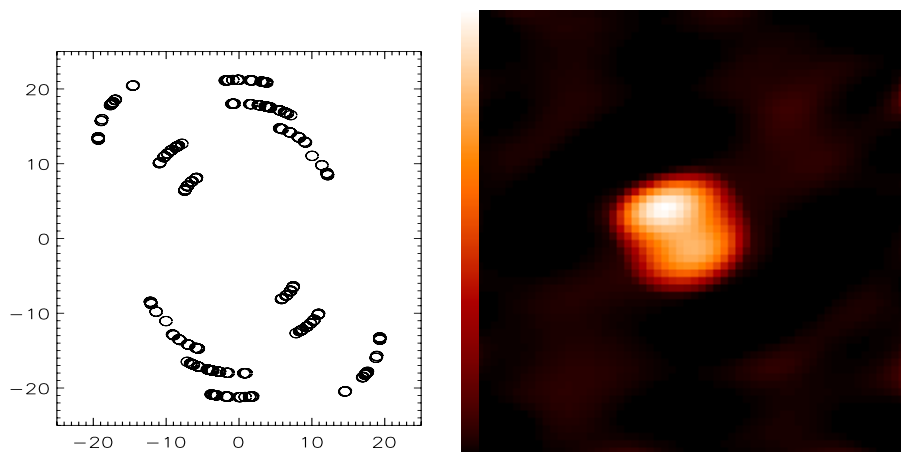


Figure 10.14. Frequency coverage (left) and reconstruction of the star χ Cygni (right)

10.6. Bibliography

- [AYE 88] AYERS G. R., DAINTY J. C., "Iterative blind deconvolution and its applications", *Opt. Lett.*, vol. 13, p. 547-549, 1988.
- [BAK 94] BAKUT P. A., KIRAKOSYANTS V. E., LOGINOV V. A., SOLOMON C. J., DAINTY J. C., "Optimal wavefront reconstruction from a Shack-Hartmann sensor by use of a Bayesian algorithm", *Opt. Commun.*, vol. 109, p. 10-15, June 1994.
- [BAL 86] BALDWIN J. E., HANIFF C. A., MACKAY, WARNER P. J., "Closure phase in high-resolution optical imaging", *Nature (London)*, vol. 320, p. 595-597, Apr. 1986.
- [BAL 96] BALDWIN J. E., BECKETT M. G., BOYSEN R. C., BURNS D., BUSCHER D. F., COX G. C., HANIFF C. A., MACKAY C. D., NIGHTINGALE N. S., ROGERS J., SCHEUER P. A. G., SCOTT T. R., TUTHILL P. G., WARNER P. J., WILSON D. M. A., WILSON R. W., "The first images from an optical aperture synthesis array: mapping of Capella with COAST at two epochs", *Astron. Astrophys.*, vol. 306, p. L13+, Feb. 1996.
- [BEN 97] BENSON J. A., HUTTER D. J., ELIAS II N. M., BOWERS P. F., JOHNSTON K. J., HAJIAN A. R., ARMSTRONG J. T., MOZURKEWICH D., PAULS T. A., RICKARD L. J., HUMMEL C. A., WHITE N. M., BLACK D., DENISON C. S., "Multichannel optical aperture synthesis imaging of zeta1 Ursae majoris with the Navy Prototype Optical Interferometer", *Astron. J.*, vol. 114, p. 1221-1226, Sep. 1997.
- [BER 03] BERGER J.-P., HAGUENAUER P., KERN P. Y., ROUSSELET-PERRAUT K., MALBET F., GLUCK S., LAGNY L., SCHANEN-DUPORT I., LAURENT E., DELBOULBE A., TATULLI E., TRAUB W. A., CARLETON N., MILLAN-GABET R., MONNIER J. D., PEDRETTI E., RAGLAND S., "An integrated-optics 3-way beam combiner for IOTA", in TRAUB W. A. (Ed.), *Interferometry for Optical Astronomy II. Proc. SPIE, Vol. 4838, pp. 1099-1106 (2003)*, p. 1099-1106, Feb. 2003.
- [BLA 03] BLANC A., MUGNIER L. M., IDIER J., "Marginal estimation of aberrations and image restoration by use of phase diversity", *J. Opt. Soc. Am. (A)*, vol. 20, num. 6, p. 1035-1045, 2003.
- [BOR 93] BORN M., WOLF E., *Principles of Optics*, Pergamon Press, 6th (corrected) edition, 1993.
- [BRE 96] BRETTE S., IDIER J., "Optimized single site update algorithms for image deblurring", in *Proc. IEEE ICIP*, Lausanne, Switzerland, p. 65-68, 1996.
- [CAS 97] CASSAING F., Analyse d'un instrument à synthèse d'ouverture optique : méthodes de cophasage et imagerie à haute résolution angulaire, PhD thesis, University of Paris XI, France, Dec. 1997.
- [CHE 07] CHENEGROS G., MUGNIER L. M., LACOMBE F., GLANC M., "3D phase diversity: a myopic deconvolution method for short-exposure images. Application to retinal imaging", *J. Opt. Soc. Am. (A)*, vol. 24, num. 5, p. 1349-1357, May 2007.
- [CON 94] CONAN J.-M., Etude de la correction partielle en optique adaptative, PhD thesis, University of Paris XI, France, Oct. 1994.

- [CON 95] CONAN J.-M., ROUSSET G., MADEC P.-Y., “Wave-front temporal spectra in high-resolution imaging through turbulence”, *J. Opt. Soc. Am. (A)*, vol. 12, num. 12, p. 1559-1570, July 1995.
- [CON 98a] CONAN J.-M., FUSCO T., MUGNIER L. M., KERSALE E., MICHAU V., “Deconvolution of adaptive optics images with imprecise knowledge of the point spread function: results on astronomical objects”, in BONACCINI D. (Ed.), *Astronomy with adaptive optics: present results and future programs*, num. 56 in ESO Conf. and Workshop Proc., Sonthofen, Germany, p. 121-132, Sep. 1998.
- [CON 98b] CONAN J.-M., MUGNIER L. M., FUSCO T., MICHAU V., ROUSSET G., “Myopic deconvolution of adaptive optics images using object and point spread function power spectra”, *Appl. Opt.*, vol. 37, num. 21, p. 4614-4622, July 1998.
- [CON 00] CONAN J.-M., FUSCO T., MUGNIER L., MARCHIS F., RODDIER C., F. RODDIER, “Deconvolution of adaptive optics images: from theory to practice”, in WIZINOWICH P. (Ed.), *Adaptive Optical Systems Technology*, vol. 4007, Munich, Germany, Proc. Soc. Photo-Opt. Instrum. Eng., p. 913-924, 2000.
- [CON 05] CONAN J.-M., ROUSSET G. (Eds.), *Multi-Conjugate Adaptive Optics for Very Large Telescopes Dossier*, vol. 6 fascicule 10 of *C. R. Physique, Académie des Sciences*, Elsevier, Paris, Dec. 2005.
- [COR 81] CORNWELL T. J., WILKINSON P. N., “A new method for making maps with unstable radio interferometers”, *Month. Not. Roy. Astr. Soc.*, vol. 196, p. 1067-1086, 1981.
- [DOH 06] DOHLEN K., BEUZIT J.-L., FELDT M., MOUILLET D., PUGET P., ANTICHI J., BARUFFOLO A., BAUDOZ P., BERTON A., BOCCALETTI A., CARBILLET M., CHARTON J., CLAUDI R., DOWNING M., FABRON C., FEAUTRIER P., FEDRIGO E., FUSCO T., GACH J.-L., GRATTON R., HUBIN N., KASPER M., LANGLOIS M., LONGMORE A., MOUTOU C., PETIT C., PRAGT J., RABOU P., ROUSSET G., SAISSE M., SCHMID H.-M., STADLER E., STAMM D., TURATTO M., WATERS R., WILDI F., “SPHERE: A planet finder instrument for the VLT”, in MCLEAN I. S., IYE M. (Eds.), *Ground-based and Airborne Instrumentation for Astronomy*, vol. 6269, Proc. Soc. Photo-Opt. Instrum. Eng., 2006.
- [FON 85] FONTANELLA J.-C., “Analyse de surface d’onde, déconvolution et optique active”, *J. of Optics (Paris)*, vol. 16, num. 6, p. 257-268, 1985.
- [FRI 65] FRIED D. L., “Statistics of a geometric representation of wavefront distortion”, *J. Opt. Soc. Am.*, vol. 55, num. 11, p. 1427-1435, 1965.
- [FUS 99] FUSCO T., VERAN J.-P., CONAN J.-M., MUGNIER L., “Myopic deconvolution method for adaptive optics images of stellar fields”, *Astron. Astrophys. Suppl.*, vol. 134, p. 1-10, Jan. 1999.
- [GER 72] GERCHBERG R. W., SAXTON W. O., “A practical algorithm for the determination of phase from image and diffraction plane pictures”, *Optik*, vol. 35, p. 237-246, 1972.
- [GLA 02] GLANC M., Applications Ophtalmologiques de l’Optique Adaptative, PhD thesis, University of Paris XI, France, 2002.

- [GLA 04] GLANC M., GENDRON E., LACOMBE F., LAFAILLE D., LE GARGASSON J.-F., P.LÉNA, "Towards wide-field retinal imaging with adaptive optics", *Opt. Commun.*, vol. 230, p. 225–238, 2004.
- [GON 76] GONSALVES R. A., "Phase retrieval from modulus data", *J. Opt. Soc. Am.*, vol. 66, num. 9, p. 961–964, 1976.
- [GON 82] GONSALVES R. A., "Phase retrieval and diversity in adaptive optics", *Opt. Eng.*, vol. 21, num. 5, p. 829–832, 1982.
- [GOO 68] GOODMAN J. W., *Introduction to Fourier Optics*, McGraw-Hill, New York, 1968.
- [GOO 85] GOODMAN J. W., *Statistical Optics*, John Wiley, New York, 1985.
- [GRA 06] GRATADOUR D., ROUAN D., MUGNIER L. M., FUSCO T., CLÉNET Y., GENDRON E., LACOMBE F., "Near-IR AO dissection of the core of NGC 1068 with NaCo", *Astron. Astrophys.*, vol. 446, num. 3, p. 813–825, Feb. 2006.
- [HAN 87] HANIFF C. A., MACKAY C. D., TITTERINGTON D. J., SIVIA D., BALDWIN J. E., "The first images from optical aperture synthesis", *Nature (London)*, vol. 328, p. 694–696, Aug. 1987.
- [HIL 04] HILL J. M., SALINARI P., "The Large Binocular Telescope project", in OSCHMANN JR. J. M. (Ed.), *Ground-based Telescopes. Proc. SPIE, Vol. 5489, pp. 603–614 (2004)*, vol. 5489 of *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, p. 603–614, Oct. 2004.
- [HOL 92] HOLMES T. J., "Blind deconvolution of speckle images quantum-limited incoherent imagery: maximum-likelihood approach", *J. Opt. Soc. Am. (A)*, vol. 9, num. 7, p. 1052–1061, 1992.
- [IDI 05] IDIER J., MUGNIER L., BLANC A., "Statistical behavior of joint least square estimation in the phase diversity context", *IEEE Trans. Image Processing*, vol. 14, num. 12, p. 2107–2116, Dec. 2005.
- [IOT] <http://tdc-www.harvard.edu/IOTA/>.
- [JEF 93] JEFFERIES S. M., CHRISTOU J. C., "Restoration of astronomical images by iterative blind deconvolution", *Astrophys. J.*, vol. 415, p. 862–874, 1993.
- [JEN 58] JENNISON R. C., "A phase sensitive interferometer technique for the measurement of the Fourier transforms of spatial brightness distribution of small angular extent", *Month. Not. Roy. Astr. Soc.*, vol. 118, p. 276–284, 1958.
- [KEC] <http://www.keckobservatory.org/>.
- [KNO 74] KNOX K. T., THOMPSON B. J., "Recovery of images from atmospherically degraded short exposure photographs", *Astrophys. J. Lett.*, vol. 193, p. L45–L48, 1974.
- [KUL 06] KULCSÁR C., RAYNAUD H.-F., PETIT C., CONAN J.-M., VIARIS DE LESEGNIO P., "Optimal control, observers and integrators in adaptive optics", *Opt. Express*, vol. 14, num. 17, p. 7464–7476, 2006.
- [LAB 70] LABEYRIE A., "Attainment of diffraction-limited resolution in large telescopes by Fourier analysing speckle patterns", *Astron. Astrophys.*, vol. 6, p. 85–87, 1970.

- [LAB 75] LABEYRIE A., “Interference fringes obtained on VEGA with two optical telescopes”, *Astrophys. J. Lett.*, vol. 196, p. L71-L75, Mar. 1975.
- [LAC 07] LACOUR S., *Imagerie des étoiles évoluées par interférométrie. Réarrangement de pupille*, PhD thesis, University of Paris VI, France, 2007.
- [LAN 92] LANE R. G., “Blind deconvolution of speckle images”, *J. Opt. Soc. Am. (A)*, vol. 9, num. 9, p. 1508-1514, 1992.
- [LAN 96] LANE R. G., “Methods for maximum-likelihood deconvolution”, *J. Opt. Soc. Am. (A)*, vol. 13, num. 10, p. 1992-1998, 1996.
- [LAN 98] LANNES A., “Weak-phase imaging in optical interferometry”, *J. Opt. Soc. Am. (A)*, vol. 15, num. 4, p. 811-824, Apr. 1998.
- [LAN 01] LANNES A., “Integer ambiguity resolution in phase closure imaging”, *Opt. Soc. Am. J. A*, vol. 18, p. 1046-1055, May 2001.
- [LAW 97] LAWSON P. R. (Ed.), *Long Baseline Stellar Interferometry*, Bellingham, SPIE Optical Engineering Press, 1997.
- [LAW 04] LAWSON P. R., COTTON W. D., HUMMEL C. A., MONNIER J. D., ZHAO M., YOUNG J. S., THORSTEINSSON H., MEIMON S. C., MUGNIER L., LE BESNERAIS G., THIÉBAUT E., TUTHILL P. G., “An interferometric imaging beauty contest”, in TRAUB W. A. (Ed.), *New Frontiers in Stellar Interferometry*, vol. 5491, Proc. Soc. Photo-Opt. Instrum. Eng., p. 886-899, 2004.
- [LER 04] LE ROUX B., CONAN J.-M., KULCSÁR C., RAYNAUD H.-F., MUGNIER L. M., FUSCO T., “Optimal control law for classical and multiconjugate adaptive optics”, *J. Opt. Soc. Am. (A)*, vol. 21, num. 7, July 2004.
- [MAR 89] MARIOTTI J.-M., “Introduction to Fourier optics and coherence”, in ALLOIN D. M., MARIOTTI J.-M. (Eds.), *Diffraction-limited Imaging with Very Large Telescopes*, vol. 274 of *NATO ASI Series C*, p. 3-31, Kluwer Academic, Cargese, France, 1989.
- [MEI 05a] MEIMON S., *Reconstruction d’images astronomiques en interférométrie optique*, PhD thesis, University of Paris XI, France, 2005.
- [MEI 05b] MEIMON S., MUGNIER L. M., LE BESNERAIS G., “Reconstruction method for weak-phase optical interferometry”, *Opt. Lett.*, vol. 30, num. 14, p. 1809-1811, July 2005.
- [MEI 05c] MEIMON S., MUGNIER L. M., LE BESNERAIS G., “A convex approximation of the likelihood in optical interferometry”, *J. Opt. Soc. Am. (A)*, Nov. 2005.
- [MER 88] MERKLE F. (Ed.), *High-resolution imaging by interferometry, part II*, num. 29 in ESO Conf. and Workshop Proc., Garching bei München, Germany, July 1988.
- [MIC 91] MICHELSON A. A., “Measurement of Jupiter’s satellites by interference”, *Nature (London)*, vol. 45, p. 160-161, Dec. 1891.
- [MIC 21] MICHELSON A. A., PEASE F. G., “Measurement of the diameter of alpha Orionis with the interferometer”, *Astrophys. J.*, vol. 53, p. 249-259, May 1921.
- [MON 03] MONNIER J. D., “Optical interferometry in astronomy”, *Reports of Progress in Physics*, vol. 66, p. 789-857, May 2003.

- [MOU 94] MOURARD D., TALLON-BOSC I., BLAZIT A., BONNEAU D., MERLIN G., MORAND F., VAKILI F., LABEYRIE A., “The GI2T interferometer on Plateau de Calern”, *Adv. Appl. Prob.*, vol. 283, p. 705-713, Mar. 1994.
- [MUG 01] MUGNIER L. M., ROBERT C., CONAN J.-M., MICHAU V., SALEM S., “Myopic deconvolution from wavefront sensing”, *J. Opt. Soc. Am. (A)*, vol. 18, p. 862-872, Apr. 2001.
- [MUG 04] MUGNIER L. M., FUSCO T., CONAN J.-M., “MISTRAL: a myopic edge-preserving image restoration method, with application to astronomical adaptive-optics-corrected long-exposure images”, *J. Opt. Soc. Am. (A)*, vol. 21, num. 10, p. 1841-1854, Oct. 2004.
- [MUG 06] MUGNIER L. M., BLANC A., IDIER J., “Phase diversity: a technique for wavefront sensing and for diffraction-limited imaging”, in HAWKES P. (Ed.), *Advances in Imaging and Electron Physics*, vol. 141, Chapter 1, p. 1-76, Elsevier, 2006.
- [MUG 07] MUGNIER L., MEIMON S., WISARD software documentation, Technical report, ONERA, 2007, European Interferometry Initiative, Joint Research Action 4, 6th Framework Programme of the EU.
- [NOL 76] NOLL R. J., “Zernike polynomials and atmospheric turbulence”, *J. Opt. Soc. Am.*, vol. 66, num. 3, p. 207-211, 1976.
- [NPO] <http://ftp.nofs.navy.mil/projects/npoi/>.
- [PER 04] PERRIN G., RIDGWAY S., MENNESSON B., COTTON W., WOILLEZ J., VERHOELST T., SCHULLER P., COUDÉ DU FORESTO V., TRAUB W., MILLAN-GALBET R., LACASSE M., “Unveiling Mira stars behind the molecules. Confirmation of the molecular layer model with narrow band near-infrared interferometry”, *Astron. Astrophys.*, vol. 426, p. 279-296, Oct. 2004.
- [PER 06] PERRIN G., WOILLEZ J., LAI O., GUÉRIN J., KOTANI T., WIZINOWICH P. L., LE MIGNANT D., HRYNEVYCH M., GATHRIGHT J., LÉNA P., CHAFFEE F., VERGNOLE S., DELAGE L., REYNAUD F., ADAMSON A. J., BERTHOD C., BRIENT B., COLLIN C., CRÉTENET J., DAUNY F., DELÉGLISE C., FÉDOU P., GOELTZENLICHTER T., GUYON O., HULIN R., MARLOT C., MARTEAUD M., MELSE B.-T., NISHIKAWA J., REESS J.-M., RIDGWAY S. T., RIGAUT F., ROTH K., TOKUNAGA A. T., ZIEGLER D., “Interferometric coupling of the Keck telescopes with single-mode fibers”, *Science*, vol. 311, p. 194-+, Jan. 2006.
- [PET 05] PETIT C., CONAN J.-M., KULCSAR C., RAYNAUD H.-F., FUSCO T., MONTRI J., RABAUD D., “Optimal control for multi-conjugate adaptive optics”, *C. R. Physique*, vol. 6, num. 10, p. 1059-1069, 2005.
- [PRI 88] PRIMOT J., ROUSSET G., FONTANELLA J.-C., “Image deconvolution from wavefront sensing: atmospheric turbulence simulation cell results”, in Merkle [MER 88], p. 683-692.
- [PRI 90] PRIMOT J., ROUSSET G., FONTANELLA J.-C., “Deconvolution from wavefront sensing: a new technique for compensating turbulence-degraded images”, *J. Opt. Soc. Am. (A)*, vol. 7, num. 9, p. 1598-1608, 1990.

- [REY 83] REY W. J., *Introduction to Robust and Quasi-robust Statistical Methods*, Springer Verlag, Berlin, 1983.
- [ROD 81] RODDIER F., "The effects of atmospherical turbulence in optical astronomy", in WOLF E. (Ed.), *Progress in Optics*, vol. XIX, p. 281-376, North Holland, Amsterdam, 1981.
- [ROD 82] RODDIER F., GILLI J. M., LUND G., "On the origin of speckle boiling and its effects in stellar speckle interferometry", *J. of Optics (Paris)*, vol. 13, num. 5, p. 263-271, 1982.
- [ROD 88a] RODDIER F., "Curvature sensing and compensation: a new concept in adaptive optics", *Appl. Opt.*, vol. 27, num. 7, p. 1223-1225, Apr. 1988.
- [ROD 88b] RODDIER F., "Passive versus active methods in optical interferometry", in Merkle [MER 88], p. 565-574, July 1988.
- [ROD 90] RODDIER N., "Atmospheric wavefront simulation using Zernike polynomials", *Opt. Eng.*, vol. 29, num. 10, p. 1174-1180, 1990.
- [ROD 98] RODDIER F., RODDIER C., GRAVES J. E., NORTHCOTT M. J., OWEN T., "Neptune cloud structure and activity: Ground based monitoring with adaptive optics", *Icarus*, vol. 136, p. 168-172, 1998.
- [ROD 99] RODDIER F. (Ed.), *Adaptive Optics in Astronomy*, Cambridge University Press, Cambridge, 1999.
- [ROU 90] ROUSSET G., FONTANELLA J.-C., KERN P., GIGAN P., RIGAUT F., LENA P., BOYER C., JAGOUREL P., GAFFARD J.-P., MERKLE F., "First diffraction-limited astronomical images with adaptive optics", *Astron. Astrophys.*, vol. 230, p. 29-32, 1990.
- [ROU 99] ROUSSET G., "Wave-front sensors", in Roddier [ROD 99], Chapter 5, p. 91-130.
- [ROU 01] ROUSSET G., MUGNIER L. M., CASSAING F., SORRENTE B., "Imaging with multi-aperture optical telescopes and an application", *C. R. Acad. Sci. Paris*, vol. IV, vol. 2, num. 1, p. 17-25, Jan. 2001.
- [SAS 85] SASIELA R. J., MOONEY J. G., "An optical phase reconstructor based on using a multiplier-accumulator approach", in *Proc. Soc. Photo-Opt. Instrum. Eng.*, vol. 551, Proc. Soc. Photo-Opt. Instrum. Eng., p. 170-176, 1985.
- [SCH 93] SCHULZ T. J., "Multiframe blind deconvolution of astronomical images", *J. Opt. Soc. Am. (A)*, vol. 10, num. 5, p. 1064-1073, 1993.
- [SHA 71] SHACK R. B., PLACK B. C., "Production and use of a lenticular Hartmann screen (abstract)", *J. Opt. Soc. Am.*, vol. 61, p. 656, 1971.
- [THI 95] THIÉBAUT E., CONAN J.-M., "Strict *a priori* constraints for maximum-likelihood blind deconvolution", *J. Opt. Soc. Am. (A)*, vol. 12, num. 3, p. 485-492, 1995.
- [THI 03] THIÉBAUT E., GARCIA P. J. V., FOY R., "Imaging with Amber/VLTI: the case of microjets", *Astrophys. Space. Sci.*, vol. 286, p. 171-176, 2003.
- [VER 97a] VERAN J.-P., Estimation de la réponse impulsionnelle et restauration d'image en optique adaptative. Application au système d'optique adaptative du Télescope Canada-France-Hawaii, PhD thesis, ENST, Paris, France, Nov. 1997.

- [VER 97b] VERAN J.-P., RIGAUT F., MAITRE H., ROUAN D., "Estimation of the adaptive optics long exposure point spread function using control loop data", *J. Opt. Soc. Am. (A)*, vol. 14, num. 11, p. 3057-3069, 1997.
- [WAL 83] WALLNER E. P., "Optimal wave-front correction using slope measurements", *J. Opt. Soc. Am. (A)*, vol. 73, num. 12, p. pp 1771-1776, Dec. 1983.
- [WEI 77] WEIGELT G., "Modified astronomical speckle interferometry 'speckle masking'", *Opt. Commun.*, vol. 21, num. 1, p. 55-59, 1977.

Chapter 11

Spectral Characterization in Ultrasonic Doppler Velocimetry

11.1. Velocity measurement in medical imaging

This chapter is devoted to *velocimetry*, i.e., the imaging of the velocity of moving structures. Velocimetry is employed in atmosphere imaging, industrial control, medical imaging, etc. In the medical field, it is essentially used to characterize blood flow and heart movements for the purpose of diagnosing cardiovascular pathologies.

These mainly concern stenosis, i.e., narrowing of the arteries connected with the presence of atheromatous plaque, and its repercussions on the organs supplied by the affected vessels (heart, brain, etc.). The resulting pathologies may be chronic (silent ischaemia) or acute (infarction). The information obtained is morphological: reduction of the cross-sectional area of the artery at the stenosis, estimation of the dimensions of the heart (cardiac hypertrophy) and monitoring of its contractions (hypokinesia, dyskinesia).

Ultrasound (US) and magnetic resonance imaging (MRI) both allow local velocity images to be acquired and thus give access to specific information that is complementary to that provided by morphological and functional imaging. Although the physical principles underlying these two techniques are radically different, the data measured and the problems posed show strong similarities. This chapter focuses on Doppler ultrasound, which is still the most delicate to interpret. We will see that the problems arising concern *spectral characterization*:

- spectral analysis and its adaptive extension (time-frequency), covered in section 11.2;
- tracking of the mean frequencies or spectral moments, dealt with in section 11.3.

As for the data measured, in both cases we have:

- short signals: only 16 to 48 samples to estimate a power spectral density (PSD) and only 4 to 8 samples to estimate the mean frequency;
- possible violation of Shannon's condition, hence possible spectral aliasing which needs to be detected and inverted;
- an unfavorable signal-to-noise ratio (SNR).

The data is globally very poor from an information standpoint. However, we also have information on the spectral, temporal and spatial coherence of the flow and structures. The haemodynamic mechanisms (viscosity of the blood) tend to organize the physiological flows. The hypothesis of the space and time variation of velocity in the flow being gradual is credible. If the flow is very turbulent, local dispersion of the velocities tends to invalidate the hypothesis but, nevertheless, the mean velocities of the erythrocytes (red corpuscles) remain similar in a plane perpendicular to the vessel axis expressing the progression of the flow in the vessel.

11.1.1. Principle of velocity measurement in ultrasound imaging

An acoustic wave having a frequency that can vary between 2 and 12 MHz, depending on the desired penetration depth, is emitted in a focused US beam by a probe. The wave is partly backscattered by the red blood cells and by mobile fibers of the tissues. Signal y_m obtained after time windowing, which enables a given depth to be isolated ("range bin" m), is thus affected by frequency shifts induced by the velocities of the backscattering structures. The Doppler shift for a single target having a constant velocity is:

$$f_d = \frac{2v \cos \theta}{c} f_e \quad (11.1)$$

where θ represents the angle between the US beam and the velocity, c is the speed of sound ($c = 1,470$ m/s) and f_e is the US frequency emitted.

11.1.2. Information carried by Doppler signals

The most intuitive way of characterizing the local complexity of a stationary flow is to represent it by a histogram of the erythrocyte velocities. A laminar flow observed locally will give a narrow histogram (the red blood cells move in the same direction at similar speeds) whereas turbulent flow downstream of a stenosis will have a much

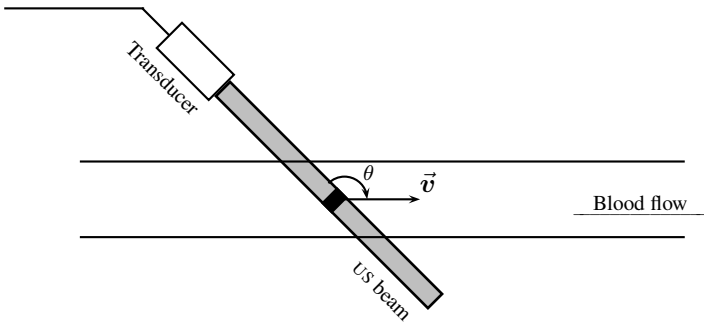


Figure 11.1. Principle of acquisition of a US signal. The transducer emits a US wave then receives the echoes reflected by the vessel walls and blood cells. Signal y_m coming from the “measurement volume” (or “range bin”) indicated in black is selected by means of an electronic gate synchronized on the emitted signal

wider histogram (the speeds and directions are very different). In medical practice, it is accepted that the power of the backscattered signal is proportional to the erythrocyte concentration. Equation (11.1) establishes the proportionality between the erythrocyte velocity and the frequency of the Doppler signal, so the histogram can be identified with the PSD $S_m(\nu)$ of the Doppler signal y_m .

The Doppler signal is non-stationary: the flow velocity and the movement of the walls vary rapidly during the cardiac ejection phase. This obliges us to use non-stationary signal analysis methods. The standard systems use sliding spectrogram or periodogram techniques. They thus process fairly long signal horizons (128 or 256 samples acquired over a duration of about 10 ms) resulting from a trade-off between the non-stationary nature of the signal and performance of the stationary periodogram methods. However, the literature [TAL 88] indicates that much shorter analysis horizons (less than 2 ms, i.e., about 16 samples) are indispensable if flow perturbations are to be characterized. The problem thus arises of tracking the spectral content of signals observed on short horizons. This is tackled in section 11.2.

Furthermore, in imaging systems, the spectral information is also reduced to a few parameters: the standard deviation of the spectrum gives an image of how dispersed the velocities are, the maximum frequency sometimes lifts ambiguities in the flow characterization, and the mean (or central) frequency, ν_m , provides a two-dimensional plot of local mean velocities. If images are to be constructed at a rate reasonable for the characterization of flow non-stationarities, the number of data points acquired has to be reduced yet again, down to eight and sometimes just four samples. We are then faced with the problem of tracking the mean frequencies of noisy signals observed on very short horizons.

Finally, in most systems, the emission is in pulsed mode in order to obtain spatial discrimination: wave trains are emitted with a recurrence frequency f_r that is limited by the desired image depth d , since an emission is held back until the deepest echoes of the previous emission have been received. The Doppler signal is thus “sampled” at the frequency $f_r = c/2d$. For high flow velocities ($f_d > f_r/2$), a spectral aliasing phenomenon occurs, which causes ambiguity in the velocity measurements. At present, the alternatives are either to lower the emission frequency at the cost of losing spatial resolution, or to reduce the angle between the ultrasound beam and the flow, at the cost of losing sensitivity. We will see how to invert the aliasing without relinquishing resolution or accuracy in section 11.3.

11.1.3. *Some characteristics and limitations*

The orders of magnitude (v varies by a few cm/s or m/s and the walls move at less than 20 cm/s) show that the Doppler shift f_d is of the order of 10^{-3} to 10^{-4} times the emitted frequency. Baseband demodulation thus allows only the useful information to be kept. The Doppler signal is thus a signal with complex values in a band of about 0 ± 20 kHz.

This signal is the result of backscattering of the acoustic wave by erythrocytes. The absolute value of the backscattering is low because the erythrocytes are small relative to the acoustic wavelength and their acoustic impedance is close to that of blood. In addition, the US wave is strongly attenuated (0.5 to 1 dB/cm/MHz) by the tissues located between the probe and the flow. Very often, the SNR is in the [0 dB, 15 dB] range.

The analysis of the spectral parameters should be approached with caution. The information collected is simply the projection of the actual velocity on the axis of the US beam, hence the presence of $\cos \theta$ in (11.1). Any dispersion in the direction of propagation of the erythrocytes and any modification of the incidence of the beam on the flow will introduce a change in the velocity measurement.

Finally, a number of artefacts are more difficult to analyze: large variation of the amplitude of the Doppler signal according to the nature of the flow, parasite echoes from multiple reflections on other structures, renewal of the erythrocytes present in the volume being measured (all the more significant at high flow speeds), etc.

11.1.4. *Data and problems treated*

The data considered in this chapter comes in the form of M complex signals $\mathcal{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_M]$ spatially juxtaposed in M “range bins”. Each \mathbf{y}_m is a vector of N samples extracted from a signal assumed to be stationary: $\mathbf{y}_m = [y_m(1), \dots, y_m(N)]^T$.

The data was acquired through apparatus capable of reproducing various characteristics of medical practice but in comfortable study conditions. The measuring system had a tube supplied with a permanent flow and equipped with a Doppler scanner (AU3 ESAOTE¹ – Italy)². The scanner operated at an emission frequency $f_e = 8$ MHz, a repetition frequency $f_r = 10$ kHz and an angle of incidence $\theta = 60^\circ$. The apparent diameter of the vessel was 22.4 mm, and it was divided into $M = 64$ range bins of 0.35 mm (see Figures 11.1 and 11.2). The system provided signals sampled at the frequency f_r in real time and for each bin simultaneously. They were encoded on 14 bits. The file processed contained 1.64 s of recording, i.e., $N_0 = 2^{14}$ samples in each distance cell. An extract of this data is shown in Figures 11.2a_r and a_i.

This experimental set-up first allowed us to use all the data in order to obtain the characteristics of the flow. As the flow was permanent, the signal recorded for each range bin was a stationary signal of N_0 samples. Dividing it into 64 signals of 256 points each enabled a mean periodogram to be calculated with a reasonable compromise between bias and variance. It could thus be assumed, at least as a first approximation, that this was the “true” sequence of the PSDs. It is represented in Figure 11.2b and will serve as an element of comparison for the rest of the chapter. Figure 11.2c shows the series of frequencies that maximize each of the periodograms. It will also serve as an element of comparison for the rest of the chapter.

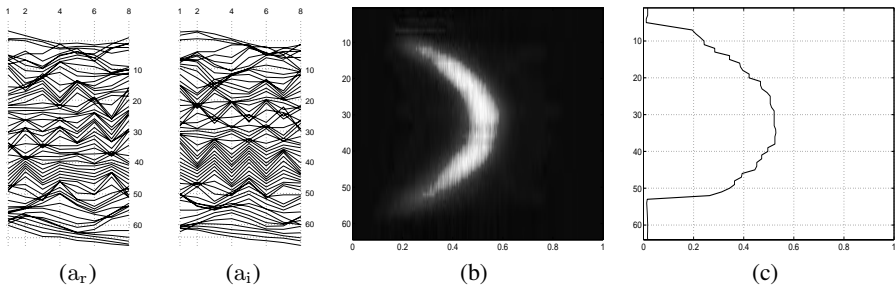


Figure 11.2. Figures a_r and a_i give the real and imaginary parts respectively of eight samples in each bin. Figures b and c show the averaged periodograms and the mean frequency sequence respectively (parabolic flow). In the four figures, the vertical axis represents depth ($m = 1$ to $m = M = 64$). The horizontal axis is time (1 to 8) for a_r and a_i, and frequency (0 to 1) for b and c

In clinical practice, the flows are obviously not permanent and are even strongly non-stationary sometimes. As explained in section 11.1.2, we chose to process signals of $N = 16$ samples in order to assess the capacities of the methods to solve spectral

1. BIOMED 2, contract no. BMH4-CT98-3782 (DG 12 - SSMI).

2. We thank P. Tortoli of the University of Florence for the acquisition of the Doppler signals.

characterization problems in conditions that were both realistic and difficult. We chose the two principle problems of this type:

(α) adaptive spectral analysis, which consists of estimating the sequence of PSD $S_m(\nu)$ of the \mathbf{y}_m . We will tackle this using long AR models in section 11.2;

(β) frequency tracking and inversion of the spectral aliasing, which means estimating the series of frequencies ν_m of signals \mathbf{y}_m . This is covered in section 11.3 on the basis of pure frequency models.

In both cases, a Gaussian Markov chain is introduced for the AR parameters (α) and the frequencies (β) in order to take spatial continuity into account. The criteria thus constructed are optimized by suitable algorithms: a Kalman smoother (α) and a Viterbi algorithm (β). The question of the hyperparameters is settled by maximum likelihood in both cases. The likelihood is optimized by a coordinate-wise descent algorithm (α) and a gradient algorithm (β). Each of the two sections ends with characteristic results³ for the signals described above.

11.2. Adaptive spectral analysis

The literature on adaptive spectral analysis contains several possible approaches: periodograms and spectrograms (and their variants), ARMA methods, AR by least squares (LS) and their adaptive extensions with sliding windows or forgetting coefficients, Wigner-Ville methods, etc. See [GUO 94, HER 97] for a broad comparative study in the US velocimetry context, the conclusion of which recommends parametric AR methods by LS. For this reason, we will turn directly to this class of methods.

In the AR spectral analysis framework, finding the sequence of PSD of signals \mathbf{y}_m requires estimation of the AR parameters $\mathbf{a}_m = [a_{mp}]$, where m is the index of the range bin under consideration $m \in \{1, \dots, M\} = \mathbb{N}_M^*$ and p is the order of coefficient ($p \in \mathbb{N}_P^*$). Let $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_M]$ be the regressor sequence and let r_m^e and r_m be the powers of the input noises and the corresponding signals.

11.2.1. Least squares and traditional extensions

In one of the M range bins, estimation of the regressor \mathbf{a}_m from the corresponding data \mathbf{y}_m by LS relies on a prediction error $\mathbf{e}_m = \mathbf{y}_m - \mathbf{Y}_m \mathbf{a}_m$ and the criterion:

$$Q_m^{\text{LS}}(\mathbf{a}_m) = \mathbf{e}_m^\dagger \mathbf{e}_m = (\mathbf{y}_m - \mathbf{Y}_m \mathbf{a}_m)^\dagger (\mathbf{y}_m - \mathbf{Y}_m \mathbf{a}_m). \quad (11.2)$$

3. The algorithms described here have been implemented in *Matlab* on a Pentium III PC, running at 450 MHz and equipped with 128 MB of RAM.

Vector \mathbf{y}_m (of size $L \times 1$) and matrix \mathbf{Y}_m (of size $L \times P - 1$) are defined according to the type of windowing [KAY 81, equation (2)], [MAR 87, p. 217]. There are four types: non-windowed (called covariance), pre-windowed, post-windowed and double windowed or pre- and post-windowed (also called autocorrelation). Depending on the case, $L = N - P$, $L = N$, $L = N + P$. This choice is important as it strongly influences the conditioning of the normal matrix and the spectral resolution, particularly when the number of data points is small [MAR 87, p. 225]. Whatever the type of windowing, minimizing the LS criterion (11.2) leads to [SOR 80]:

$$\hat{\mathbf{a}}_m^{\text{LS}} = \arg \min_{\mathbf{a}_m} Q_m^{\text{LS}}(\mathbf{a}_m) = (\mathbf{Y}_m^\dagger \mathbf{Y}_m)^{-1} \mathbf{Y}_m^\dagger \mathbf{y}_m. \quad (11.3)$$

For the set of M windows, adaptive LS techniques (ALS) work by taking account of the criteria (11.2) around the current bin in a window of variable width or by geometrical weighting⁴.

There are a number of drawbacks to these algorithms, connected with the choice of the parameters, at least in our context:

- The LS methods can only be used in combination with a principle of parsimony that limits the order of the model [AZE 86] and thus avoids parasite peaks in the spectrum⁵. This compromise can be found automatically through criteria such as FPE [AKA 70], AIC [AKA 74], CAT [PAR 74] or MDL [RIS 78], but they become inefficient when there is not enough data [ULR 76].

- Even if the order of the model is adjusted empirically, from a small amount of data, the acceptable orders remain too low to describe the great variety of spectra that can be encountered in velocimetry. We will see that the method presented allows high-order models to be estimated.

- From a spatial point of view, in the ALS method framework, no automatic method for adjusting the trade-off exists, either for estimating the width of the window or for the forgetting coefficient.

To find a remedy for these disadvantages, we will look into regularized techniques in the next subsection.

11.2.2. Long AR models – spectral smoothness – spatial continuity

11.2.2.1. Spatial regularity

The idea is to restate the problem including the notions of spatial regularity and spectral smoothness within the criterion itself. To do this, we generalize the seminal

4. This process directly brings an idea of spatial regularity into the sequence of spectra.

5. This limitation of the order is a roundabout way of inducing spectral smoothness.

work of Kitagawa and Gersch [KIT 85] to construct a measure of the distance between two AR spectra. Starting from the expression for the PSD:

$$S_m(\nu) = \frac{r_m^e}{|1 - A_m(\nu)|^2} \text{ with } A_m(\nu) = \sum_{p=1}^P a_{mp} e^{2j\pi\nu p},$$

we measure the spectral distance between S_m and $S_{m'}$ by using the Sobolev distance of order k between functions A_m and $A_{m'}$:

$$D_k(m, m') = \int_0^1 \left| \frac{d^k}{d\nu^k} (A_m(\nu) - A_{m'}(\nu)) \right|^2 d\nu.$$

It can be easily shown that a quadratic form can be obtained:

$$D_k(m, m') = (\mathbf{a}_m - \mathbf{a}_{m'})^\dagger \mathbf{\Delta}_k (\mathbf{a}_m - \mathbf{a}_{m'}), \quad (11.4)$$

where $\mathbf{\Delta}_k$ is a simply diagonal matrix $\mathbf{\Delta}_k = \text{diag}[1^{2k}, 2^{2k}, \dots, P^{2k}]$ called the smoothness matrix of order k .

11.2.2.2. Spectral smoothness

To measure the spectral smoothness, we just need to measure the distance to a constant spectrum, i.e., $A_{m'} = 0$, which again gives a quadratic form, initially proposed by Kitagawa and Gersch:

$$D_k(m) \propto \mathbf{a}_m^\dagger \mathbf{\Delta}_k \mathbf{a}_m. \quad (11.5)$$

Note 1 *Strictly speaking, this is not a measure of spectral distance or spectral smoothness, since the D_{ks} are not built from the PSD S_m but from functions A_m . Nevertheless, it does measure the spatial regularity and spectral smoothness in a certain sense. Furthermore, its quadratic nature greatly simplifies the optimization with respect to the \mathbf{a}_{ms} (section 11.2.2.4) and the question of estimating the hyperparameters (section 11.2.4).*

11.2.2.3. Regularized least squares

Starting from the two expressions (11.4)-(11.5) and the LS criteria (11.2), we construct the regularized LS criterion (LSReg), as in (2.5) of Chapter 2:

$$\begin{aligned} Q^{\text{Reg}}(\mathbf{A}) &= \sum_{m=1}^M \frac{1}{r_m^e} (\mathbf{y}_m - \mathbf{Y}_m \mathbf{a}_m)^\dagger (\mathbf{y}_m - \mathbf{Y}_m \mathbf{a}_m) \\ &+ \frac{1}{r_s} \sum_{m=1}^M \mathbf{a}_m^\dagger \mathbf{\Delta}_k \mathbf{a}_m + \frac{1}{r_d} \sum_{m=1}^{M-1} (\mathbf{a}_m - \mathbf{a}_{m+1})^\dagger \mathbf{\Delta}_k (\mathbf{a}_m - \mathbf{a}_{m+1}) \end{aligned} \quad (11.6)$$

which has three terms: the first measures fidelity to the data, the second the spectral smoothness of each spectrum and the third the spatial regularity. The relative weight of each term is determined by the powers of the input noises r_m^e and, above all, the spectral parameter (r_s) and spatial parameter (r_d). We will also use $\lambda_s = 1/r_s$ and $\lambda_d = 1/r_d$, so that the constraints increase with the parameters.

11.2.2.4. Optimization

Given its quadratic structure, several competing options can be considered for the minimization of (11.6); see Chapter 2, section 2.2.2, which is devoted to this question. It is possible to obtain the minimum explicitly by solving a sparse linear system of dimension $MP \times MP$. The criterion in question is convex and differentiable, so gradient techniques are also a possibility [BER 95]. However, with a view to carrying out the processing on-line, we opted for Kalman filtering (KF) and Kalman smoothing (KS), which was also Kitagawa and Gersch's initial point of view in [KIT 85]. See Chapter 4, a part of which covers KF and KS.

11.2.3. Kalman smoothing

11.2.3.1. State and observation equations

To use this alternative, it is necessary to express the model in a state representation form. We will not return to the more general Kalman formulation and it will be seen that the following form is sufficient for optimization purposes.

The evolution of the successive regressors \mathbf{a}_m is guided by the state model:

$$\mathbf{a}_{m+1} = \alpha_m \mathbf{a}_m + \boldsymbol{\varepsilon}_m, \quad (11.7)$$

where each $\boldsymbol{\varepsilon}_m$ is a zero-mean, circular, complex vector with covariance matrix $P_m^\varepsilon = r_m^\varepsilon \boldsymbol{\Delta}_k^{-1}$, and sequence $\boldsymbol{\varepsilon}_m$, for $m \in \mathbb{N}_M^*$, is spatially white. This is a generalized version of the one proposed by Kitagawa and Gersch [KIT 85].

The state model also brings in the mean and the initial covariance of the state: the zero vector and $P^a = r^a \boldsymbol{\Delta}_k^{-1}$ respectively.

The observation equation is simply the recurrence equation that guides the AR model in each range bin, in matrix form:

$$\mathbf{y}_m = \mathbf{Y}_m \mathbf{a}_m + \mathbf{e}_m. \quad (11.8)$$

Each \mathbf{e}_m is a zero-mean, circular, complex vector of covariance $r_m^e \mathbf{I}_L$. The sequence \mathbf{e}_m , $m \in \mathbb{N}_M^*$, is also spatially white. This, too, is a generalization of the form proposed in [KIT 85].

11.2.3.2. Equivalence between parameterizations

To implement the KS, the equations for which are given in Chapter 4, section 4.5.1, it is necessary to determine its parameters (r^a and $\alpha_m, r_m^\varepsilon$ for $m \in \mathbb{N}_{M-1}^*$). They are determined according to r_d, r_s so that the associated KS effectively minimizes criterion (11.6). [JAZ 70, p. 150] gives the criterion minimized by the KS associated with (11.7)-(11.8):

$$Q^{\text{KS}}(\mathbf{A}) = \sum_{m=1}^M \frac{1}{r_m^\varepsilon} (\mathbf{y}_m - \mathbf{Y}_m \mathbf{a}_m)^\dagger (\mathbf{y}_m - \mathbf{Y}_m \mathbf{a}_m) + \sum_{m=1}^{M-1} \frac{1}{r_m^\varepsilon} (\mathbf{a}_{m+1} - \alpha_m \mathbf{a}_m)^\dagger \Delta_k (\mathbf{a}_{m+1} - \alpha_m \mathbf{a}_m) + \frac{1}{r^a} \mathbf{a}_1^\dagger \Delta_k \mathbf{a}_1. \quad (11.9)$$

By developing and identifying (11.6) and (11.9), we establish the link between the two sets of parameters in the form of a descending recurrence:

① initialization ($m = M - 1$):

$$\alpha_{M-1} = (1 + \rho)^{-1} \quad \text{and} \quad r_{M-1}^\varepsilon = r_d \alpha_{M-1};$$

② recursion ($m = M - 2, \dots, 1$):

$$\alpha_m = (2 + \rho - \alpha_{m+1})^{-1} \quad \text{and} \quad r_m^\varepsilon = r_d \alpha_m;$$

③ the last step gives the initial power:

$$r^a = r_d (1 + \rho - \alpha_1)^{-1} \quad \text{with} \quad \rho = r_d / r_s > 0.$$

These equations allow the coefficients of the KS (r^a and the $\alpha_m, r_m^\varepsilon$) to be calculated in advance according to r_d, r_s so as to minimize (11.6).

Note 2 It can be shown [GIO 01] that the above system admits a stationary limit and that the corresponding “stationary” criterion differs from the homogeneous criterion (11.6) only by the two terms connected with the first and last regressors, which are both proportional to $\alpha(\alpha - 1)/r^\varepsilon$. The simpler, stationary form is thus most often used in practice.

11.2.4. Estimation of hyperparameters

The method has $M + 4$ hyperparameters: k , the order of smoothness; P , the order of the AR model; r_m^ε for $m \in \mathbb{N}_M^*$, the sequence of the prediction errors, λ_s for spectral continuity and λ_d for spatial continuity.

The problem of estimating the hyperparameters is one of the questions that have not been completely resolved yet. See Chapter 8, which is devoted to this question. The approach chosen here is to maximize the likelihood (ML) of the hyperparameters. The quadratic character of criterion (11.6) allows it to have a Bayesian interpretation in Gaussian terms and enables the observation law to be explicitly deduced given the hyperparameters $p(\mathbf{y}_1, \dots, \mathbf{y}_M; k, P, r_1^e, r_2^e, \dots, r_M^e, \lambda_s, \lambda_d)$ (see Chapter 3). In this context, a completely satisfactory approach would be to maximize the likelihood with respect to the $M + 4$ hyperparameters. For the sake of computational efficiency, the likelihood is only maximized with respect to λ_s and λ_d .

The order parameters are fixed independently of the data. The order P of the AR model no longer influences the form of the spectra if it is chosen large enough, e.g. $P > N/2$. This is why it is kept at its maximum value $P = N - 1$ in practice (hence the expression “long AR”). The smoothness order also has little influence on the form of the spectra obtained provided that it is chosen to be non-zero. In practice, it is set at $k = 1$.

Parameters r_m^e for the power of the input noises play the role of weighting the data of each range bin in the criterion for LSReg (11.6). In practice, they are replaced by the powers of signals r_m in order to simplify the estimation procedure. These parameters can be estimated independently from the data by the standard empirical estimator of the power $\hat{r}_m = \mathbf{y}_m^\dagger \mathbf{y}_m / N$. Practically, they have only a weak influence on the shape of the PSD.

We now come to the two parameters that principally influence the shape of the depth-frequency map: λ_s and λ_d , which are both set automatically by ML. We know how to calculate the likelihood from the sub-products of the KF. Up to some constants, the co-log-likelihood (CLL) reads:

$$CLL(\lambda_s, \lambda_d) = \sum_{m=1}^M \log |\mathbf{R}_m| + \mathbf{e}_m^\dagger \mathbf{R}_m^{-1} \mathbf{e}_m,$$

which is to be minimized with respect to (λ_s, λ_d) . This calculation requires two sub-products of the KF (\mathbf{R}_m and \mathbf{e}_m) and, in particular, the inversion and calculation of the determinant of \mathbf{R}_m . This is a square matrix of size L that varies from 1 to $2N - 1$ depending on the type of windowing chosen. In practice, this matrix remains small and it does not appear crucial to use specific inversion algorithms.

As far as the minimization of the CLL is concerned, several approaches are in competition but none of them can guarantee that the global minimum will be obtained. This aspect is specifically covered in Chapter 8. The model used is a coordinate-wise descent method with golden section directional search [BER 95].

11.2.5. Processing results and comparisons

This subsection is devoted to processing results based on data presented in section 11.1.4. The method described above is compared to the periodogram method (used in commercial systems). In particular, we have excluded the LS methods as their order selecting methods are not reliable, as explained earlier.

11.2.5.1. Hyperparameter tuning

The two important parameters (λ_s, λ_d) were adjusted automatically by ML. First, the CLL was calculated on a logarithmic grid of 100×100 values of (λ_s, λ_d) . The corresponding contours, given in Figure 11.3, are pretty regular and have a single, clearly marked minimum at $\hat{\lambda}_s = -1.26$ and $\hat{\lambda}_d = 2.14$ (on a \log_{10} scale). These values can be obtained in practice in only 2.35 s using the descent algorithm described above. Some typical trajectories are also shown in Figure 11.3.

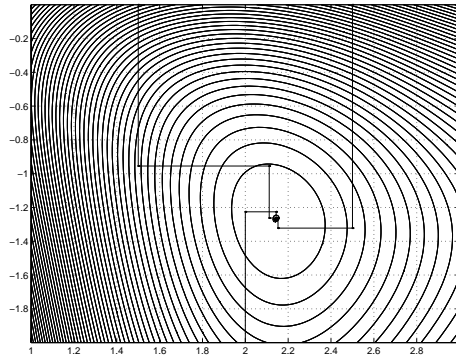


Figure 11.3. $CLL(\lambda_s, \lambda_d)$ contours to be minimized in (λ_s, λ_d) . The minimum is indicated by a star (*). The spectral parameter λ_s is shown on the vertical axis and the spatial parameter λ_d on the horizontal axis (\log_{10} scale in both cases). The figure also shows the trajectories of the optimization algorithm for three different initializations

Note 3 It is noteworthy that, for variations of less than a decade in the hyperparameters, the change in the PSD map is very slight. This aspect is particularly important for qualifying the overall robustness of the method. Unlike the AR model order in standard LS methods, which can have an abrupt effect on the shape of the spectra, the choice of (λ_s, λ_d) offers more flexibility for automatic or handmade adjustments.

11.2.5.2. Qualitative comparison

As far as the spectra themselves are concerned, the typical results obtained by periodogram and the method presented here are shown in Figure 11.4. A simple qualitative

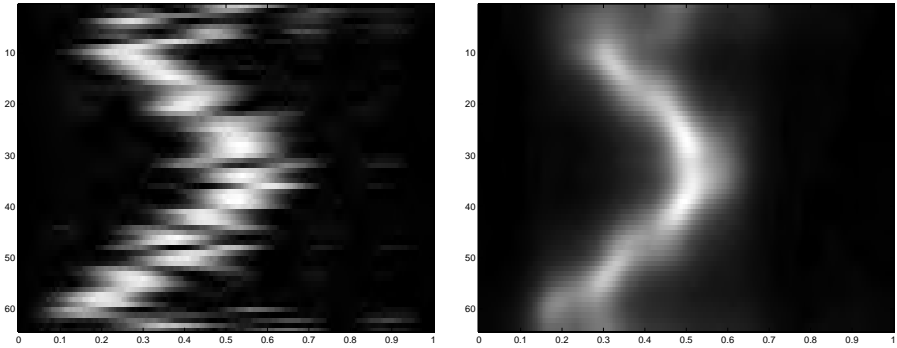


Figure 11.4. *Depth-frequency contours obtained, left, by standard periodogram; right, by the method presented (with hyperparameters by ML)*

comparison with the reference of Figure 11.2 allows several conclusions to be drawn. First, the gain relative to the periodogram is obvious: the spatial regularity (in the vertical depth direction) clearly improves the series of spectra that are in conformity with the reference of Figure 11.2. Secondly, the frequency dynamics is respected: from 0.2 “at the edge” of the vessel to 0.6 at the center. Third, the spectral resolution is also increased and fits the reference better.

11.3. Tracking spectral moments

We now come to the second problem, that of frequency tracking, possibly beyond Shannon’s frequency limit. The signals to be analyzed are the same as before but the number of data points available per range bin is generally lower, e.g. $N = 8$, for the reasons given in section 11.1.2. These same signals are now modeled by a pure sine wave drowned in noise. Using notations a_m and ν_m for the complex amplitude and the frequency, and \mathbf{b}_m for the measuring or modeling noise:

$$\mathbf{y}_m = a_m \mathbf{z}(\nu_m) + \mathbf{b}_m \quad \text{with} \quad \mathbf{z}(\nu_m) = [1, e^{j2\pi\nu_m}, \dots, e^{j2\pi\nu_m(N-1)}]^T \quad (11.10)$$

This model is well known in spectral analysis but gives rise to two observations: first, it is clearly periodic with respect to ν_m . In one sense, this characteristic is the cornerstone of the problem discussed here: it expresses the spectral aliasing and constitutes the key to its inversion. Secondly, although it is linear with respect to the amplitude a_m , it is not linear with respect to the frequency ν_m : the problem to be dealt with is therefore nonlinear.

Let us also note the vectors of the frequencies and amplitudes as $\boldsymbol{\nu} = [\nu_1, \dots, \nu_M]^T$ and $\mathbf{a} = [a_1, \dots, a_M]^T$. Finally, the “true parameters” are noted

with a star: $\boldsymbol{\nu}^*$, \mathbf{a}^* , etc. What we need to do is to construct an estimator $\hat{\boldsymbol{\nu}}$ for the parameters of interest $\boldsymbol{\nu}^*$. The \mathbf{a}^* parameters are less interesting and are called *nuisance parameters*). The method described is Bayesian and is based on several elements:

- marginalization of the nuisance parameters;
- modeling of the sequence of frequencies $\boldsymbol{\nu}$ by a Markov chain so as to take the spatial continuity into account;
- choice of a chain of discrete states allowing specific algorithms to be used (for the calculation of both the solution $\hat{\boldsymbol{\nu}}$ and the hyperparameters).

11.3.1. Proposed method

11.3.1.1. Likelihood

With the hypothesis that the \mathbf{b}_m are Gaussian, zero-mean, white, uniform, of variance r_b and spatially independent, it is easy to construct the likelihood of the set of frequency and amplitude parameters as the product:

$$p(\mathcal{Y} | \boldsymbol{\nu}, \mathbf{a}) = \prod_{m=1}^M p(\mathbf{y}_m | \nu_m, a_m) = (\pi r_b)^{-NM} \exp \{ -CLL(\boldsymbol{\nu}, \mathbf{a}) / r_b \} \quad (11.11)$$

where the CLL has the form of an LS criterion:

$$CLL(\boldsymbol{\nu}, \mathbf{a}) = \sum_{m=1}^M (\mathbf{y}_m - a_m \mathbf{z}(\nu_m))^\dagger (\mathbf{y}_m - a_m \mathbf{z}(\nu_m))$$

11.3.1.2. Amplitudes: prior distribution and marginalization

Defining the law for parameters $\boldsymbol{\nu}$ and \mathbf{a} necessitates the construction of a joint law in $(\boldsymbol{\nu}, \mathbf{a})$. The absence of information concerning the links between amplitudes and frequencies naturally leads to a separable choice:

$$p(\boldsymbol{\nu}, \mathbf{a}) = p(\boldsymbol{\nu}) p(\mathbf{a}) . \quad (11.12)$$

For the amplitudes, a separable choice is also dictated by the absence of information on a possible interdependence between range bins. As far as the form of the law is concerned, since the marginalization of the amplitudes is to be carried out, we resort to a Gaussian law:

$$p(\mathbf{a}) = (\pi r_a)^{-M} \exp \{ -\mathbf{a}^\dagger \mathbf{a} / r_a \} . \quad (11.13)$$

Given the separability of likelihood (11.11) and prior law (11.13), the marginalization of the amplitudes yields:

$$p(\mathcal{Y} | \boldsymbol{\nu}) = \prod_{m=1}^M \int_{a_m} p(\mathbf{y}_m | \nu_m, a_m) p(a_m) da_m = \prod_{m=1}^M p(\mathbf{y}_m | \nu_m) . \quad (11.14)$$

In addition, since (11.10) is linear with respect to a_m and b_m and since a_m and b_m are Gaussian and independent; the law for $(\mathbf{y}_m | \nu_m)$ is also Gaussian, zero-mean and of covariance:

$$\mathbf{R}_m = E(\mathbf{y}_m \mathbf{y}_m^\dagger) = r_a \mathbf{z}(\nu_m) \mathbf{z}(\nu_m)^\dagger + r_b \mathbf{I}_N.$$

The expression for the law of $(\mathbf{y}_m | \nu_m)$ obviously brings in its determinant and its inverse, which can be written explicitly in the form:

$$\mathbf{R}_m^{-1} = r_b^{-1} \mathbf{I}_N - N \alpha \mathbf{z}(\nu_m) \mathbf{z}(\nu_m)^\dagger \quad \text{and} \quad |\mathbf{R}_m| = r_b^{N-1} (r_b + N r_a),$$

with $\alpha = N r_a / (r_b (N r_a + r_b))$. The complete law can thus be written:

$$p(\mathbf{y}_m | \nu_m) = \pi^{-N} |\mathbf{R}_m^{-1}| \exp \{ -\mathbf{y}_m^\dagger \mathbf{R}_m^{-1} \mathbf{y}_m \} \quad (11.15)$$

$$= \beta \exp \{ -\gamma_m + N \alpha P_m(\nu_m) \} \quad (11.16)$$

where P_m is the periodogram of signal \mathbf{y}_m at frequency ν_m :

$$P_m(\nu_m) = (\mathbf{z}(\nu_m)^\dagger \mathbf{y}_m)^\dagger (\mathbf{z}(\nu_m)^\dagger \mathbf{y}_m) = \frac{1}{N} \left| \sum_{n=1}^N y_m(n) e^{2j\pi \nu_m n} \right|^2,$$

$$\text{and} \quad \beta = \pi^{-N} r_b^{1-N} / (N r_a + r_b), \quad \gamma_m = \mathbf{y}_m^\dagger \mathbf{y}_m / r_b.$$

Finally, the joint law for the set of observations, given the frequencies, can be written as product (11.14):

$$p(\mathcal{Y} | \boldsymbol{\nu}) = \beta^M \exp \{ -\gamma \} \exp \{ -\alpha \text{CLML}(\boldsymbol{\nu}) \} \quad \text{with} \quad \gamma = \sum_{m=1}^M \gamma_m \quad (11.17)$$

where the co-log-marginal-likelihood (CLML) $\text{CLML}(\boldsymbol{\nu})$ is the opposite of the sum of the periodograms in each range bin.

$$\text{CLML}(\boldsymbol{\nu}) = - \sum_{m=1}^M P_m(\nu_m) \quad (11.18)$$

Note 4 The fundamental property of this function, mentioned in the introduction, is its 1-periodic character with respect to each of the variables: $\forall k_m \in \mathbb{Z}, m = 1, \dots, M$:

$$\text{CLML}(\nu_1, \nu_2, \dots, \nu_m) = \text{CLML}(\nu_1 + k_1, \nu_2 + k_2, \dots, \nu_m + k_m) \quad (11.19)$$

In consequence, the information contributed by the data leaves the set of frequencies undetermined.

11.3.1.3. Frequencies: prior law and posterior law

As announced in the introduction, the frequency law is built with discrete states. We postulate a minimum value ν_m and a maximum value ν_M for the frequencies and regularly discretize the interval $[\nu_m, \nu_M]$ on an arbitrarily fine grid of P values. The possible values for the frequency are denoted ν^p , for $p \in \mathbb{N}_P$.

In contrast to the amplitude law, the law chosen for the frequencies takes the idea of spatial continuity into account through a Markov chain associated with a quadratic Gibbs energy:

$$CLP(\boldsymbol{\nu}) = \sum_{m=1}^{M-1} (\nu_{m+1} - \nu_m)^2, \quad (11.20)$$

where CLP is used for *Co-Log-Prior*. We obtain the probabilities of transitions between the states of the chain by “discretizing and renormalizing” a Gaussian law of variance r_ν :

$$\mathbb{P}_m(p, q) = \Pr(\nu_{m+1} = \nu^p \mid \nu_m = \nu^q) = \frac{\exp\{-(\nu^p - \nu^q)^2/2r_\nu\}}{\sum_{p'=1}^P \exp\{-(\nu^{p'} - \nu^q)^2/2r_\nu\}} \quad (11.21)$$

which is independent of m . A uniform initial probability is attributed to the frequencies, i.e.,

$$\mathbb{P}(p) = \Pr(\nu_1 = \nu^p) = 1/P.$$

Finally, in what follows, the “probability⁶ of observations” will be denoted $\mathbb{O}_m(p) = p(\mathbf{y}_m \mid \nu_m = \nu^p)$.

The prior information and the information provided by the data is merged using Bayes’ rule, which gives the posterior law for $\boldsymbol{\nu}$:

$$p(\boldsymbol{\nu} \mid \mathcal{Y}) \propto p(\mathcal{Y} \mid \boldsymbol{\nu}) p(\boldsymbol{\nu}) \propto \exp\{-\alpha CLPL(\boldsymbol{\nu})\}$$

where the *Co-Log-Posterior-Likelihood* (CLPL) can be written in the form:

$$CLPL(\boldsymbol{\nu}) = - \sum_{m=1}^M P_m(\nu_m) + \lambda \sum_{m=1}^{M-1} (\nu_{m+1} - \nu_m)^2 \quad (11.22)$$

with $\lambda = 1/2\alpha r_\nu$. In a deterministic framework, the CLPL is a regularized LS criterion. It contains two terms that measure the confidence we have in the measurements and in the *a priori* idea of spatial regularity respectively. The regularization parameter λ (which depends on the hyperparameters $\mathbf{r} = [r_a, r_b, r_\nu]$) adjusts the compromise between the two.

6. Strictly speaking, this is a density and not a probability but we will nevertheless use this notation as it is the usual one in the literature on Markov chains.

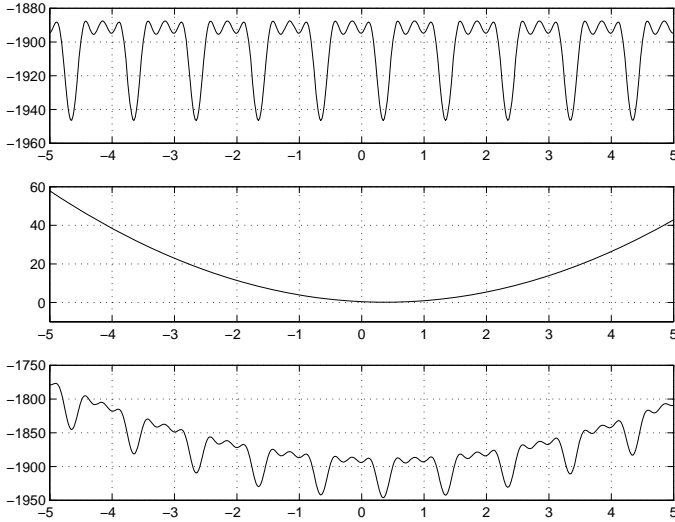


Figure 11.5. Typical form of the criteria as a function of one of the ν_m ($m = 50$). From top to bottom: $CLML(\nu)$ (1-periodic), $CLP(\nu)$ (quadratic) and their sum $CLPL(\nu)$. The regularization breaks the periodicity and removes the indeterminations

Note 5 This is the “regularized counterpart” of note 4. While the $CLML$ is 1-periodic in all directions ν_m , $m = 1, \dots, M$, i.e., it verifies (11.19), the $CLPL$ does not have this property. The regularization term removes the indeterminations.

However, a global indetermination remains. $CLML(\nu)$ is 1-periodic with respect to each ν_m and the regularization term only depends on the successive differences between frequencies, so it is globally insensitive to a constant level. We then have $\forall k_0 \in \mathbb{Z}$,

$$CLPL([\nu_1, \nu_2, \dots, \nu_m]) = CLPL([\nu_1 + k_0, \nu_2 + k_0, \dots, \nu_m + k_0]).$$

Two frequency profiles differing by a constant integer level remain equiprobable *a posteriori*. This indetermination is removed by specifying that the first frequency is within $[-1/2, +1/2)$.

All that remains is to choose a punctual estimator $\hat{\nu}$ for ν^* . Our first choice was the maximum *a posteriori* (MAP), i.e., the minimizer of regularized criterion (11.22):

$$\hat{\nu}^{\text{MAP}} = \arg \max_{\nu} p(\nu | \mathcal{Y}) = \arg \min_{\nu} CLPL(\nu). \quad (11.23)$$

There are possible alternatives: the marginal maximum *a posteriori* (MMAP), the mean *a posteriori*, etc. They are not presented here.

11.3.1.4. Viterbi algorithm

The Viterbi algorithm gives the MAP (11.23) at low computing cost. It is a traditional dynamic programming algorithm that exactly minimizes criterion (11.22) step by step to obtain (11.23). Its principle is to minimize the criterion with respect to ν_M (depending on ν_{M-1}), then with respect to ν_{M-1} (depending on ν_{M-2}) and so on to gradually approach the desired minimization. For a detailed description, see [FOR 73, RAB 86]. We will simply note here that the observation probability structure $\mathbb{O}_m(p)$ given by (11.18) and (11.17) allows all the probabilities to be calculated by M fast Fourier transform calculations on P points.

11.3.2. Likelihood of the hyperparameters

The estimation of the hyperparameters $\mathbf{r} = [r_a, r_b, r_\nu]$ by maximum likelihood makes use of the fact that the frequencies ν are probabilized to marginalize the joint density in ν and \mathcal{Y} with respect to ν so as to obtain the likelihood of the parameters attached to the data:

$$\begin{aligned} HL_{\mathcal{Y}}(\mathbf{r}) &= \Pr(\mathcal{Y}; \mathbf{r}) = \sum_{\nu} \Pr(\mathcal{Y}, \nu) \\ &= \sum_{p_1=1}^P \dots \sum_{p_M=1}^P \Pr(\mathcal{Y}, \nu_1 = \nu^{p_1}, \dots, \nu_M = \nu^{p_M}) . \end{aligned} \quad (11.24)$$

In general, we work with the opposite of the logarithm of the likelihood, denoted $HCLL$ for *Hyperparameters-Co-Log-Likelihood*:

$$HCLL_{\mathcal{Y}}(\mathbf{r}) = -\log HL_{\mathcal{Y}}(\mathbf{r}),$$

which is minimized with respect to the vector of the hyperparameters \mathbf{r} :

$$\hat{\mathbf{r}}^{\text{ML}} = \arg \min_{\mathbf{r}} HCLL_{\mathcal{Y}}(\mathbf{r}).$$

The known properties of $HCLL_{\mathcal{Y}}(\mathbf{r})$ do not allow a global optimization. A first approach could be to use a coordinate-wise descent algorithm as in the case of adaptive spectral analysis (section 11.2). Here, we have chosen to implement gradient descent techniques. The two following sections are devoted to the calculation of $HCLL_{\mathcal{Y}}(\mathbf{r})$ and its gradient.

11.3.2.1. Forward-Backward algorithm

The sum in (11.24) extends over P^M states of the chain and cannot be calculated in practice. However, the “Forward” algorithm presented in [FOR 73, RAB 86] performs the step-wise marginalization of the joint law and gives access to the likelihood for a relatively low computing cost.

This algorithm, in its normalized form (recommended by [DEV 85] for numerical stability reasons), is based on the probabilities:

$$\tilde{\mathcal{F}}_m(p) = \frac{\Pr(\mathcal{Y}_1^m, \nu_m = \nu^p)}{\Pr(\mathcal{Y}_1^m)} \quad \text{and} \quad \tilde{\mathcal{B}}_m(p) = \frac{\Pr(\mathcal{Y}_{m+1}^M | \nu_m = \nu^p)}{\Pr(\mathcal{Y}_{m+1}^M | \mathcal{Y}_1^m)},$$

the partial observations at instants m to m' being denoted: $\mathcal{Y}_m^{m'} = [\mathbf{y}_m, \dots, \mathbf{y}_{m'}]$. The likelihood can be deduced from the sub-products of the “Forward” phase. The Forward-Backward algorithm also gives the posterior marginal probabilities of the states of the Markov chain, given an observation sequence and the model parameters (\mathbb{P} , \mathbb{P} and \mathbb{O}), (which would allow the MMAP to be determined):

$$p_m(p) = \Pr(\nu_m = \nu^p | \mathcal{Y}) = \tilde{\mathcal{F}}_m(p) \tilde{\mathcal{B}}_m(p) \quad (11.25)$$

together with the double marginal posterior which will be useful for calculating the gradient of the likelihood [LEV 83]:

$$\begin{aligned} p_m(i_{m-1}, i_m) &= \Pr(\nu_{m-1} = \nu^{i_{m-1}}, \nu_m = \nu^{i_m} | \mathcal{Y}) \\ &= \mathcal{N}_m \tilde{\mathcal{F}}_{m-1}(p) \tilde{\mathcal{B}}_m(q) \mathbb{P}(p, q) \mathbb{O}_m(q). \end{aligned} \quad (11.26)$$

11.3.2.2. Likelihood gradient

The gradient calculation is based on the properties of the auxiliary function (generally denoted Q) of the EM (*Expectation Maximization*) algorithm [BAU 70, LIP 82]. It is built on two sets of hyperparameters \mathbf{r} and \mathbf{r}' , by “completing” data \mathcal{Y} by object ν to be marginalized:

$$Q(\mathbf{r}, \mathbf{r}') = \mathbb{E}_{\nu}(\log \Pr(\nu, \mathcal{Y}; \mathbf{r}') | \mathcal{Y}; \mathbf{r}) = \sum_{\nu} \log \Pr(\nu, \mathcal{Y}; \mathbf{r}') \Pr(\nu | \mathcal{Y}; \mathbf{r}).$$

Here, we obtain the following expression for Q :

$$\begin{aligned} Q(\mathbf{r}, \mathbf{r}') &= \sum_{m=2}^M \sum_{i_{m-1}=1}^P \sum_{i_m=1}^P p_m(i_{m-1}, i_m) \log \mathbb{P}'(i_{m-1}, i_m) \\ &\quad + \sum_{p=1}^P \mathbb{P}(p) \log \mathbb{P}'(p) + \sum_{m=1}^M \sum_{i_m=1}^P p_m(i_m) \log \mathbb{O}'_m(i_m) \end{aligned} \quad (11.27)$$

– $(\mathbb{P}', \mathbb{P}', \mathbb{O}')$ and $(\mathbb{P}, \mathbb{P}, \mathbb{O})$ are the characteristics of the model for hyperparameters \mathbf{r}' and \mathbf{r} , respectively;

– $p_m(i_m)$ and $p_m(i_{m-1}, i_m)$ are the posterior marginal laws defined by (11.25) and (11.26) for hyperparameters \mathbf{r} .

The standard estimation strategy using the EM algorithm does not apply directly here as the chain is not parametrized by its natural parameters (\mathbb{P}, \mathbb{p}) , but by the hyperparameters \mathbf{r} . Function Q is still of great interest, however, because it possesses the following property:

$$\left. \frac{\partial Q(\mathbf{r}, \mathbf{r}')}{\partial \mathbf{r}'} \right|_{\mathbf{r}'=\mathbf{r}} = - \frac{\partial HCLLy(\mathbf{r})}{\partial \mathbf{r}}.$$

This property gives the gradient of $HCLLy(\mathbf{r})$ when (11.27) is derived:

$$\frac{\partial Q}{\partial r'_a} = \sum_{m=1}^M \sum_{i_m=1}^P p_m(i_m) \frac{\partial \log \mathbb{O}'_m(i_m)}{\partial r'_a} \quad (11.28)$$

$$\frac{\partial Q}{\partial r'_b} = \sum_{m=1}^M \sum_{i_m=1}^P p_m(i_m) \frac{\partial \log \mathbb{O}'_m(i_m)}{\partial r'_b} \quad (11.29)$$

$$\frac{\partial Q}{\partial r'_\nu} = \sum_{m=2}^M \sum_{i_{m-1}=1}^P \sum_{i_m=1}^P p_m(i_{m-1}, i_m) \frac{\partial \log \mathbb{P}'(i_{m-1}, i_m)}{\partial r'_\nu}. \quad (11.30)$$

Derivatives \mathbb{O}' and \mathbb{P}' are obtained by deriving (11.15) and (11.21) respectively:

$$\begin{aligned} \frac{\partial \log \mathbb{O}'}{\partial r_a} &= -\frac{N}{Nr_a + r_b} - \frac{Nr_b}{(Nr_a + r_b)^2} P_m(\nu_m^i) \\ \frac{\partial \log \mathbb{O}'}{\partial r_b} &= -\frac{N-1}{r_b} - \frac{1}{Nr_a + r_b} + \frac{1}{r_b^2} \mathbf{y}_m^\dagger \mathbf{y}_m + \frac{Nr_a}{(Nr_a + r_b)^2} P_m(\nu_m^i) \\ \frac{\partial \log \mathbb{P}'}{\partial r_\nu} &= \frac{1}{2r_\nu^2} \left((\nu^{i_m} - \nu^{i_{m-1}})^2 - \sum_{q=1}^P (\nu^q - \nu^{i_{m-1}})^2 \mathbb{P}(i_{m-1}, q) \right). \end{aligned}$$

11.3.3. Processing results and comparisons

This subsection presents some typical results. As in the case of AR adaptive spectral analysis, the section is divided into two parts: first, the estimation of the hyperparameters, then the reconstruction of the frequencies. The signals processed are the same as in the adaptive spectral analysis except for the fact that we have only kept one sample in two to “simulate” a real spectral aliasing situation.

11.3.3.1. Tuning the hyperparameters

First of all, the hyperparameter likelihood HCLL was calculated on a grid of $25 \times 25 \times 25$ values and the likelihood contours are given in Figure 11.6. Like that in Figure 11.3, the function is regular and has a single, clearly marked minimum: $\hat{r}_a^{\text{ML}} = 0.292$, $\hat{r}_b^{\text{ML}} = -0.700$ and $\hat{r}_\nu^{\text{ML}} = -2.583$ (on a logarithmic scale). It is also

worth noting that a variation of $1/10$ (still on the logarithmic scale) produces an imperceptible modification in the corresponding frequency profile, which demonstrates a certain robustness of the method.

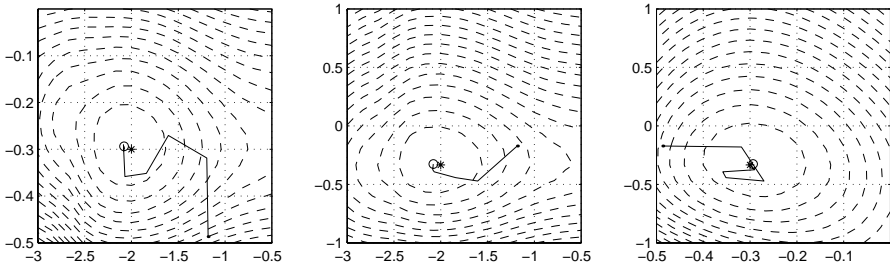


Figure 11.6. Likelihood of the hyperparameters: typical behavior. HCLL contour lines (—), minimum (*), initializations (•), trajectory of the descent algorithm (---) and minimum reached (○). All the figures are on a logarithmic scale and from left to right: on the horizontal axis r_v , r_b , and on the vertical axis r_b , r_a , r_a

As far as the optimization itself is concerned, several descent directions were compared: usual gradient, Vignes' correction, bisector correction, and the pseudo-conjugate direction of Polak-Ribière [BER 95]. Our investigations show that, as was to be expected, the direction of the gradient generates zig-zag trajectories in the parameter space, whereas this is not so for the three corrected directions. These three corrections allow a 25% to 40% gain in calculation time relative to the direction of the gradient, with a clear advantage for the Polak-Ribière pseudo-conjugate direction. Three line search techniques were also compared: dichotomic search, and quadratic and cubic interpolation. The first was the fastest. Finally, about 3 s sufficed for the optimum to be reached. The convergence of the algorithm is illustrated in Figure 11.6.

11.3.3.2. Qualitative comparison

Figure 11.7 compares typical results. Figure 11.7a was obtained by maximum likelihood, i.e., by choosing the frequency that maximized each of the 64 periodograms (on $\nu \in [0, 1]$). By construction, this solution obviously does not allow frequencies to be tracked outside $\nu \in [0, 1]$. It is particularly irregular, and an algorithm progressing in the standard way (by gradual approximation) cannot track the frequencies beyond $\nu = 1$, as Figure 11.7b shows. In contrast, the solution shown in Figure 11.7c is much more satisfactory:

- the effect of the regularization is clear. The solution is more regular and incontestably closer to the reference of Figure 11.2;
- even beyond the limit frequency $\nu = 1$, the MAP tracks the true frequencies correctly;

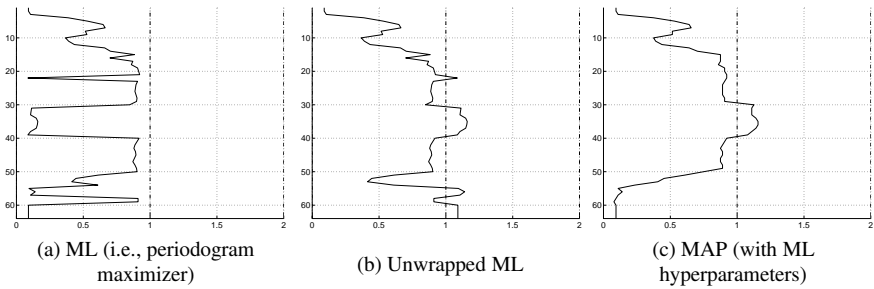


Figure 11.7. Comparison of estimated frequency profiles

– the dynamics are respected: frequencies up to $\nu = 1.15$ can be read (i.e., double the reference frequencies of Figure 11.2, since we subsampled once).

In addition, it is obtained in an entirely automatic way (the hyperparameters are at maximum likelihood) and only takes a tenth of a second of work for the Viterbi algorithm.

11.4. Conclusion

The developments of this chapter (see also [GIO 02, BER 01]) were chosen for their general character. The concepts developed for the spectral analysis have also been used for analyzing meteorological radar clutter [GIO 01] and could be used in other fields such as speech analysis. The methods proposed for estimating the mean velocity have been used for characterizing skin tissue by its acoustic attenuation [GIO 94]. The tracking capacity beyond the Shannon frequency is directly transposable to magnetic resonance velocimetry [HER 99], where it can significantly improve the SNR on the velocity images and reduce the duration of the clinical study.

Specific improvements can be made to these methods by introducing more specific prior information. For example, the possible break in flow continuity at the vessel wall is not taken into consideration but it could be. For parabolic laminar flows, the velocity decreases regularly from the center of the flow towards the wall. The hypothesis of a connection between the flow velocity and the vessel wall makes sense. However, in plug flows there are very high velocities near the vessel wall and the possibility of large discontinuities in the velocity near the wall must be maintained. Joint detection (at the same time as the velocity estimation) of the limits of the vessel can thus be envisaged, allowing an exception to be made to spatial continuity.

It has been established that there are variations in the Doppler signal amplitude during the cardiac cycle. An increase is observed during the acceleration of the blood,

with a maximum occurring early after peak systole. The flow regimen (laminar or turbulent) also influences the Doppler signal amplitude. It would thus be possible to take this phenomenon into account to further adapt the temporal regularization.

Experiment and expert assessment of the results obtained point out balances between the various alternatives for data processing, particularly regularization.

The first element is a balance between a) the quality of the results and b) the computing cost and method complexity. This requires efficient algorithms that are suited to the calculations to be made (Kalman filter, Viterbi algorithm, gradient, etc.).

The second element concerns the “visual” aspect of the results. It encourages the use of convex functionals. Regularization by a more discriminating functional (such as Blake and Zisserman’s truncated quadratic [BLA 87]) gives a “binary” aspect to the maps that is not very helpful for subsequent interpretation.

The third element is the sensitivity of the solution to the hyperparameters and the capacity to estimate these hyperparameters. Using likelihood criteria such as those presented here seems to be an efficient way of tuning the hyperparameters. However, in routine use, it is indispensable to offer the user a set of suitable hyperparameters straight away. The authors’ experience suggests that the choices proposed above lead to methods that are sufficiently robust with respect to their hyperparameters and that, for a given type of application, a single set of hyperparameters provides significant, if not optimal, improvement in the results.

11.5. Bibliography

- [AKA 70] AKAIKE H., “Statistical predictor identification”, *Ann. Inst. Stat. Math.*, vol. 22, p. 207-217, 1970.
- [AKA 74] AKAIKE H., “A new look at the statistical model identification”, *IEEE Trans. Automat. Contr.*, vol. AC-19, num. 6, p. 716-723, Dec. 1974.
- [AZE 86] AZENCOTT R., DACUNHA-CASTELLE D., *Series of Irregular Observations: Forecasting and Model Building*, Springer Verlag, New York, NY, 1986.
- [BAU 70] BAUM L. E., PETRIE T., SOULES G., WEISS N., “A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains”, *Ann. Math. Stat.*, vol. 41, num. 1, p. 164-171, 1970.
- [BER 95] BERTSEKAS D. P., *Nonlinear Programming*, Athena Scientific, Belmont, MA, 1995.
- [BER 01] BERTHOMIER C., HERMENT A., GIOVANNELLI J.-F., GUIDI G., POURCELOT L., DIEBOLD B., “Multigate Doppler signal analysis using 3-D regularized long AR modeling”, *Ultrasound Med. Biol.*, vol. 27, num. 11, p. 1515-1523, 2001.

- [BLA 87] BLAKE A., ZISSERMAN A., *Visual Reconstruction*, The MIT Press, Cambridge, MA, 1987.
- [DEV 85] DEVIJVER P. A., “Baum’s forward-backward algorithm revisited”, *Pattern Recognition Letters*, vol. 3, p. 369-373, Dec. 1985.
- [FOR 73] FORNEY G. D., “The Viterbi algorithm”, *Proc. IEEE*, vol. 61, num. 3, p. 268-278, Mar. 1973.
- [GIO 94] GIOVANNELLI J.-F., IDIER J., QUERLEUX B., HERMENT A., DEMOMENT G., “Maximum likelihood and maximum a posteriori estimation of Gaussian spectra. Application to attenuation measurement and color Doppler velocimetry”, in *Proc. Int. Ultrasonics Symp.*, vol. 3, Cannes, France, p. 1721-1724, Nov. 1994.
- [GIO 01] GIOVANNELLI J.-F., IDIER J., DESODT G., MULLER D., “Regularized adaptive long autoregressive spectral analysis”, *IEEE Trans. Geosci. Remote Sensing*, vol. 39, num. 10, p. 2194-2202, Oct. 2001.
- [GIO 02] GIOVANNELLI J.-F., IDIER J., BOUBERTAKH R., HERMENT A., “Unsupervised frequency tracking beyond the Nyquist limit using Markov chains”, *IEEE Trans. Signal Processing*, vol. 50, num. 12, p. 1-10, Dec. 2002.
- [GUO 94] GUO Z., DURAND J.-G., LEE H. C., “Comparison of time-frequency distribution techniques for analysis of simulated Doppler ultrasound signals of the femoral artery”, *IEEE Trans. Biomed. Eng.*, vol. BME-41, num. 4, p. 332-342, Apr. 1994.
- [HER 97] HERMENT A., GIOVANNELLI J.-F., DEMOMENT G., DIEBOLD B., DELOUCHE A., “Improved characterization of non-stationary flows using a regularized spectral analysis of ultrasound Doppler signals”, *Journal de Physique III*, vol. 7, num. 10, p. 2079-2102, Oct. 1997.
- [HER 99] HERMENT A., MOUSSEAU E., DE CESARE A., JOLIVET O., DUMEE P., TODD-POKROPEK A., BITTOUN J., GUGLIELMI J. P., “Spatial regularization of flow patterns in magnetic resonance velocity mapping”, *J. of Magn. Reson. Imaging*, vol. 10, num. 5, p. 851-860, 1999.
- [JAZ 70] JAZWINSKI A. H., *Stochastic Process and Filtering Theory*, Academic Press, New York, NY, 1970.
- [KAY 81] KAY S. M., MARPLE S. L., “Spectrum analysis – A modern perspective”, *Proc. IEEE*, vol. 69, num. 11, p. 1380-1419, Nov. 1981.
- [KIT 85] KITAGAWA G., GERSCH W., “A Smoothness Priors Time-Varying AR Coefficient Modeling of Nonstationary Covariance Time Series”, *IEEE Trans. Automat. Contr.*, vol. AC-30, num. 1, p. 48-56, Jan. 1985.
- [LEV 83] LEVINSON S. E., RABINER L. R., SONDHI M. M., “An introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognition”, *Bell Syst. Tech. J.*, vol. 62, num. 4, p. 1035-1074, Apr. 1983.
- [LIP 82] LIPORACE L. A., “Maximum likelihood estimation for multivariate observations of Markov sources”, *IEEE Trans. Inf. Theory*, vol. 28, p. 729-734, Sep. 1982.
- [MAR 87] MARPLE S. L., *Digital Spectral Analysis with Applications*, Prentice-Hall, Englewood Cliffs, NJ, 1987.

- [PAR 74] PARZEN E., "Some recent advances in time series modeling", *IEEE Trans. Automat. Contr.*, vol. AC-19, num. 6, p. 723-730, Dec. 1974.
- [RAB 86] RABINER L. R., JUANG B. H., "An introduction to hidden Markov models", *IEEE ASSP Mag.*, vol. 3, num. 1, p. 4-16, 1986.
- [RIS 78] RISSANEN J., "Modeling by shortest data description", *Automatica*, vol. 14, p. 465-471, 1978.
- [SOR 80] SORENSON H. W., *Parameter Estimation*, Marcel Dekker, New York, NY, 1980.
- [TAL 88] TALHAMI H. E., KITNEY R. I., "Maximum likelihood frequency tracking of the audio pulsed Doppler ultrasound signal using a Kalman filter", *Ultrasound Med. Biol.*, vol. 14, num. 7, p. 599-609, 1988.
- [ULR 76] ULRYCH T. J., CLAYTON R. W., "Time series modelling and maximum entropy", *Phys. Earth Planet. Interiors*, vol. 12, p. 188-200, 1976.

Chapter 12

Tomographic Reconstruction from Few Projections

12.1. Introduction

The aim of tomography is to non-destructively reconstruct a map of a parameter that is characteristic of an object, e.g. its density. Its principle is based on analyzing the interaction between the object and radiation (X-ray, electronic or optical, for example) that is made to propagate through it. The characteristic quantity we are trying to find can be obtained from the observations by inversion of the equations of transport in the material.

Tomography has many fields of application. In medicine, systems such as ultrasound scanners or MRI allow us to observe internal clues to pathologies (tumours, aneurisms, etc.). In nuclear medicine, the behavior of a tracer inside the body can be followed by functional imaging so as to show up possible dysfunctions of organs or tissues. In industry, tomography is used in the development of manufacturing processes (welding, molding, etc.), in production quality control, and for safety and security (checking the contents of luggage). It is also an important investigation tool in many research fields: architecture of bone structure, flow analysis, etc.

In several configurations, the number of projections available is limited, e.g. because of restrictions on the dose of radiation that can be transmitted to the patient in medical applications, or because of testing rates to be respected or the transient nature of the phenomenon to be observed in industry. Moreover, geometrical constraints

may limit the distribution of the projections to certain angles. This chapter concentrates particularly on reconstruction methods suited to such contexts. The methods will be presented for the case where the irradiation is by X-ray transmission.

We will first present the relationship between the observations and the attenuation data we are trying to find. Then we will recall the analytical 2D and 3D reconstruction methods, point out their limits in the case of a small number of projections, and introduce the discrete reconstruction formalism. In the subsections that follow, we will present reconstruction approaches suited to a discrete framework and a limited number of projections. Finally, we will use some examples to illustrate the results obtained by the different approaches.

12.2. Projection generation model

If an object composed of homogeneous material is illuminated by X-rays and a narrow monoenergetic ray crosses the object, the intensity of the ray measured as it leaves the object can be obtained from:

$$I = I_0 \exp \{-\mu L\} \quad (\text{Beer-Lambert law}) \quad (12.1)$$

where I_0 is the intensity of the input ray, L the distance covered by the ray in the object, and μ the linear attenuation coefficient of the object, which depends on the density of the material, its nuclear composition and the energy of the X-ray flux. This relationship can be generalized to a cross-section of a non-homogeneous object perpendicular to the axis oz crossed by the rays.

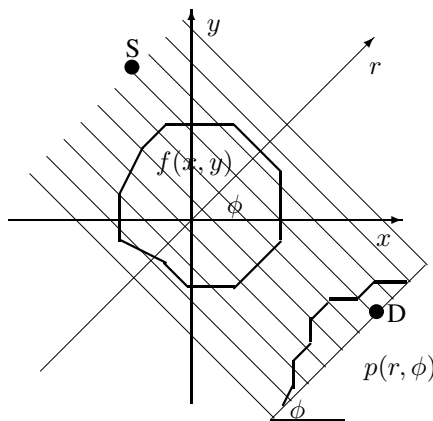


Figure 12.1. X-ray tomography

If we characterize the object by the distribution of its linear attenuation coefficient μ , which is a continuous function of the two space variables $\mu(x, y) = f(x, y)$; we have:

$$I = I_0 \exp \left\{ - \int_L f(x, y) dl \right\} \implies -\log \left(\frac{I}{I_0} \right) = \int_L f(x, y) dl \quad (12.2)$$

where dl is the basic length of path L . By noting $p = -\log I/I_0$ and moving the emitter (source S) and receiver (detector D) in parallel along a straight line making an angle ϕ with the ox axis (see Figure 12.1), we obtain what is called a *projection* $p(r, \phi)$:

$$p(r, \phi) = \int_L f(x, y) dl = \iint_D f(x, y) \delta(r - x \cos \phi - y \sin \phi) dx dy \quad (12.3)$$

X-ray tomography image reconstruction consists of finding an estimation $\hat{f}(x, y)$ of $f(x, y)$ from the projections $p(r, \phi_i)$, $i = 1, \dots, M$. This can easily be extended to the 3D case where the aim is to find an estimation $\hat{f}(x, y, z)$ of $f(x, y, z)$ from the X-ray images $p(r_1, r_2, \vec{u}_i)$, $i = 1, \dots, M$, where \vec{u}_i represents the direction of projection i . In this chapter, we will particularly take the context where we have few projections and they are limited to angles ϕ_i .

12.3. 2D analytical methods

If we consider the ideal case where function $p(r, \phi)$ is known perfectly for all values of r and ϕ , the reconstruction problem comes down to the inversion of the Radon transform (RT) \mathcal{R} [DEA 83]:

$$p(r, \phi) = \iint f(x, y) \delta(r - x \cos \phi - y \sin \phi) dx dy \quad (12.4)$$

for which the analytical expression of the inverse is:

$$f(x, y) = \frac{1}{2\pi^2} \int_0^\pi \int_0^\infty \frac{\partial p(r, \phi) / \partial r}{r - x \cos \phi - y \sin \phi} dr d\phi. \quad (12.5)$$

In practice, we only have the measurements of $p(r, \phi)$ for discrete values of ϕ and r . Thus, the integrals have to be approximated by sums. Two major difficulties then become apparent: the approximation of $\partial p(r, \phi) / \partial r$ and the approximate calculation of the integral in r (because of the singularity of its kernel). In addition, although the discretization along the r axis can be reasonably fine, the discretization of angle ϕ is generally more parsimonious. These constraints have resulted in the direct application of the expression for the inverse of the RT being unused in practice for many years, as

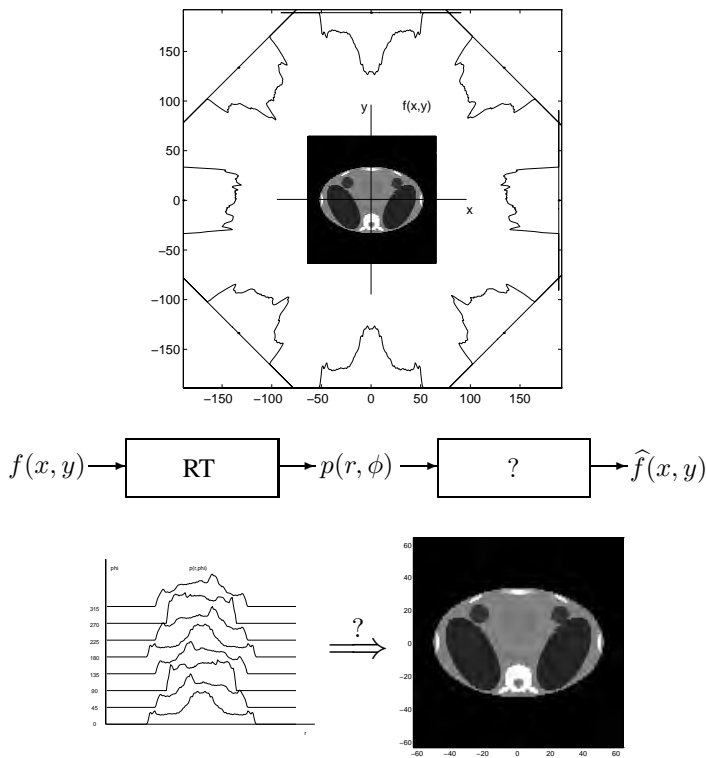


Figure 12.2. Problem of X-ray tomography image reconstruction in 2D

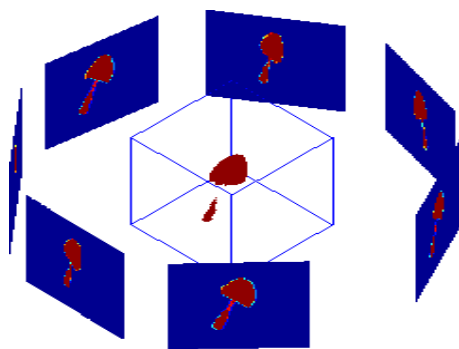


Figure 12.3. Problem of X-ray tomography image reconstruction in 3D

outside the ideal, theoretical case, direct application of the inversion formula does not give convincing results.

A considerable number of methods have been put forward [BRO 78, BUD 79, CHO 74, HER 80, HER 87, MOH 88, NAT 80] with the aim of obtaining more satisfactory approximate solutions. Some operate directly on the equations, others use auxiliary transformations: Hilbert transform (HT), or Fourier transform (FT). These methods as a whole can be summed up by defining the following operators:

$$\begin{aligned}
 - \text{derivation } \mathcal{D}: & \quad \bar{p}(r, \phi) = \partial p(r, \phi) / \partial r, \\
 - \text{HT}\mathcal{H}: & \quad g(r', \phi) = \frac{1}{\pi} \int_0^\infty \frac{\bar{p}(r, \phi)}{(r - r')} dr, \\
 - \text{backprojection (BP) } \mathcal{B}: & \quad f(x, y) = \frac{1}{2\pi} \int_0^\pi g(x \cos \phi + y \sin \phi, \phi) d\phi, \\
 - \text{1D FT } \mathcal{F}_1: & \quad P(\Omega, \phi) = \int p(r, \phi) e^{-j\Omega r} dr, \\
 - \text{1D inverse FT } \mathcal{F}_1^{-1}: & \quad p(r, \phi) = \frac{1}{2\pi} \int P(\Omega, \phi) e^{j\Omega r} d\Omega, \\
 - \text{2D FT } \mathcal{F}_2: & \quad F(\omega_x, \omega_y) = \iint f(x, y) e^{-j(\omega_x x + \omega_y y)} dx dy, \\
 - \text{2D inverse FT } \mathcal{F}_2^{-1}: & \quad f(x, y) = \frac{1}{4\pi^2} \iint F(\omega_x, \omega_y) e^{-j(\omega_x x + \omega_y y)} d\omega_x d\omega_y.
 \end{aligned}$$

With these definitions, the following analytical relations can be demonstrated:

$$\begin{aligned}
 f &= \mathcal{B} \mathcal{H} \mathcal{D} \mathcal{R} f = \mathcal{B} \mathcal{F}_1^{-1} |\Omega| \mathcal{F}_1 \mathcal{R} f = \mathcal{B} \mathcal{C}_1 \mathcal{R} f \\
 &= \mathcal{F}_2^{-1} |\Omega| \mathcal{F}_2 \mathcal{B} \mathcal{R} f = \mathcal{C}_2 \mathcal{B} \mathcal{R} f
 \end{aligned}$$

where $\mathcal{C}_1 = \mathcal{F}_1^{-1} |\Omega| \mathcal{F}_1$ and $\mathcal{C}_2 = \mathcal{F}_2^{-1} |\Omega| \mathcal{F}_2$ are convolution operators:

$$\begin{aligned}
 \mathcal{C}_1 p(r, \phi) &= (h_1 \star p)(r, \phi), \quad h_1(r) = \int |\Omega| e^{-j\Omega r} d\Omega \\
 \mathcal{C}_2 b(x, y) &= (h_2 \star b)(x, y), \quad h_2(x, y) = \iint \sqrt{\omega_x^2 + \omega_y^2} e^{-j(\omega_x x + \omega_y y)} d\omega_x d\omega_y.
 \end{aligned}$$

Finally, we should also mention another relation known as the *projection slice theorem*:

$$F(\omega_x, \omega_y) = P(\Omega, \phi) \quad \text{for} \quad \omega_x = \Omega \cos \phi \quad \text{and} \quad \omega_y = \Omega \sin \phi \quad (12.6)$$

which is at the heart of many inversion techniques in tomography and turns the RT inversion problem into a *Fourier synthesis* problem [MOH 88]. These relationships summarize the various algorithms that are used to obtain approximate solutions for the RT inversion, and thus for the image reconstruction problem in X-ray tomography. They are shown schematically below:

– direct inversion of RT:

$$p(r, \phi) \rightarrow \boxed{\mathcal{D}} \rightarrow \bar{p}(r, \phi) \rightarrow \boxed{\mathcal{H}} \rightarrow g(r', \phi) \rightarrow \boxed{\mathcal{B}} \rightarrow f(x, y)$$

– backprojection of filtered projections:

$$p(r, \phi) \rightarrow \boxed{\mathcal{F}_1} \rightarrow \boxed{\text{filter } |\Omega|} \rightarrow \boxed{\mathcal{F}_1^{-1}} \rightarrow g(r', \phi) \rightarrow \boxed{\mathcal{B}} \rightarrow f(x, y)$$

– backprojection and filtering by convolution:

$$p(r, \phi) \rightarrow \boxed{\text{filter 1D } h_1(r)} \rightarrow g(r', \phi) \rightarrow \boxed{\mathcal{B}} \rightarrow f(x, y)$$

– backprojection followed by 2D filtering:

$$p(r, \phi) \rightarrow \boxed{\mathcal{B}} \rightarrow b(x, y) \rightarrow \boxed{\mathcal{F}_2} \rightarrow \boxed{\begin{array}{c} \text{filter} \\ |\Omega| = \sqrt{\omega_x^2 + \omega_y^2} \end{array}} \rightarrow \boxed{\mathcal{F}_2^{-1}} \rightarrow f(x, y)$$

– backprojection followed by 2D convolution filtering:

$$p(r, \phi) \rightarrow \boxed{\mathcal{B}} \rightarrow b(x, y) \rightarrow \boxed{\text{filter 2D } h_2(x, y)} \rightarrow f(x, y)$$

– padding and interpolation in the Fourier domain and 2D inverse FT:

$$p(r, \phi) \rightarrow \boxed{\mathcal{F}_1} \rightarrow P(\Omega, \phi) \rightarrow \boxed{\begin{array}{c} \text{interpolation} \\ \omega_x = \Omega \cos \phi \\ \omega_y = \Omega \sin \phi \end{array}} \rightarrow F(\omega_x, \omega_y) \rightarrow \boxed{\mathcal{F}_2^{-1}} \rightarrow f(x, y)$$

It is worth noting, however, that we come up against two difficulties when implementing these methods numerically:

– the projections have limited supports, both in the angles and in their finite number. The integrals are approximated by sums and the various transforms are approximated by their discrete versions;

– the final image is often represented in Cartesian coordinates in the form of pixels or voxels, and there is necessarily an interpolation stage in the central backprojection step. This interpolation stage is more explicit in the methods using Fourier synthesis, which pass through the Fourier domain. The interpolation is itself an inverse problem. It makes prior information on the object necessary, a fact often eclipsed in the algorithms.

We have not mentioned the specific methods which, by using certain properties of the object $f(x, y)$, transform the RT inversion problem into the inversion of another transform, often of lower dimension. For example, if the object possesses revolution symmetry ($f(x, y) = f(\rho)$ and $p(r, \phi) = p(r)$), the problem can be reduced to the inversion of the Abel transform in one dimension:

$$p(r) = \int_0^r \frac{f(\rho)}{\sqrt{r - \rho}} d\rho.$$

12.4. 3D analytical methods

A first approach in 3D reconstruction is to carry out successive reconstructions of 2D slices using the reconstruction techniques mentioned above and starting from projections obtained on linear strip sensors. However, this is beset by practical difficulties as it involves successive rotations of the detection system around the object followed by translation movements slice by slice.

Thus, for an acquisition system composed of an X-ray source and a linear sensor, helicoidal acquisition geometry has been proposed: while the acquisition system is rotating, the patient's bed moves at a constant velocity. Provided that there is at least one half to one turn of the acquisition system for a bed movement of the thickness of reconstruction slice, the reconstruction techniques can be adapted to yield a 3D object. The difference between this and the slice-by-slice geometry is that we have to take account of the fact that one projection is involved in several reconstructed slices [CRA 90, KAL 95]. To shorten the acquisition time, current systems are evolving towards multiline or even two-dimensional sensors. The problem of an acquisition ray intersecting several slices is then even trickier.

In certain configurations, the whole reconstruction zone can be handled by one 2D detector. The analytical reconstruction algorithms must take the conical geometry into account, i.e., the fact that a projection ray crosses several slices of the object to be reconstructed. Algorithms have been proposed for this type of geometry: [FEL 84] is suitable for small angles of aperture of the beams and [GRA 91] deals with larger angles.

12.5. Limitations of analytical methods

Analytical methods have been used successfully in medicine but their use remains more limited in other applications such as non-destructive evaluation. This can be explained by stricter acquisition constraints in non-destructive evaluation: the angles of incidence, the number of measurements and the signal-to-noise ratio (SNR) are often limited. In consequence, the analytical solutions are rarely satisfactory. Figure 12.4 illustrates these difficulties.

It is also worth mentioning that, even in medical imaging, the number of projections is limited and the acquisition system has to be optimized, since a great effort is made to keep the quantity of X-rays received by the patient to a minimum. This is the case, for example, in helicoidal tomography, where traditional methods cannot be applied directly. To make up for the small amount of data, prior information has to be added. This means developing methods more specific to the class of objects to be examined: objects where variations are smooth, objects with abrupt changes, objects that are continuous or piecewise constant, binary objects, composed of

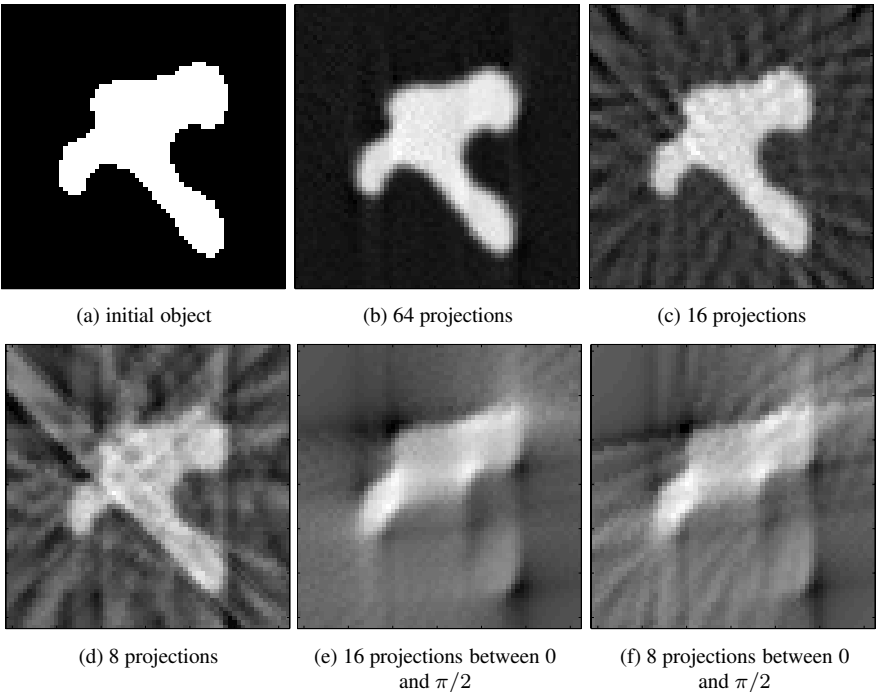


Figure 12.4. Five cases of reconstruction by filtered backprojection in 2D: 64, 16 and 8 projections spread between 0 and π , and 16 and 8 projections between 0 and $\pi/2$

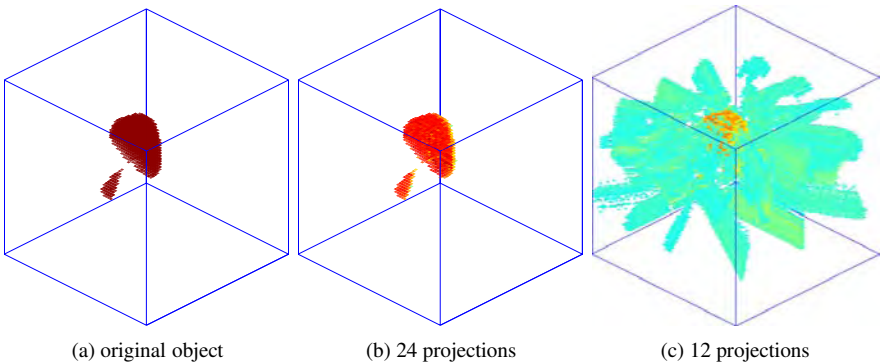


Figure 12.5. Two cases of reconstruction by filtered backprojection in 3D: 24 and 12 projections, distributed between 0 and 2π around the object

geometrical shapes (ellipses or polygons in 2D, ellipsoids or polyhedra in 3D), etc. Analytical methods do not readily allow prior information to be introduced and used. This type of prior information can be handled more easily by a formal description of the object and discrete modeling, which thus seems more suitable.

12.6. Discrete approach to reconstruction

Since there is a finite amount of data and the calculations are performed numerically, there is no alternative but to discretize the problem, in the broad sense of the term. Three approaches are possible:

- represent the object by an appropriate parametric model, model the connection between the data and the parameters of the model and estimate the parameters directly from the data;
- model the image by projecting it onto a basis of appropriate functions with a reasonable number of coefficients to be estimated;
- discretize the object by a set of pixels or voxels using the finest desired resolution and bring in prior information suited to the nature of the object under examination: continuous, piecewise continuous, sparse, binary, etc.

An example of methods based on the first approach is the modeling of an object by superimposing a finite number of ellipsoids and the estimation of their parameters $\{(x_i, y_i, z_i), (a_i, b_i, c_i), (\theta_i, \phi_i), f_i\}$ by least squares or maximum likelihood. The main difficulty lies in the fact that the relationship between the data and these parameters is not linear (although explicit and analytical). We thus have to make use of non-trivial optimization techniques to calculate the solution.

The last two of the three approaches mentioned above have one point in common: projection of the object $f(x, y)$ on a basis of functions:

$$f(x, y) \simeq \sum_{n=1}^N x_n b_n(x, y).$$

In the latter case, the basis functions, $b_n(x, y)$, are indicative of the supports of the pixels or voxels (zero-order splines). This projection allows the problem to be transformed into a discrete system of linear equations:

$$\begin{aligned} & \text{for } i = 1, \dots, M_1; j = 1, \dots, M_2, \\ p(r_j, \phi_i) & \simeq \iint f(x, y) \delta(r_j \cos \phi_i - r_j \sin \phi_i) dx dy \\ & = \sum_{n=1}^N x_n \iint b_n(x, y) \delta(r_j \cos \phi_i - r_j \sin \phi_i) dx dy. \end{aligned}$$

By reorganizing and renaming the data in the form $y_m = p(r_j, \phi_i)$, $m = (i - 1) \times M_1 + j$, we find $\mathbf{y} \simeq \mathbf{Ax}$, where $\mathbf{y} = [y_1, \dots, y_M]^T$ and $\mathbf{x} = [x_1, \dots, x_N]^T$ are column vectors containing the data and the unknown coefficients of the decomposition respectively, and \mathbf{A} is the matrix with elements:

$$A_{mn} = \iint b_n(x, y) \delta(r_j \cos \phi_i - r_j \sin \phi_i) dx dy,$$

where m corresponds to the observation index $p(r_j, \phi_i)$. The problem thus comes down to obtaining a satisfactory approximate solution for an equation of the type $\mathbf{y} = \mathbf{Ax} + \mathbf{b}$, where \mathbf{b} represents both the errors connected with the various approximations and the measurement noise proper. We can thus use all the methods described in the preceding chapters, while taking advantage of the particular structure of matrix \mathbf{A} and the specific models for noise \mathbf{b} . We consider some of these below.

Of course, the choice of the basis functions $b_n(x, y)$ (pixels, natural pixels, harmonic functions, wavelets, etc.) has important consequences for the structure and properties of matrix \mathbf{A} and for the meaning and computational complexity of its elements.

Choice of basis functions

Two approaches can be distinguished: either the basis is chosen according to a prior on the object, independently of the geometry (e.g. the pixels or voxels), or it is induced by the geometry (e.g. the natural pixels) [GAR 87]. In the first case, the elements of \mathbf{x} can have a physical meaning but the elements of matrix \mathbf{A} are often more costly to calculate. In the second case, the calculation is less costly but the basis is not necessarily orthogonal and complete and it is more difficult to give a physical meaning to the elements of \mathbf{x} .

In the first case, three categories of functions can again be distinguished: global functions (e.g. Fourier series), local functions (e.g. splines) and hybrid functions (e.g. wavelets).

In all these cases, the meaning of the elements of \mathbf{x} depends on this choice. In the same way, the meaning of the elements and the structure of matrix \mathbf{A} depend on the choice and the geometry of the acquisition of the projections. For example, in conic geometry, for complete data (uniform angular coverage between 0 and 2π), the matrix \mathbf{A} will have a block-circulant structure.

From now on, to present the principle of the inversion methods, we will use the general case without taking this feature into consideration. Of course, it is absolutely necessary to take it into account in implementation, so as to obtain algorithms with reasonable computing costs.

12.7. Choice of criterion and reconstruction methods

As we have just seen, the discrete formulation of the problem leads us to look for a satisfactory solution $\hat{\mathbf{x}}$ for $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$. Matrix \mathbf{A} has large dimensions and is often ill-conditioned or even singular. Its conditioning depends particularly on the basis chosen for the object decomposition, the geometry and the ratio of the number of independent data to the number of parameters describing the object. In the preceding chapters, we have seen that it is often illusory to think that we can obtain a satisfactory solution to the problem by looking for an exact solution $\mathbf{A}^{-1}\mathbf{y}$, a least squares solution $\arg \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2$, or even a generalized inverse solution $\mathbf{A}^\dagger \mathbf{y}$. A satisfactory solution can only be obtained by somehow introducing prior information on the solution. This may be done through a judicious choice of the basis functions so that the object can actually be described by a small number of parameters that are simply estimated in the least squares sense. However, this approach is too restrictive and specific: a basis of functions has to be built for a particular object. The alternative, when these functions are chosen independently of the operator (pixels or voxels for example), is the regularization approach that allows more generic priors to be expressed.

In what follows, we will focus more on the methods that define the solution as the optimizer of a regularized criterion. Three categories of solutions can thus be distinguished:

- 1) Those defined as the unconstrained minimizer of a composite criterion:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} J(\mathbf{x}) = \mathcal{Q}(\mathbf{y}, \mathbf{A}\mathbf{x}) + \lambda \mathcal{F}(\mathbf{x}, \mathbf{x}^0)$$

where \mathcal{Q} and \mathcal{F} are generally two distances. Various expressions can be obtained for \mathcal{Q} and \mathcal{F} depending on the hypotheses on the nature of the noise, the data and the nature of the object. A traditional example is $J(\mathbf{x}) = \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 + \lambda \|\mathbf{D}\mathbf{x}\|^2$, where \mathbf{D} is a derivation operator matrix to make the solution smooth. In this case, $\hat{\mathbf{x}}$ can be expressed analytically:

$$\hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{A} + \lambda \mathbf{D}^T \mathbf{D})^{-1} \mathbf{A}^T \mathbf{y},$$

and this expression can be interpreted in terms of analytical methods if we observe that \mathbf{A}^T corresponds to a backprojection operation and $(\mathbf{A}^T \mathbf{A} + \lambda \mathbf{D}^T \mathbf{D})^{-1}$ to a filtering operation in 2D.

- 2) Those defined as the constrained minimizer of a simple criterion:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \mathcal{F}(\mathbf{x}, \mathbf{x}^0) \quad \text{s. t. } \mathbf{y} = \mathbf{A}\mathbf{x}$$

where \mathcal{F} is also generally a distance or a measure of divergence between \mathbf{x} and a default solution \mathbf{x}^0 . A traditional example is $\mathbf{x}^0 = 0$ and $\mathcal{F}(\mathbf{x}) = \|\mathbf{D}\mathbf{x}\|^2$. In this case too, the solution has an analytical expression:

$$\hat{\mathbf{x}} = (\mathbf{D}^T \mathbf{D})^{-1} \mathbf{A}^T (\mathbf{A}(\mathbf{D}^T \mathbf{D})^{-1} \mathbf{A}^T)^{-1} \mathbf{y},$$

which is equivalent to a backprojection of the filtered projections. Another traditional example consists of choosing:

$$\mathcal{F}(\mathbf{x}, \mathbf{x}^0) = KL(\mathbf{x}, \mathbf{x}^0) = \sum_j x_j \log x_j / x_j^0 + x_j - x_j^0$$

where $KL(\mathbf{x}, \mathbf{x}^0)$ is the Kullback-Leibler divergence of \mathbf{x} with respect to \mathbf{x}^0 . This brings us to the maximum entropy methods.

3) Those defined as the minimizer of a probabilistic criterion:

$$\hat{\mathbf{x}} = \arg \min_{\tilde{\mathbf{x}}} \bar{C}(\tilde{\mathbf{x}})$$

where $\bar{C}(\tilde{\mathbf{x}})$ is the expectation of a cost function $C(\tilde{\mathbf{x}}, \mathbf{x})$ with respect to the posterior probability law $p(\mathbf{x} | \mathbf{y})$:

$$\bar{C}(\tilde{\mathbf{x}}) = \int C(\tilde{\mathbf{x}}, \mathbf{x}) p(\mathbf{x} | \mathbf{y}) d\mathbf{x}$$

It is the choice of this cost function that brings us to different estimation structures. With a Gaussian hypothesis for the noise and object laws and a quadratic cost, we find the previous solutions again. However, this probabilistic approach allows us to go beyond deterministic solutions through other choices for the prior laws and other cost functions, or through the idea of *marginalization*.

Types of estimators

In the Bayesian probabilistic approach, according to the choice made for the cost function $C(\tilde{\mathbf{x}}, \mathbf{x})$, a variety of expressions can be obtained for estimator $\hat{\mathbf{x}}$ depending on the data. Among these estimators, the following are of particular interest:

– maximum *a posteriori* (MAP):

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} p(\mathbf{x} | \mathbf{y}) = \arg \max_{\mathbf{x}} p(\mathbf{x}, \mathbf{y});$$

– posterior mean (PM):

$$\hat{\mathbf{x}} = E(\mathbf{x} | \mathbf{y}) = \int \mathbf{x} p(\mathbf{x} | \mathbf{y}) d\mathbf{x} = \frac{\int \mathbf{x} p(\mathbf{x}, \mathbf{y}) d\mathbf{x}}{\int p(\mathbf{x}, \mathbf{y}) d\mathbf{x}};$$

– marginal maximum *a posteriori* (marginal MAP or Marginal Posterior Mode, MPM):

$$\hat{f}_j = \arg \max_{x_j} p(x_j | \mathbf{y}).$$

One of the elements that help in the choice of these various estimators is the computing cost. Depending on the type of estimator chosen, it is necessary to perform either an optimization or an integration. The optimization is either multivariate, as in the MAP case, or scalar, as in the marginal MAP case. Note, too, that in the marginal MAP case, there is an integration step that is the calculation of the marginal laws [BOU 96, DIN 90, GEM 87, SAQ 98].

Except for the linear Gaussian case, it is often difficult to find analytical expressions for these integrals and the criteria to be optimized are non-quadratic. Special attention thus needs to be paid to the integration and optimization aspects. For this reason, the next two subsections describe a number of algorithms that implant these different estimators.

12.8. Reconstruction algorithms

A reconstruction algorithm can be defined as a sequence of operations carried out on projections in order to construct an image. An algorithm that contented itself with this definition without stating the method on which the algorithm was based or the properties of the images obtained by using it would be of little value. We will thus restrict ourselves to algorithms that optimize a deterministic or probabilistic criterion. In the first subsection below, we give an overview of the optimization algorithms that can be used for convex criteria in particular and, in the second, we take a very brief look at algorithms using probabilistic criteria.

12.8.1. Optimization algorithms for convex criteria

We have seen that an important family of methods for solving inverse problems defines a solution to the problem by minimizing a criterion that is least squares $J(\mathbf{x}) = \|\mathbf{y} - \mathbf{Ax}\|^2$, regularized least squares $\|\mathbf{y} - \mathbf{Ax}\|^2 + \lambda\Phi(\mathbf{x})$ or, more generally, $\mathcal{Q}(\mathbf{y}, \mathbf{Ax}) + \lambda\mathcal{F}(\mathbf{x}, \mathbf{x}_0)$, with \mathcal{Q} and \mathcal{F} two distances or two measures of divergence. Similarly, in probabilistic approaches, we find MAP $J(\mathbf{x}) = -\log p(\mathbf{x} | \mathbf{y})$ or marginal MAP $J(x_j) = -\log p(x_j | \mathbf{y})$ criteria. In this section, we give a synthetic presentation of the optimization algorithms that can be used particularly for convex criteria. The convex criteria can be subdivided into three groups of increasing generality and complexity: i) strictly convex and quadratic, ii) continuous, strictly convex but non-quadratic, iii) continuous, convex but possibly non-derivable at a few points.

In the first case, the solution exists, is unique and is a linear function of the data. It can be calculated by any descent algorithm. It is even possible to obtain an explicit, direct solution.

In the second case, the solution exists and is unique but is not a linear function of the data. It can easily be calculated using any descent algorithm.

In the third case, the solution generally exists but may not be unique. Its calculation requires some caution and the solution is not a linear function of the data.

Let us not forget that a non-convex criterion may be multimodal, which means that we need to resort to global optimization algorithms. The solution is not generally a linear function of the data and the solution is not easy to calculate.

From an algorithmic point of view, the algorithms for calculating the optimum solution can be divided into two categories: those that change the set of unknowns \mathbf{x} at each iteration and those that change a single element x_j (or a block of elements) at a time. The algorithms of each category can again be subdivided into those that use the whole set of data \mathbf{y} at each iteration and those that use only one data item y_i (or a block of data) at each iteration.

12.8.1.1. Gradient algorithms

An important family of algorithms for optimizing a criterion updates the estimate $\mathbf{x}^{(k-1)}$ additively in a descent direction (often the opposite direction to that of the gradient of the criterion).

To look more closely at the traditional descent methods in tomographic reconstruction, let us consider the special case when the criterion to be optimized has the generic form:

$$J(\mathbf{x} | \mathbf{y}) = \sum_i q_{y_i}(z_i) + \lambda \sum_j \phi_j(t_j),$$

with $z_i = [\mathbf{A}\mathbf{x}]_i = \mathbf{a}_{i*}^T \mathbf{x}$ and $t_j = [\mathbf{D}\mathbf{x}]_j = x_j - x_{j-1}$, where $\mathbf{a}_{i*}^T = [a_{i1}, \dots, a_{iN}]$ is row i of matrix \mathbf{A} , \mathbf{D} a finite differences matrix, q_y and ϕ_j convex functions that reach their minimum at y and 0, respectively, and λ a regularization parameter.

To distinguish between the various classes of algorithms mentioned above, we will use the following notation:

– for the algorithms that use all the data \mathbf{y} and update all the unknowns \mathbf{x} at each iteration, we write:

$$\mathbf{g}(\mathbf{x} | \mathbf{y}) = \left[\frac{\partial J(\mathbf{x} | \mathbf{y})}{\partial x_1}, \dots, \frac{\partial J(\mathbf{x} | \mathbf{y})}{\partial x_N} \right]^T = \mathbf{A}^T \mathbf{q}'_y(z) + \lambda \mathbf{D}^T \phi'(t)$$

for the gradient of $J(\mathbf{x} | \mathbf{y})$ with respect to \mathbf{x} , with $\phi'(t) = [\phi'_1(t_1), \dots, \phi'_N(t_N)]$ and $\mathbf{q}'_y = [q'_{y_1}, \dots, q'_{y_M}]$;

– for those that use only a single data item y_i but update all the unknowns \mathbf{x} :

$$J(\mathbf{x} | y_i) = q_{y_i}(z_i) + \lambda \sum_j \phi_j(t_j),$$

$$\mathbf{g}(\mathbf{x} | y_i) = q'_{y_i}(z_i) \mathbf{a}_{i*} + \lambda \mathbf{D}^T \phi'(t);$$

– for the algorithms that use all the data but only update a single variable x_k at each iteration:

$$J(x_k | \mathbf{y}; \mathbf{x}_{\setminus k}) = J(\mathbf{x} | \mathbf{y})$$

$$g(x_k | \mathbf{y}; \mathbf{x}_{\setminus k}) = \frac{\partial J(\mathbf{x} | \mathbf{y})}{\partial x_k} = \sum_i a_{ik} q'_{y_i}(z_i) + \lambda \sum_j d_{jk} \phi'_j(t_j);$$

– for the algorithms that use a single data item y_i and update only a single variable x_k at each iteration:

$$J(x_k | y_i; \mathbf{x}_{\setminus k}) = q_{y_i}(z_i) + \lambda \sum_j \phi_j(t_j)$$

$$g(x_k | y_i; \mathbf{x}_{\setminus k}) = \frac{\partial J(x_k | y_i; \mathbf{x}_{\setminus k})}{\partial x_k} = a_{ik} q'_{y_i}(z_i) + \lambda \sum_j d_{jk} \phi'_j(t_j).$$

We are now going to look at a number of special cases used in tomography. Here, a_{ij} represents the length of the path of ray i in pixel j . We will observe that most of the conventional algorithms in tomography are in fact gradient algorithms applied to least squares criteria: $q_{y_i}(z_i) = (y_i - z_i)^2$, or are based on a Kullback-Leibler divergence: $q_{y_i}(z_i) = KL(y_i, z_i)$.

12.8.1.2. SIRT (Simultaneous Iterative Relaxation Techniques)

Proposed by [GIL 72], this is a gradient algorithm operating on $J(\mathbf{x} | \mathbf{y}) = \|\mathbf{y} - \mathbf{Ax}\|^2$:

$$\mathbf{x}^{(k)} = \mathbf{x}^{(k-1)} + \alpha^{(k)} \mathbf{D} \mathbf{A}^T (\mathbf{y} - \mathbf{Ax}^{(k-1)}), \quad k = 1, 2, \dots$$

with $\mathbf{D} = \text{Diag}\{1/\sum_i a_{ij}\}$. For $\alpha^{(k)}$, there are several possible variants: fixed, optimal, etc.

The calculations performed by this algorithm at each iteration are mainly a projection \mathbf{Ax} and a backprojection $\mathbf{A}^T(\mathbf{y} - \mathbf{Ax})$. When α is fixed, it should verify $0 < \alpha \|\mathbf{DA}^T \mathbf{A}\| < 2$. The word *simultaneous* signifies that all the data is used at each iteration of the algorithm.

12.8.1.3. ART (Algebraic Reconstruction Technique)

This algorithm, widely used in tomography, corresponds to the minimization of $J(\mathbf{x} | y_i) = (y_i - \mathbf{a}_{i*}^T \mathbf{x})^2$ by a variable or optimal step-size gradient algorithm:

$$\mathbf{x}^{(k)} = \mathbf{x}^{(k-1)} + \alpha^{(k)} \frac{y_i - \mathbf{a}_{i*}^T \mathbf{x}^{(k-1)}}{\|\mathbf{a}_{i*}\|^2} \mathbf{a}_{i*}, \quad k = 1, 2, \dots; \quad i = 1 \bmod M,$$

with $\mathbf{x}^{(0)} = \mathbf{0}$. It is also worth noting that, when $\alpha^{(k)} = 1$, $\mathbf{x}^{(k)}$ can be obtained from $\mathbf{x}^{(k-1)}$ in the direction of projection of $\mathbf{x}^{(k-1)}$ on the subset defined by the

equation $y_i = \mathbf{a}_{i*}^T \mathbf{x}$, $i = k \bmod M$. At each update of \mathbf{x} , only one data item is used. There are also versions in which certain constraints (e.g. positivity) are imposed on the solution at each iteration. ART is a special case of projection onto convex sets (POCS), originally presented by Kaczmarz [KAC 37].

The order in which the data is used is crucial for the efficiency of the algorithm. The underlying idea is that the successive iterations should be as independent as possible [HER 93, MAT 96].

12.8.1.4. ART by blocks

ART modifies the values of pixels \mathbf{x} by backprojecting the differences between the observed and calculated values of a single projection ray. A variant backprojects the difference for a whole block of projection rays. The updating equation thus becomes:

$$\mathbf{x}^{(k)} = \mathbf{x}^{(k-1)} + \alpha^{(k)} \frac{\mathbf{A}_{i_k}^T (\mathbf{y}_{i_k} - \mathbf{A}_{i_k} \mathbf{x}^{(k-1)})}{\|\mathbf{A}_{i_k}^T \mathbf{A}_{i_k}\|},$$

where \mathbf{y}_{i_k} is a block of data and $\mathbf{A}_{i_k}^T$ the matrix corresponding to the ray projection equations for the block.

ART corresponds to the case where i_k is scalar and the S-ART corresponds to the case where i_k represents the set of rays for a projection direction. Herman [HER 93] showed that the reconstruction could be accelerated by choosing a successively orthogonal series of blocks. This type of method is known as *Ordered Subset M ART*, where M indicates the number of block projection directions.

12.8.1.5. ICD (Iterative Coordinate Descent) algorithms

These are relaxation algorithms (Chapter 2). For the basic case $J(\mathbf{x} | \mathbf{y}) = \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2$, this structure is equivalent to that of the Gauss-Seidel algorithm [PAT 99].

12.8.1.6. Richardson-Lucy algorithm

Let us consider the following criterion:

$$J(\mathbf{x} | \mathbf{y}) = KL(\mathbf{y}, \mathbf{A}\mathbf{x}) = - \sum_i y_i \log \mathbf{a}_{i*}^T \mathbf{x} / y_i + y_i - \mathbf{a}_{i*}^T \mathbf{x},$$

for which we have:

$$\begin{aligned} \frac{\partial J(\mathbf{x} | \mathbf{y})}{\partial x_j} &= \sum_i a_{ij} \frac{y_i}{\mathbf{a}_{i*}^T \mathbf{x}} - 1, \\ \frac{\partial^2 J(\mathbf{x} | \mathbf{y})}{\partial x_j^2} &= - \sum_i a_{ij}^2 \frac{y_i}{(\mathbf{a}_{i*}^T \mathbf{x})^2}. \end{aligned}$$

The approximation $(\partial^2 J(\mathbf{x} | \mathbf{y}) / \partial x_j^2)^{-1} \simeq -x_j / \sum_i a_{ij}$ gives an approximate Newton algorithm that can be written:

$$x_j^{(k)} = x_j^{(k-1)} \frac{1}{\sum_i a_{ij}} \sum_i a_{ij} \frac{y_i}{\sum_\ell a_{i\ell} x_\ell^{(k-1)}}, \quad k = 1, 2, \dots; \quad j = k \bmod N,$$

and is known as the Richardson-Lucy algorithm [LUC 74, RIC 72]. It can also be interpreted as an EM algorithm in a probabilistic approach with Poisson distributions [SHE 82] or as a projection onto convex sets (POCS) algorithm using the Kullback-Leibler divergence [SNY 92].

12.8.2. Optimization or integration algorithms

Probabilistic models make it easier to describe situations where the image to be reconstructed is a real quantity that is continuous, piecewise constant or takes discrete values. A typical example is Markov modeling (see Chapter 7). In these cases, it is often necessary to optimize a criterion (MAP or marginal MAP estimators), or to calculate a posterior mean (PM estimator). We have already described a number of algorithms for optimizing criteria when they are convex. In this subsection, we will focus on the algorithms that are used either to optimize multimodal criteria or to calculate PM estimators.

As mentioned above, stochastic approaches lead to the construction of a posterior probability distribution of the type $\exp \{-J(\mathbf{x})\}$ over the whole set of reconstructions, which expresses both the fidelity to data (likelihood) and the regularization terms. The reconstruction is defined as an estimator coming from this distribution: MAP estimator (maximum of the probability distribution), MPM estimator (maximum of the marginal posterior distribution), or PM estimator (mean of the posterior distribution), for example. According to the type of prior considered (Markov, non-convex function), the above estimators cannot be calculated directly from the distribution, $\exp \{-J(\mathbf{x})\}$, but are determined by sampling. Two types of algorithms are mainly used for this:

- the Gibbs sampler is particularly suitable for Markov priors as the expression for the conditional probabilities is then simple. This is presented in section 7.4.2. By simulating a series of configurations representative of the probability distribution P , we can estimate the marginals of P , and thus define the MPM estimator of the posterior marginal maximum. Similarly, this algorithm enables the mean of the posterior distribution to be estimated;

- the simulated annealing algorithm can construct a random series having a law that converges towards a uniform distribution on the minima of P . It thus enables the MAP to be estimated (see section 7.4.2.3).

12.9. Specific models for binary objects

The methods we have presented so far are very general and can be applied to most image reconstruction problems. However, in some applications, particularly in NDE, we want to reconstruct the image of a defect (air pocket) in a homogeneous medium (metal). The problem then comes down to the reconstruction of a binary object. Three categories of methods for handling this problem can be distinguished:

1) Methods that model the object by a set of binary voxels [BOU 96] using Markov modeling for the binary image and finally estimate this image using one of the estimators mentioned in the preceding sections. In these methods, the direct model linking the projections to the voxels is linear, but it is often difficult to take account of the prior information on the closing of the defect's outline using a simple Markov model.

2) Other methods propose a direct reconstruction of the closed contour of the object, either by modeling the contour as the solution to a partial derivative equation (active contours, *Snakes*), or by modeling it as the passage of a higher dimension function to zero (*level-set methods*) [SAN 96]. The main drawbacks of these methods are (i) the high computing cost (e.g. because of the updating of the function representing the wavefront in the level-set method) and (ii) the difficulty of implementation and the lack of tools for choosing the step for the propagation of the wavefront.

3) Finally, there are methods that model the contour by deformable geometrical shapes, the parameters of which are estimated from the data [AMI 93, BAT 98, HAN 94]. Several categories of geometrical models have been put forward: elliptical or super-quadratic curves or surfaces, harmonics and curves or surfaces described by splines. The first is too simple and restrictive, the second is broader and is suitable for star-shaped curves and surfaces, and the last is more general. The price to be paid is an increase in the number of parameters.

A comparison of these various methods can be found in [MOH 97], together with a specific method for cases where the object is modeled by a polygon in 2D or a polyhedron in 3D (see [SOU 04] for a detailed presentation of the latter approach).

12.10. Illustrations

The principal objective of these illustrations is to show what can be achieved using the different methods, particularly in the case where there are very few projections at limited angles.

12.10.1. 2D reconstruction

We chose a binary object with dimensions 128×128 (Figure 12.6a), simulated five projections for the angles $-\pi/4$, $-\pi/8$, 0 , $\pi/8$, $\pi/4$, added Gaussian white noise

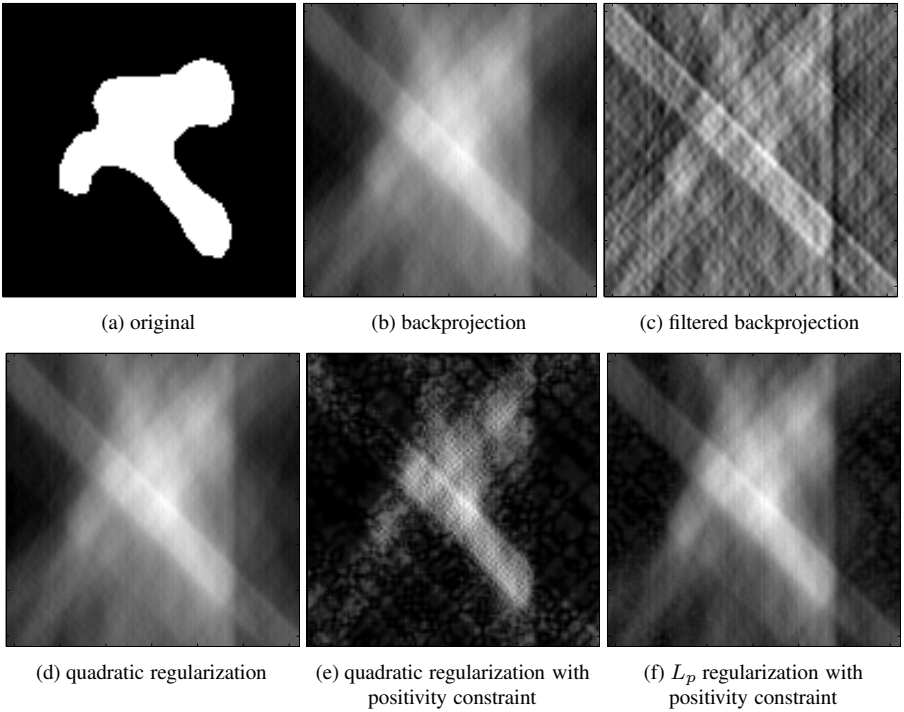


Figure 12.6. *Comparison of 2D reconstruction methods*

to obtain an SNR of 20 dB and then used this data to perform reconstructions using the various methods.

12.10.2. 3D reconstruction

We chose a binary object with dimensions $128 \times 128 \times 128$ (Figure 12.7). Nine projections were distributed uniformly, either between 0 and π or between 0 and $\pi/2$. In both cases, we simulated the data without and with noise (SNR of 20 dB). We then tested the various methods.

The reconstructions by ART (Figure 12.8) show strong perturbations induced by the noise when regularization is not used. In the next reconstruction, by ICM (Figure 12.9), we used a regularization particularly well suited to the binary situation: an Ising model (see section 7.3.4.1), which favors compact zones. In this reconstruction, the effects of the noise are significantly attenuated. However, the cap of the mushroom is separated from its stalk.

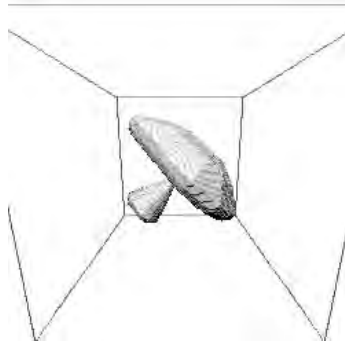


Figure 12.7. *Original object for 3D tests*

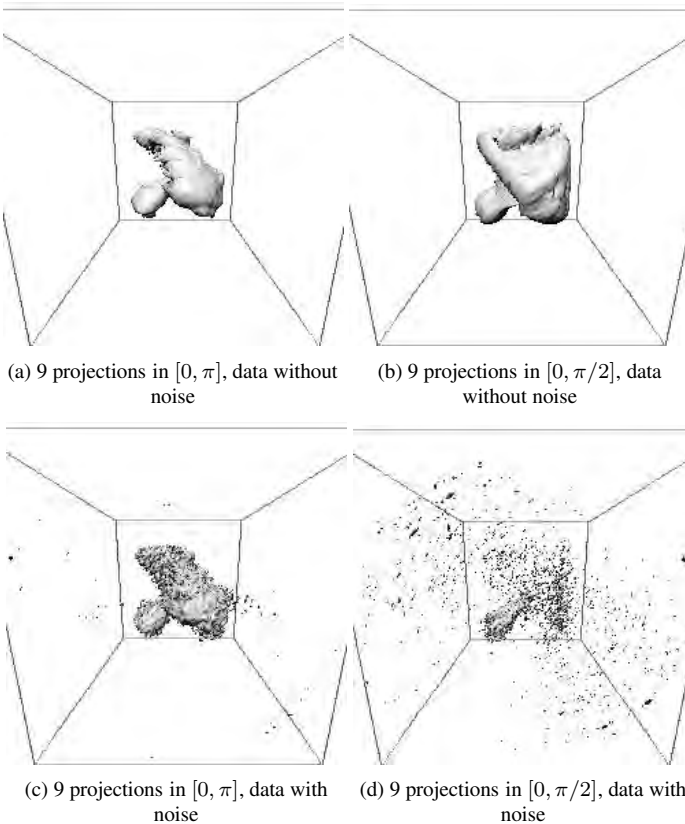


Figure 12.8. *Reconstruction by ART without regularization. Thresholding was applied for display purposes*

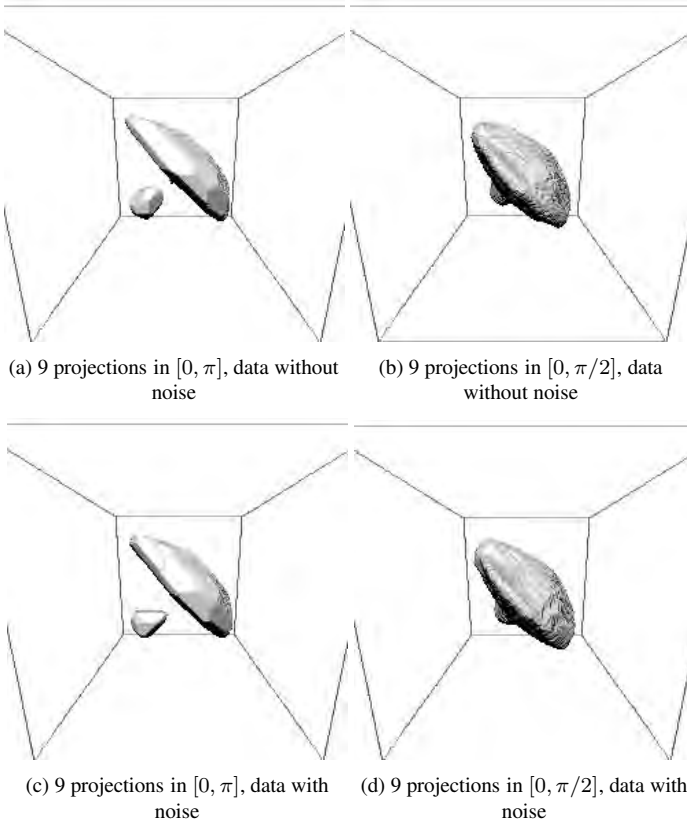


Figure 12.9. *Reconstruction by ICM and regularization by an Ising model*

12.11. Conclusions

The aim of this chapter was to give an overview of the various reconstruction methods in X-ray tomography, focusing on the methods that can be used in difficult situations: when the number of projections is small or the angular coverage is very restricted, i.e., when conventional analytical methods are unable to provide satisfactory results. These more sophisticated methods obviously induce higher computing costs, which, even today, limit their practical use in medical systems or in everyday NDE. However, the trend for the future will always be to reduce the amount of radiation transmitted to a patient or to reduce the cost and acquisition time in NDE applications. For these reasons, it will be necessary to use the more sophisticated methods. Nevertheless, it is still a good idea to continue developing more specific methods with reasonable computing cost for specific applications. This is true, for example, for the

methods recently developed for reconstructing the closed surface of a compact, homogeneous object directly from X-ray data, without going through a voxel reconstruction step.

12.12. Bibliography

- [AMI 93] AMIT Y., MANBECK K. M., “Deformable template models for emission tomography”, *IEEE Trans. Medical Imaging*, vol. 12, num. 2, p. 260-268, June 1993.
- [BAT 98] BATTLE X. L., CUNNINGHAM G. S., HANSON K. M., “Tomographic reconstruction using 3D deformable models”, *Phys. Med. Biol.*, vol. 43, p. 983-990, 1998.
- [BOU 96] BOUMAN C. A., SAUER K. D., “A unified approach to statistical tomography using coordinate descent optimization”, *IEEE Trans. Image Processing*, vol. 5, num. 3, p. 480-492, Mar. 1996.
- [BRO 78] BROOKS R. A., WEISS G. H., TALBERT A. J., “A new approach to interpolation in computed tomography”, *J. Comput. Assist. Tomogr.*, vol. 2, p. 577-585, 1978.
- [BUD 79] BUDINGER T. F., GULLBERG W. L., HUESMAN R. H., “Emission computed tomography”, in HERMAN G. T. (Ed.), *Image Reconstruction from Projections: Implementation and Application*, New York, NY, Springer Verlag, p. 147-246, 1979.
- [CHO 74] CHO Z. H., “General views on 3-D image reconstruction and computerized transverse axial tomography”, *IEEE Trans. Nuclear Sciences*, vol. 21, p. 44-71, 1974.
- [CRA 90] CRAWFORD C. R., KING K. F., “Computed tomography scanning with simultaneous patient translation”, *Med. Phys.*, vol. 17, num. 6, p. 967-982, Jan. 1990.
- [DEA 83] DEANS S. R., *The Radon Transform and Some of its Applications*, Wiley Interscience, New York, NY, 1983.
- [DIN 90] DINTEN J.-M., Tomographie à partir d’un nombre limité de projections : régularisation par champs markoviens, PhD thesis, University of Paris XI, Jan. 1990.
- [FEL 84] FELDKAMP L. A., DAVIS L. C., KRESS J. W., “Practical cone-beam algorithm”, *J. Opt. Soc. Am. (A)*, vol. 1, num. 6, p. 612-619, 1984.
- [GAR 87] GARNERO L., Reconstruction d’images tomographiques à partir d’un ensemble limité de projections, Doctoral thesis, University of Paris XI, Jan. 1987.
- [GEM 87] GEMAN S., MCCLURE D., “Statistical methods for tomographic image reconstruction”, in *Proc. 46th Session of the ICI, Bulletin of the ICI*, vol. 52, p. 5-21, 1987.
- [GIL 72] GILBERT P., “Iterative methods for the three-dimensional reconstruction of an object from projections”, *J. Theor. Biol.*, vol. 36, p. 105-117, 1972.
- [GRA 91] GRANGEAT P., “Mathematical framework of cone beam 3D reconstruction via the first derivative of the Radon transform”, in HERMAN G. T., LOUIS A. K., NATTERER F. (Eds.), *Mathematical Methods in Tomography*, vol. 1497, New York, Springer Verlag, p. 66-97, 1991.

- [HAN 94] HANSON K. M., CUNNINGHAM G. S., JENNINGS G. R. J., WOLF D. R., "Tomographic reconstruction based on flexible geometric models", in *Proc. IEEE ICIP*, vol. 2, Austin, TX, p. 145-147, Nov. 1994.
- [HER 80] HERMAN G. T., *Image Reconstruction from Projections. The Fundamentals of Computerized Tomography*, Academic Press, New York, NY, 1980.
- [HER 87] HERMAN G. T., TUY H. K., LANGENBERG K. J., SABATIER P. C., *Basic Methods of Tomography and Inverse Problems*, Adam Hilger, Bristol, UK, 1987.
- [HER 93] HERMAN G. T., MEYER L. B., "Algebraic reconstruction techniques can be made computationally efficient", *IEEE Trans. Medical Imaging*, vol. 12, p. 600-609, 1993.
- [KAC 37] KACZMARZ S., "Angenaherte Auflosung von Systemen linearer Gleichungen", *Bull. Intern. Acad. Polon. Sci. Lett.*, vol. A-35, p. 355-357, 1937.
- [KAL 95] KALENDER W. A., "Principles and Performance of Spiral CT", in GOODMAN L. W., FOWLKES J. B. (Eds.), *Medical CT and Ultrasound: Current Technology and Applications*, Madison, WI, Advanced Medical Publ., p. 379-410, 1995.
- [LUC 74] LUCY L. B., "An iterative technique for the rectification of observed distributions", *Astron. J.*, vol. 79, num. 6, p. 745-754, 1974.
- [MAT 96] MATEJ S., LEWITT R. M., "Practical considerations for 3-D image reconstruction using spherically symmetric volume elements", *IEEE Trans. Medical Imaging*, vol. 15, p. 68-78, 1996.
- [MOH 88] MOHAMMAD-DJAFARI A., DEMOMENT G., "Maximum entropy reconstruction in X ray and diffraction tomography", *IEEE Trans. Medical Imaging*, vol. MI-7, num. 4, p. 345-354, 1988.
- [MOH 97] MOHAMMAD-DJAFARI A., "Shape reconstruction in X-ray tomography", in *Proc. SPIE*, vol. 3170, San Diego, CA, p. 240-251, July 1997.
- [NAT 80] NATTERER F., "Efficient implementation of optimal algorithms in computerized tomography", *Math. Methods Appl. Sci.*, vol. 2, p. 545-555, 1980.
- [PAT 99] PATRIKSSON M., *Nonlinear Programming and Variational Inequality Problems. A Unified Approach*, Applied Optimization, Kluwer Academic Publishers, Dordrecht, The Netherlands, May 1999.
- [RIC 72] RICHARDSON W. H., "Bayesian-based iterative method of image restoration", *J. Opt. Soc. Am.*, vol. 62, p. 55-59, Jan. 1972.
- [SAN 96] SANTOSA F., "A level-set approach for inverse problems involving obstacles", *ESAIM : COCV*, vol. 1, p. 17-33, Jan. 1996.
- [SAQ 98] SAQUIB S. S., BOUMAN C. A., SAUER K. D., "ML parameter estimation for Markov random fields with applications to Bayesian tomography", *IEEE Trans. Image Processing*, vol. 7, num. 7, p. 1029-1044, July 1998.
- [SHE 82] SHEPP L. A., VARDI Y., "Maximum likelihood reconstruction for emission tomography", *IEEE Trans. Medical Imaging*, vol. MI-1, p. 113-122, 1982.

- [SNY 92] SNYDER D. L., SCHULZ T. J., O'SULLIVAN J. A., "Deblurring subject to nonnegativity constraints", *IEEE Trans. Signal Processing*, vol. 40, num. 5, p. 1143-1150, May 1992.
- [SOU 04] SOUSSEN C., MOHAMMAD-DJAFARI A., "Polygonal and polyhedral contour reconstruction in computed tomography", *IEEE Trans. Image Processing*, vol. 13, num. 11, p. 1507-1523, Nov. 2004.

Chapter 13

Diffraction Tomography

13.1. Introduction

In this chapter, we will look at the use of the Bayesian approach to solve the *diffraction tomography* problem. For this type of problem, the measurements collected, the waves scattered by an object, depend on the physical parameters of the object in a nonlinear way. Many works have considered linear approximations of the propagation phenomenon – the best known being Born and Rytov’s approximations – but we do not intend to follow this path. We will insist on the solution to the problem that takes the nonlinear model into account.

For a nonlinear inverse problem, if we consider an explicit relationship between the parameters of object \mathbf{x} and measurements \mathbf{y} in the form $\mathbf{y} = \mathcal{A}(\mathbf{x})$, the inversion consists of deducing parameters \mathbf{x} from measured data \mathbf{y} . A natural approach to this problem would be to minimize the least squares criterion:

$$J(\mathbf{x}) = \|\mathbf{y} - \mathcal{A}(\mathbf{x})\|^2.$$

In the case of ill posed problems, as illustrated in Chapter 1, this type of solution is not suitable because of its great sensitivity to variations in the measurements, which inevitably contain errors. The Bayesian framework enables a regularized solution to be defined for this problem by using a probabilistic model of the object sought, as in the case of linear inverse problems.

Nonlinearity is, by its very nature, a non-property. Here, we will not go into a generalization of Bayesian regularization for nonlinear inverse problems but will insist on

the specificities of the diffraction tomography problem and how they can be exploited to define and optimize criteria. We will present the wave propagation equations which allow an integral direct model to be deduced in the form of two coupled equations. We will describe the *method of moments*, which is often used to discretize the direct integral model and gives an algebraic model, again in the form of two coupled equations. Estimation in the maximum *a posteriori* (MAP) sense will thus enable us, from this model, to define regularized solutions as minimizing certain criteria. The nonlinearity of the direct model makes these criteria non-convex.

This criterion minimization framework will allow us to look at most of the methods used to solve this problem from the same point of view. We will start by presenting methods that optimize the criteria locally by successive linearizations of the direct model. Then we will point out the connection with techniques that aim to simultaneously reconstruct the object parameters and the field in the domain of the object. Finally, we will stress the importance of using global optimization techniques in particularly difficult situations where the criteria have local minima.

13.2. Modeling the problem

The generic term *diffraction tomography* is used here in connection with a large number of imaging modes for which the phenomenon measured is related to the diffraction of a wave in an inhomogeneous medium. We consider that the object to be imaged is symmetric about one of its axes (cylindrical symmetry) so as to bring the problem down to two dimensions. Moreover, measurements are taken in a domain D_M outside domain D_O in which the object is present, hence the connection with conventional X-ray tomography. Let us also stress the fact that the objective is to construct an image (grid of values) of the physical characteristics of the object, unlike what is done in a number of works where only the outline of the object or the field in the domain of the object is considered.

We will only present here, and consider in what follows, applications involving electromagnetic waves. Nevertheless, it is interesting to note that strong similarities exist with certain applications using acoustic waves (e.g. ultrasonic tomographic imaging [KAK 88]).

13.2.1. Examples of diffraction tomography applications

Applications of diffraction tomography techniques are to be found in many areas, such as biomedical engineering, non-destructive evaluation of conducting materials and geophysical exploration.

13.2.1.1. Microwave imaging

In the 1970s, X-ray tomography enabled images to be taken of the human body. The fact that ionizing radiation can be harmful encouraged studies of other forms of energy such as microwaves (at low power levels), ultrasound, and magnetic resonance. Research on active microwave imaging has developed greatly since the early 1980s [BOL 83].

The aim of such imaging techniques is to determine the propagation characteristics of an inhomogeneous medium (the human body or at least a part of it), with spatial variation of the conductivity $\sigma(\mathbf{r})$ and dielectric permittivity $\epsilon(\mathbf{r})$ ($\mathbf{r} \in D_O$). This inhomogeneous medium is surrounded by a homogeneous medium (e.g. water) having known characteristics σ_0 and ϵ_0 . The geometric configurations envisaged (Figure 13.1a) are taken directly from those of X-ray tomography. Unlike in the case for X-rays, diffraction phenomena cannot be neglected for this mode of imaging; the waves can no longer be assumed to travel through the medium in straight lines, which complicates the imaging problem considerably [KAK 88]. The frequencies involved in this type of application vary between a hundred or so MHz and several hundred GHz.

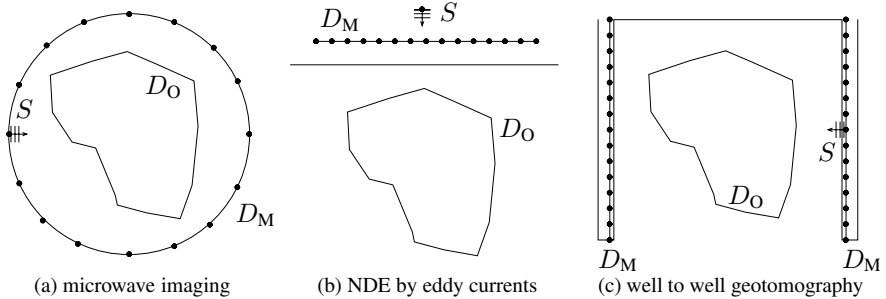


Figure 13.1. Various tomography configurations: a plane wave is emitted by a source S and propagates in object domain D_O . The scattered field is measured by a finite number of receivers placed in domain D_M

13.2.1.2. Non-destructive evaluation of conducting materials using eddy currents

The aim of the non-destructive evaluation (NDE) of conducting materials is to detect and study the defects in these materials. In such a process, an imaging technique can characterize the defects and intervenes after a detection step. For this, an electromagnetic wave is emitted from a homogeneous medium known to be non-conducting (generally air) and having known characteristics $\sigma = 0$ and ϵ_0 . The measurements of the field that has propagated through the conducting medium (Figure 13.1b) are acquired in the same homogeneous medium. The physical characteristic to be imaged is

the conductivity $\sigma(\mathbf{r})$ of a possible defect in the conducting object which has known characteristics ($\sigma_0 = 10^7$ S/m and ϵ_0) [ZOR 91]. The frequency range considered covers kHz to MHz.

13.2.1.3. Geophysical exploration

Geotomography has proved to be an interesting tool in geophysics, for instance when searching for petroleum oil or monitoring its extraction. An example of a possible application is *well to well tomography* where two wells are bored, one on either side of a region to be explored (Figure 13.1c). Electromagnetic waves are emitted by a source placed in one of the wells and the measurements are taken by receivers located in the two wells (measuring the waves reflected and transmitted by the medium). The aim is to image the conductivity $\sigma(\mathbf{r})$ and permittivity $\epsilon(\mathbf{r})$ of the medium [HOW 86]. The frequencies used are of the order of one MHz.

13.2.2. Modeling the direct problem

We will not go into the details of the development of the integral equations to model the imaging problem but we will insist on the hypotheses used and the form of the resulting coupled equations.

13.2.2.1. Equations of propagation in an inhomogeneous medium

We are going to take the case of electromagnetic waves, e.g. in microwave tomography, but similar expressions can be written in other applications and for acoustic waves [COL 92].

Let us consider the propagation of *plane* electromagnetic waves in the *harmonic regime* of pulsation ω in *non-magnetic* media. An inhomogeneous medium is surrounded by an ambient homogeneous medium (having known characteristics: permittivity ϵ_0 , conductivity σ_0 , magnetic permeability of vacuum $\mu_0 = 4\pi \times 10^{-7}$ N.A⁻²). We can adopt a two-dimensional framework by considering that the object is uniform about one of its axes (*cylindrical symmetry*) and the incident wave is polarized along this axis (*transverse magnetic* configuration). This will allow us to work with scalar fields. In this framework, Maxwell's equations give ϕ , the component of the electric field along the axis of the object, from:

$$\Delta\phi(\mathbf{r}) + k^2(\mathbf{r})\phi(\mathbf{r}) = -j\omega\mu_0 J(\mathbf{r}) \quad (13.1)$$

where \mathbf{r} is the position vector in the 2-D space, Δ is the Laplacian, and J is induced by the source. The wave number $k^2(\mathbf{r}) = \omega^2\mu_0 (\epsilon(\mathbf{r}) + j\sigma(\mathbf{r})/\omega)$ of the inhomogeneous object at \mathbf{r} is connected to its electrical permittivity $\epsilon(\mathbf{r})$ and its conductivity $\sigma(\mathbf{r})$ (the surrounding homogeneous medium is characterized by its wave number k_0^2). From the equation above, we can deduce a propagation equation in integral form:

$$\phi(\mathbf{r}) = \phi_0(\mathbf{r}) + \int_D \mathcal{G}(\mathbf{r}, \mathbf{r}') (k^2(\mathbf{r}') - k_0^2) \phi(\mathbf{r}') d\mathbf{r}', \quad (13.2)$$

where ϕ_0 is the incident field; \mathcal{G} is the free space Green's function which, in the two-dimensional case, can be written:

$$\mathcal{G}(\mathbf{r}, \mathbf{r}') = \frac{j}{4} H_0^{(1)}(k_0 |\mathbf{r} - \mathbf{r}'|),$$

where $H_0^{(1)}$ is the zero-order Hankel transform of the first kind.

13.2.2.2. Integral modeling of the direct problem

We have just described the scalar equations corresponding to the propagation of waves in an inhomogeneous medium. Now, we are going to describe the propagation phenomenon involved in the imaging problem. To do this, we will consider the near-field geometrical configuration shown in Figure 13.2.

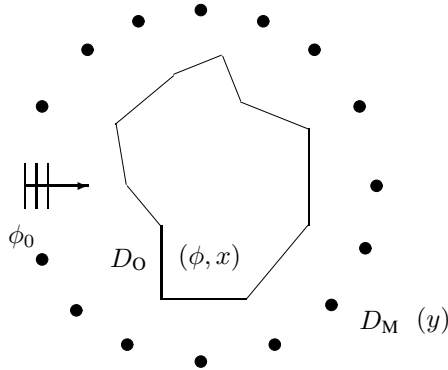


Figure 13.2. Chosen two-dimensional configuration

The measurements are acquired by sensors placed in a discrete set of positions denoted D_M in the surrounding medium¹. From now on we will use $y(\mathbf{r}_i)$ to denote the scattered field at the measuring points $\mathbf{r}_i \in D_M$. The total field in the object domain will be noted $\phi(\mathbf{r})$, $\mathbf{r} \in D_O$. The object is characterized by its complex contrast $x(\mathbf{r}) = k^2(\mathbf{r}) - k_0^2$. We will now describe the propagation of the waves from the source to the measuring points, the incident field in the object domain $\phi_0(\mathbf{r})$, $\mathbf{r} \in D_O$ – corresponding to the emitted wave – being known.

1. It is sometimes necessary to take the measurements outside the surrounding medium. It is thus no longer possible to work with the free space Green's functions and the interfaces between the two media have to be taken into account through their own Green's functions (see for example [ZOR 91] on NDE using eddy currents).

Using equation (13.2), the scattered field at the measuring points can be written in terms of the contrast and the total field in the object:

$$y(\mathbf{r}_i) = \int_{D_0} \mathcal{G}(\mathbf{r}_i, \mathbf{r}') x(\mathbf{r}') \phi(\mathbf{r}') d\mathbf{r}', \quad \mathbf{r}_i \in D_M. \quad (13.3)$$

This equation is often called the *observation equation*. Field ϕ in the object domain verifies an equation of the same type:

$$\phi(\mathbf{r}) = \phi_0(\mathbf{r}) + \int_{D_0} \mathcal{G}(\mathbf{r}, \mathbf{r}') x(\mathbf{r}') \phi(\mathbf{r}') d\mathbf{r}', \quad \mathbf{r} \in D_0, \quad (13.4)$$

sometimes called the *coupling equation*. Note that this equation is implicit in ϕ .

The direct problem is modeled by the relation connecting the inhomogeneous object $x(\mathbf{r})$, $\mathbf{r} \in D_0$ with the scattered field at the measuring points $y(\mathbf{r}_i)$, $\mathbf{r}_i \in D_M$ and is thus written in the form of the two coupled equations (13.3)–(13.4). This model is nonlinear as the scattered field is linearly dependent on the product of the contrast and the field in the object domain (13.3), and the field itself depends on the contrast (13.4).

As a general rule, there is no analytical expression for the solution of the direct problem but various numerical methods exist for solving it [COL 92]. In section 13.3 we will present the discretization of the direct model in integral form and a numerical method for solving the direct problem will follow on naturally from this.

13.3. Discretization of the direct problem

Much of the work in *mathematical physics* on solving inverse problems in general – and the problem of diffraction tomography in particular – is presented in a functional framework. The discretization is only carried out in a final step, for the numerical calculation of a functionally defined solution. A. Tarantola, for example, writes, “*This is a quite general conclusion in any inverse problem theory: discretization, if any, must be reserved for the final computations, and not for developing the formulae*” [TAR 82]. In *statistical data processing*, on the other hand, the discretization is generally taken into account as early as the direct model. This is the framework we have chosen here and we will point out the advantages for the problem at hand. To obtain an algebraic model, we will use the method of moments, as described briefly below, to discretize the coupled equations (13.3)–(13.4).

13.3.1. Choice of algebraic framework

The deliberate choice of using an algebraic rather than a functional framework is tricky to justify and can give rise to considerable discussion. For the problem posed here, several arguments can be put forward:

- the measurements are taken using a finite number of sensors and are thus discrete by nature, whereas a functional framework very often assumes that the measured values are known continuously around the object;
- apart from rare special cases where the object can be parametrized in a simple way (e.g. if we know we are dealing with a homogeneous disk), the solution is calculated numerically;
- for the diffraction tomography problem, the nonlinear equations are not easy to handle and it is interesting to observe, as we will point out in section 13.5.1, that several methods suggested in a functional framework, starting from distinct ways of reasoning, lead to solutions that are functionally different but numerically identical;
- finally, the equations to be handled are simpler in an algebraic framework, even though they are not linear. It is then easier to focus on the fact that the problem is ill posed and on the way to regularize it by introducing prior models elaborated on the solution sought. Note that the probabilistic framework may also be understood from a functional point of view as in [TAR 87] but the models then become rather difficult to handle outside the Gaussian case.

13.3.2. Method of moments

The *method of moments* is a generic method that allows linear equations to be discretized so that they can be solved numerically when their solution cannot be reached in the functional domain. An excellent overview of this method, also showing that most discretization techniques can be interpreted as methods of moments, is to be found in [HAR 87].

The general framework for using this method is to solve a functional equation of the type: $Lf = g$, where L is a linear operator, g a known function and f the unknown function to be determined. The aim is to discretize this equation so as to obtain a matrix relation of the form $\mathbf{L}\mathbf{f} = \mathbf{g}$ which can be solved numerically. To do this, function f is represented as a linear combination of a set of functions $\{b_1, b_2, \dots, b_n\}$, called *basis functions*:

$$f = \sum_{i=1}^n f_i b_i,$$

where coefficients f_i are to be determined. An approximate solution is considered in which the number n of basis functions is finite. The linear equation to be solved thus becomes:

$$\sum_{i=1}^n f_i L b_i = g.$$

By projecting this relation on to a set of functions $\{t_1, t_2, \dots, t_n\}$, called *testing functions*, the equation can be written in matrix form:

$$\mathbf{L}\mathbf{f} = \mathbf{g}, \quad \text{with } L_{i,j} = \langle t_i, L b_j \rangle \text{ and } g_i = \langle t_i, g \rangle;$$

where $\mathbf{L} = (L_{i,j})$ is a matrix of size $n \times n$ and $\mathbf{f} = (f_i)$, $\mathbf{g} = (g_i)$ are column vectors of length n .

The choice of the basis and testing functions is a tricky problem in the method of moments. The choice should be based on practical considerations, such as ease of calculation of the various scalar products, but also on physical considerations so as to take account of the specificity of the problem being dealt with. Finally, these choices will have a direct influence on the structure of the matrix \mathbf{L} and thus on the ease with which the algebraic problem can be solved numerically. We note that, for testing functions of the form $t_i(r) = r^i$, $\mathbf{g}_i = \int r^i g(r) dr$ is the i th order moment of g , hence the name given to the method.

13.3.3. Discretization by the method of moments

Solving the direct problem, i.e., calculating the scattered field at the measurement points that corresponds to a given contrast function, does not fall directly within the application framework of the method of moments since the relation linking the object with the data is not linear. Nevertheless, the method can be applied to each of the coupled equations (13.3-13.4). We will use the discretization performed in [HOW 86] (called the *volume current method* in the article) which is often employed for this problem.

Equation (13.4) is discretized by considering a square grid on object domain D_O . For basis functions $\{b_i\}_{i=1, \dots, n_O}$ and testing functions $\{t_i\}_{i=1, \dots, n_O}$, we take the indicative functions on the square regions D_i (pixels), of side c , of D_O (ϕ , ϕ_0 and x are assumed constant over these regions and represented by column vectors of length n_O). Thus, the algebraic equation corresponding to (13.4) can be written in the form:

$$\phi = \phi_0 + \mathbf{G}_O \mathbf{X} \phi, \quad (13.5)$$

in which \mathbf{X} is a diagonal matrix $\mathbf{X} = \text{Diag} \{x_i\}_{i=1, \dots, n_O}$, and \mathbf{G}_O is a matrix of size $n_O \times n_O$, the elements of which verify:

$$(\mathbf{G}_O)_{i,j} = \frac{1}{c^2} \int_{D_i} \int_{D_j} \mathcal{G}(\mathbf{r}, \mathbf{r}') d\mathbf{r}' d\mathbf{r}.$$

Similarly, relation (13.3) can be written in the algebraic form:

$$\mathbf{y} = \mathbf{G}_M \mathbf{X} \phi, \quad (13.6)$$

with $\mathbf{y} = (y(\mathbf{r}_i))_{i=1, \dots, n_M}$, where n_M is the number of measurement points, \mathbf{G}_M a matrix of size $n_M \times n_O$, the elements of which can be written:

$$(\mathbf{G}_M)_{i,j} = \frac{1}{c^2} \int_{D_M} \int_{D_j} \mathcal{G}(\mathbf{r}, \mathbf{r}') d\mathbf{r}' t_i(\mathbf{r}) d\mathbf{r},$$

where t_i are testing functions on measurement domain D_M .

In [HOW 86] it is suggested that the integrals of the Green's function should be approached by integrating, not over the square regions D_i , but over disks having an equivalent surface area. This gives an analytical formula for the integrals and avoids the use of numerical integration methods. With this approach, the discretization errors can be considered to be negligible if c is of the order of a tenth of the wavelength in the surrounding homogeneous medium [HOW 86].

The direct model was presented in integral form with the coupled equations (13.3)-(13.4). In algebraic form, we obtain the coupled equations (13.5)-(13.6). This model can be expressed in an explicit form directly linking the object with the measurements without bringing in the field in the object domain. For a given object of contrast x , equation (13.5) allows field ϕ to be written in the form:

$$\phi = (\mathbf{I} - \mathbf{G}_O \mathbf{X})^{-1} \phi_0.$$

By replacing this relation in equation (13.6), the explicit relation linking x to y can be deduced:

$$y = \mathcal{A}(x), \quad \text{with } \mathcal{A}(x) = \mathbf{G}_M \mathbf{X} (\mathbf{I} - \mathbf{G}_O \mathbf{X})^{-1} \phi_0. \quad (13.7)$$

Note that, to solve the direct problem, equation (13.7) brings in the inversion of a matrix. In an equivalent way (mathematically speaking but not in terms of computing cost), equation (13.5) requires a linear system of n_O equations with n_O unknowns to be solved. Nevertheless, this inversion does not pose any numerical problems as the linear system is well conditioned.

Finally, it is worth noting that these relations were established for vectors and matrices with complex values. It is sometimes useful to take the case of real variables, by separating the real and imaginary parts. In this case, exactly the same matrix relationships can be obtained by modifying the definition of matrices \mathbf{G}_M and \mathbf{G}_O . These relationships can also be generalized to the multi-source case where we consider measurements acquired for distinct emitting sources [CAR 96].

13.4. Construction of criteria for solving the inverse problem

We now have an algebraic model of the direct problem which allows us to calculate scattered field y at the measuring points from a knowledge of contrast x of the object, possibly using a calculation of the field at object ϕ . Now we are going to turn our attention to solving the inverse problem by reconstructing contrast x (object) from measurements of scattered field y (data). Like most inverse problems, this one is ill-posed. In particular, the solution is very unstable with respect to variations, even very slight ones, in the data [COL 92]. Moreover, even in a functional framework, where

measurements are assumed to be known at all points of a continuous domain D_M surrounding the object, the uniqueness of the solution is not proven. Starting from this algebraic model, we will use the Bayesian framework to construct criteria defining regularized solutions of this inverse problem.

To do this, we need to have a model for the errors that perturb the measurements (modeling errors, discretization error and noise on the measurements). As no particular information is available here on the statistics of measurement noise, unlike in some applications (see Chapter 14, for example), we will consider zero-mean, circular, white, Gaussian errors of variance σ_M^2 . As pointed out in section 3.2, we are not saying that the perturbations actually are Gaussian but the Gaussian *hypothesis* is the least compromising choice we can make for the error distribution, the errors being considered to have a mean value of zero and finite variance.

We also need a prior model of the object. It should be noted that the object to be imaged, \mathbf{x} , has complex values. Given the specific physical meanings of the real part (connected with the permittivity of the object) and the imaginary part (connected with its conductivity), we will consider a separable probability density here

$$p(\mathbf{x}) = p_R(\text{Re}\{\mathbf{x}\}) p_I(\text{Im}\{\mathbf{x}\}).$$

As we want to reconstruct images composed of homogeneous zones, we will consider densities corresponding to a Gibbs-Markov model, as presented in Chapter 7, and written in the form²:

$$p(\mathbf{x}) \propto \exp \{ -\Phi(\mathbf{x})/T \}.$$

However, it is prudent to content ourselves with a convex function Φ at first, to avoid increasing the minimization difficulties due to the nonlinearity of the direct model. To get our ideas clear, let us consider a traditional generalized Gauss-Markov model [BOU 93] (of potential function $|\cdot|^p$, $1 < p \leq 2$).

13.4.1. First formulation: estimation of \mathbf{x}

The parameters to be estimated for the imaging problem are the values of the contrast function on the image pixels, i.e., vector \mathbf{x} .

The likelihood of the object, which corresponds to the direct distribution of the data, can be written:

$$p(\mathbf{y} | \mathbf{x}) = (\pi \sigma_M^2)^{-n_M/2} \exp \left\{ -\frac{1}{\sigma_M^2} \|\mathbf{y} - \mathcal{A}(\mathbf{x})\|^2 \right\}.$$

2. From now on, we will, incorrectly, write the probability density of the object in this form, the separation between the real and imaginary parts and any differences of parameters of these densities being implicit.

Bayes' rule enables the information provided by the measurements and the prior information to be fused into the posterior density of the parameters to be estimated:

$$p(\mathbf{x} | \mathbf{y}) = p(\mathbf{y} | \mathbf{x}) p(\mathbf{x}) / p(\mathbf{y}),$$

where $p(\mathbf{y})$ is a normalization coefficient independent of \mathbf{x} .

The maximum *a posteriori* estimate maximizes the posterior distribution:

$$\hat{\mathbf{x}}^{\text{MAP}} = \arg \max_{\mathbf{x}} p(\mathbf{x} | \mathbf{y}),$$

or, in other words, minimizes the criterion:

$$J^{\text{MAP}}(\mathbf{x}) = \|\mathbf{y} - \mathcal{A}(\mathbf{x})\|^2 + \lambda \Phi(\mathbf{x}),$$

with, here:

$$\mathcal{A}(\mathbf{x}) = \mathbf{G}_M \mathbf{X} (\mathbf{I} - \mathbf{G}_O \mathbf{X})^{-1} \phi_0,$$

and $\lambda = \sigma_M^2/T$ playing the role of the regularization parameter governing the trade-off between fidelity to the data and the prior model. Note that the minimum should be understood in the sense of the global minimum of the criterion. As a shortcut, the J^{MAP} criterion will often be called the maximum *a posteriori* criterion in what follows, or simply the MAP criterion.

13.4.2. Second formulation: simultaneous estimation of \mathbf{x} and ϕ

The problem can also be considered from another point of view by trying to estimate the object, \mathbf{x} , and the field at the object, ϕ , simultaneously from the data. Starting with the same hypotheses as above (additive zero-mean, circular, white Gaussian noise on the measurements), this time we will exploit the relationship linking the object to the measurements using the coupled equations (13.5)-(13.6).

Bayes' rule allows us to express the joint posterior distribution of \mathbf{x} and ϕ in the form:

$$p(\mathbf{x}, \phi | \mathbf{y}) = p(\mathbf{y} | \mathbf{x}, \phi) p(\phi | \mathbf{x}) p(\mathbf{x}) / p(\mathbf{y}). \quad (13.8)$$

In this relation, we are only interested in the three terms of the numerator. The denominator term $p(\mathbf{y})$ is independent of ϕ and \mathbf{x} and does not come into the calculation of the MAP estimate. Let us examine these three terms one by one:

– as the model used for the measurement errors and equation (13.6), the first term can be written:

$$p(\mathbf{y} | \mathbf{x}, \phi) \propto \exp \left\{ -\frac{1}{\sigma_M^2} \|\mathbf{y} - \mathbf{G}_M \mathbf{X} \phi\|^2 \right\};$$

– the second term corresponds to the probability density of ϕ for known \mathbf{x} . ϕ being the total field at the object, it is uniquely determined by the second of the coupled equations (13.5) when \mathbf{x} is known. This leads us to consider a probability measure (and not a probability density) of the type:

$$\delta(\phi - \phi_0 - \mathbf{G}_O \mathbf{X} \phi), \quad (13.9)$$

where δ is Dirac's delta function;

– the last term corresponds to the prior model for the object.

To be able to develop equation (13.8) using probability densities, we are going to use a neat calculation trick. Consider an error e_0 that perturbs the coupling equation (13.5). If we consider these perturbations to be zero-mean, circular, white and Gaussian with a variance σ_0^2 and independent of object \mathbf{x} , we can show that the probability density of ϕ , knowing \mathbf{x} , can be written in the form:

$$p(\phi | \mathbf{x}) \propto \exp \left\{ -\frac{1}{\sigma_0^2} \|\phi - \phi_0 - \mathbf{G}_O \mathbf{X} \phi\|^2 \right\}.$$

The joint \mathbf{x} and ϕ posterior density can be expressed in the form:

$$p(\mathbf{x}, \phi | \mathbf{y}) \propto \exp \{ -J^{\text{MAPJ}}(\mathbf{x}, \phi) \},$$

with:

$$J^{\text{MAPJ}}(\mathbf{x}, \phi) = \frac{1}{\sigma_M^2} \|\mathbf{y} - \mathbf{G}_M \mathbf{X} \phi\|^2 + \frac{1}{\sigma_0^2} \|\phi - \phi_0 - \mathbf{G}_O \mathbf{X} \phi\|^2 + \frac{\Phi(\mathbf{x})}{T}. \quad (13.10)$$

The estimate in the MAP sense of the pair (\mathbf{x}, ϕ) maximizes $p(\mathbf{x}, \phi | \mathbf{y})$, or, in other words, minimizes criterion J^{MAPJ} . We recall that the criterion thus obtained is based on hypotheses – perturbations of the coupling equation (13.5) – that have no physical or statistical justification. They are simply ad hoc hypotheses that allow the J^{MAPJ} criterion to be established. It can be seen that, if these perturbations are applied at the level of data \mathbf{y} , the hypotheses are equivalent to considering errors of the form $\mathbf{e}_M + \mathbf{G}_M \mathbf{X} (\mathbf{I} - \mathbf{G}_O \mathbf{X})^{-1} \mathbf{e}_0$ in which \mathbf{e}_M and \mathbf{e}_0 are random, independent, zero-mean, Gaussian variables, the errors considered on the data depending on object \mathbf{x} . To break free of these hypotheses, we have to make variance σ_0^2 of perturbations \mathbf{e}_0 tend towards zero. The Gaussian distribution $p(\phi | \mathbf{x})$ thus tends towards the Dirac distribution (13.9) in the sense of generalized functions. In this case, the weighting coefficient $1/\sigma_0^2$ of the second term of the J^{MAPJ} criterion tends towards infinity, which is the same as saying that the estimate in the MAP sense of the pair (\mathbf{x}, ϕ) minimizes the criterion:

$$J^{\text{MAPC}}(\mathbf{x}, \phi) = \|\mathbf{y} - \mathbf{G}_M \mathbf{X} \phi\|^2 + \lambda \Phi(\mathbf{x}), \quad (13.11)$$

(with $\lambda = \sigma_M^2/T$) under the bilinear constraints:

$$\phi - \phi_0 - \mathbf{G}_O \mathbf{X} \phi = 0. \quad (13.12)$$

In general, from a probabilistic point of view, the solution in \mathbf{x} of the estimate in (\mathbf{x}, ϕ) in the joint MAP sense has no reason to correspond to the estimate of \mathbf{x} in the MAP sense. It is obvious that we arrive at the same solution here because the deterministic relation linking ϕ and \mathbf{x} , and the J^{MAPC} criterion is none other than the J^{MAP} criterion in which we have set $\phi = (\mathbf{I} - \mathbf{G}_0\mathbf{X})^{-1}\phi_0$, the relationship we took into account as a constraint linking \mathbf{x} and ϕ .

13.4.3. Properties of the criteria

We have defined a regularized solution to the problem of diffraction tomography, bringing us to the minimization of a criterion with or without constraints. It is clear that, before trying to implement optimization techniques, it is important to study the properties of these criteria.

As we saw in Chapter 3, in the case of a linear problem in \mathbf{x} with a Gaussian model of additive noise, the data fidelity term is quadratic in \mathbf{x} and thus convex; any non-convexity comes from the prior model of the unknowns. In our case, the non-convexity of the criteria presented and of the constraint comes from the nonlinearity of the model of the direct problem (recall that we chose a prior model giving a convex function Φ). We are thus not completely sure that these criteria are unimodal, which does not necessarily mean that they have local minima³. Apart from the situations where we can make use of a linear approximation of the direct problem, two types of situation can be envisaged: the first, relatively *favorable*, for which, although they are not convex, the criteria do not have local minima; the second, more *difficult*, for which local minima are present. The latter appear in particular when the contrast function takes high values and the number of data values is small (possibly even lower than the number of unknowns). Depending on the type of situation, we can see already that it is not possible to envisage implementing the same optimization techniques.

13.5. Solving the inverse problem

Since the late 1980s, many methods have been put forward for solving this nonlinear inverse problem. They are generally presented in a functional framework and it is not always very easy to compare them from a theoretical point of view. The Bayesian algebraic framework adopted in this chapter, i.e., the minimization of the criteria defined in section 13.4, allows us to take a unified view of most of the approaches used. The aim here is not to go into the details of the various inversion methods employed but, above all, to present their strengths and weaknesses, and points of comparison in terms of optimization and regularization. Three classes of methods can be distinguished, which we will study below.

3. Recall that convexity is only a sufficient condition for unimodality.

13.5.1. Successive linearizations

The first methods employed to solve this inverse problem used successive affine approximations of the model of the direct problem and thus successively solved *linear* inverse problems. Let us consider a two-step iterative process:

- 1) affine approximation of the model of the direct problem around a current value for the object \mathbf{x}_n : $\mathcal{A}(\mathbf{x}) = \mathbf{A}_n \mathbf{x} + \mathbf{b}_n$;
- 2) calculation of \mathbf{x}_{n+1} solution of the linear inverse problem corresponding to this approximation.

Note that this type of approach can be envisaged for solving any nonlinear inverse problem, even if no convergence property can be established in general.

These methods have generally been proposed on the basis of approximations of wave propagation equations and thus of the direct model in integral form. To compare them, it is necessary to study the affine approximation and the regularization performed.

13.5.1.1. Approximations

From a theoretical point of view, the most consistent of the affine approximations of a function \mathcal{A} in the neighborhood of a point \mathbf{x}_n is given by its first order Taylor expansion⁴ around \mathbf{x}_n :

$$\mathcal{A}(\mathbf{x}) = \mathcal{A}(\mathbf{x}_n) + \nabla_{\mathbf{x}} \mathcal{A}(\mathbf{x}_n)(\mathbf{x} - \mathbf{x}_n) + \mathcal{O}((\mathbf{x} - \mathbf{x}_n)^2).$$

In the algebraic framework, the calculation of $\mathbf{A}_n \triangleq \nabla_{\mathbf{x}} \mathcal{A}(\mathbf{x}_n)$ can be done simply. If $\phi_n = (\mathbf{I} - \mathbf{G}_0 \mathbf{X}_n)^{-1} \phi_0$ represents the field at the object corresponding to contrast \mathbf{x}_n and Ψ_n is the corresponding diagonal matrix, \mathbf{A}_n can be written: $\mathbf{A}_n = \mathbf{G}_M (\mathbf{I} + \mathbf{X}_n (\mathbf{I} - \mathbf{G}_0 \mathbf{X}_n)^{-1} \mathbf{G}_0) \Psi_n$.

Nevertheless, such an approximation has not always been used. The *Born iterative method* [WAN 89] (BIM), which aims to solve each of the coupled equations (13.3)-(13.4) successively for one of the variables (contrast or field at the object), considers an approximation via a matrix of form $\mathbf{A}_n = \mathbf{G}_M \Psi_n$ corresponding to a rougher linearization of the model (note that, in this case, $\mathbf{b}_n = 0$ since $\mathcal{A}(\mathbf{x}_n) = \mathbf{A}_n \mathbf{x}_n$).

The above type of approximation is, however, taken into account in the *distorted Born iterative method* (DBIM) [CHE 90] and the Newton-Kantorovitch method

4. Note, however, that such an expansion cannot be defined so simply for functions of complex variables, with complex values. In fact, the gradient operator is not defined for such functions. Nevertheless, this relation can be applied and calculated from an equivalent real notation (separating the real and imaginary parts) of the direct model equations.

(NKM) [JOA 91], which have proved to be the same although they were put forward using distinct approaches.

Note that the computing cost for approximation matrices is more or less equivalent to that of solving the direct problem.

13.5.1.2. *Regularization*

The regularization performed by the BIM method is of the Tikhonov-type, i.e., via quadratic penalization on \mathbf{x} . This method can thus be directly interpreted in terms of minimizing J^{MAP} .

The NKM and DBIM methods introduce a penalization that is still quadratic but on $\mathbf{x} - \mathbf{x}_n$ and not directly on \mathbf{x} .

13.5.1.3. *Interpretation*

In the framework proposed in section 13.4, a successive linearization scheme can be envisaged for minimizing the J^{MAP} criterion [CAR 97a]:

- 1) calculation of matrix \mathbf{A}_n and vector \mathbf{b}_n corresponding to the affine approximation of the direct problem around a current value for object \mathbf{x}_n ;
- 2) minimization of the criterion $J_n = \|\mathbf{y} - \mathbf{A}_n \mathbf{x} + \mathbf{b}_n\|^2 + \lambda \Phi(\mathbf{x})$.

At each iteration of this algorithm, criterion J^{MAP} is approximated by a convex criterion J_n having the same value at \mathbf{x}_n and the same slope at this point. Of course, there is no theoretical guarantee of convergence for this algorithm and it may diverge. Similarly, any convergence that may exist and the stationary point reached depend on how the algorithm is initialized. However, when the algorithm does converge towards a stationary point \mathbf{x}_∞ , this point is a stationary point of the J^{MAP} criterion (i.e., its gradient is zero).

Note that the algorithm corresponding to BIM also consists of approximating the J^{MAP} criterion (for $\Phi(\mathbf{x}) = \|\mathbf{x}\|^2$) by a series of convex criteria having the same value at \mathbf{x}_n but not having the same slope at \mathbf{x}_n . Thus, even when this algorithm converges, the solution does not necessarily correspond to a stationary point of J^{MAP} . The regularization performed on $\mathbf{x} - \mathbf{x}_n$ for DBIM and NKM does not allow them to be interpreted in this framework but it has been observed [CHE 90] that DBIM diverges more easily than BIM, apparently because of this regularization and, for NKM, a scheme to adapt the regularization parameters had to be introduced in [JOA 91] to avoid this drawback.

To conclude on these methods, it is worth noting that all the regularization methods for linear inverse problems can be used to solve nonlinear problems by successive linearizations (in particular the well known truncated singular value decomposition

(TSVD) as proposed in [ERI 96]). However, such methods only try to stabilize each linear inverse problem. In contrast, the successive linearization scheme to minimize the criterion defined in section 13.4 allows the nonlinear problem to be regularized by introducing prior information on the object under study.

13.5.2. Joint minimization

A second approach used for solving this nonlinear problem aims at the simultaneous calculation of \mathbf{x} and ϕ that satisfy the coupled equations (13.3)-(13.4). For this, a joint criterion on \mathbf{x} and ϕ is defined, having the form:

$$K(\mathbf{x}, \phi) = \alpha_M \|\mathbf{y} - \mathbf{G}_M \mathbf{X} \phi\|^2 + \alpha_O \|\phi - \phi_0 - \mathbf{G}_O \mathbf{X} \phi\|^2 + \lambda \Phi(\mathbf{x}, \phi). \quad (13.13)$$

Note that, here, we are in an algebraic framework whereas the methods based on such a criterion have usually been introduced in a functional framework. Such criteria appeared for the diffraction tomography problem in [KLE 92], in parallel with [SAB 93], and have been widely used since then.

Certain differences have become apparent between the proposed methods, which we will look at more closely now:

- various values have been proposed for parameters α_M and α_O present in the criterion (e.g. in [KLE 92] and [SAB 93]), without the values being justified theoretically;
- differences also appear in the regularization term sometimes introduced into the criterion. At the beginning, no regularization was taken into consideration [KLE 92, SAB 93]. Then a joint penalization on \mathbf{x} and ϕ was proposed (e.g. in [BAR 94]). Finally, a regularization on \mathbf{x} only was considered, with an energy Φ corresponding to Gibbs-Markov models (e.g. in [CAO 95] and [BER 95]);
- finally, various optimization techniques have been used in order to calculate this solution: gradient-type local optimization (e.g. in [KLE 92, SAB 93, BAR 94]); or global stochastic optimization based on simulated annealing [CAO 95] (for a non-convex function Φ).

The link between criteria K (13.13) and J^{MAPJ} (13.10), presented in section 13.4.2 as an artifice of calculation, is obvious (the only difference coming from the presence of ϕ in function Φ). Furthermore, this is the approach used in [CAO 95], considering the errors on each of the coupled equations, without justification of these hypotheses. If regularization only concerns \mathbf{x} , criterion (13.13) corresponds to the J^{MAPC} criterion penalized by the error of the norm on constraint (13.12). The result of such minimization is not guaranteed to verify constraint (13.12) and the solution thus does not correspond to that defined by the joint approach in section 13.4.2. To this end, an algorithm for optimization under constraints, based on the augmented Lagrangian has been proposed [CAR 96]. Note that, as the convexity conditions of the criterion and of the constraint are not fulfilled, such an algorithm has no guarantee of convergence.

Finally, we observe that the calculation of the criterion, and thus its minimization, do not require the direct problem to be solved. The criterion computing cost is thus reasonable compared with the solution of the direct problem. The price to be paid to be able to have this reduced computing cost is the multiplication of the number of unknowns since \mathbf{x} and ϕ are sought simultaneously (recall that, in a multi-source case, ϕ is the field in the object domain for each source, and the number of unknowns is thus multiplied by the number of sources plus one).

13.5.3. Minimizing MAP criterion

A third approach to the problem is to directly minimize the least squares criterion $\|\mathbf{y} - \mathcal{A}(\mathbf{x})\|^2$, regularized (i.e., the J^{MAP} criterion defined in section 13.4.1) or not. Note that, for such an approach, it is not necessary to define operator \mathcal{A} explicitly; it is sufficient to have an algorithm that allows the direct problem to be solved. However, knowledge of the direct problem model and the highlighting of some of its particular structures can enable specific optimization algorithms to be implemented.

The importance of regularization has been amply stated in this book. Nevertheless, it should be pointed out that many works define the solution as minimizing the non-regularized criterion. For example, in [GAR 91], a simulated annealing global optimization algorithm is used. As the simulations are performed without noise on the measurements (some call this the *inverse crime*), this algorithm has given good results, even without regularization. In [HAR 95], a conjugate gradient algorithm is proposed in a functional framework and reconstructions from noisy measurements are presented. Regularization is not explicitly introduced in the criterion but the algorithm is stopped before the solution ceases to be satisfactory.

As we saw in section 13.4.3, the nonlinearity of the direct model makes the J^{MAP} criterion non-convex and local minima may be present. Thus, prudence is necessary before implementing an algorithm to minimize this criterion.

In the most *favorable* situations, for which the least squares criterion is unimodal – the same is true for the J^{MAP} criterion when function Φ is itself convex – the methods of minimization by successive linearizations of section 13.5.1 and by optimization under constraints of section 13.5.2 generally enable the solution to be calculated. In addition, such methods have a considerably lower computation cost than the conventional gradient optimization techniques, which require the direct problem to be solved for each evaluation of the J^{MAP} criterion. Thus, it does not seem useful to try to minimize the J^{MAP} criterion directly in these situations, except if a prior model with a non-convex function Φ is taken into account [LOB 97].

In more *difficult* situations for which the criterion has local minima, the use of global optimization techniques to minimize the J^{MAP} criterion have to be envisaged.

Indeed, local optimization techniques – in particular those presented in sections 13.5.1 and 13.5.2 when they converge – may get stuck in a local minimum. This problem is raised in [GAR 91], where the amount of available data is small and the interest of a global optimization technique, simulated annealing, is pointed out. In [CAR 95] a deterministic global optimization algorithm, based on gradual non-convexity (GNC), is proposed and, in [CAR 97b], an algorithm for coordinate-wise global optimization (*Iterative Coordinate Descent* ICD) is put forward. Both these algorithms take advantage of particular structures of direct model $\mathcal{A}(\mathbf{x})$ to try to carry out the optimization globally. Nevertheless, theoretical convergence towards the global minimum is not guaranteed.

As an illustration, we show the results of simulations in a difficult configuration (see [CAR 96] for more details on these simulations). We have at our disposal measurements taken at 8 sensors for waves emitted from 8 distinct positions. The measurements contain noise (signal-to-noise ratio 20 dB) and the object has been discretized with a grid of 11×11 pixels (121 unknowns for 64 data points). The maximum contrast is 5.5. In Figure 13.3, we have represented the permittivity of the original object,

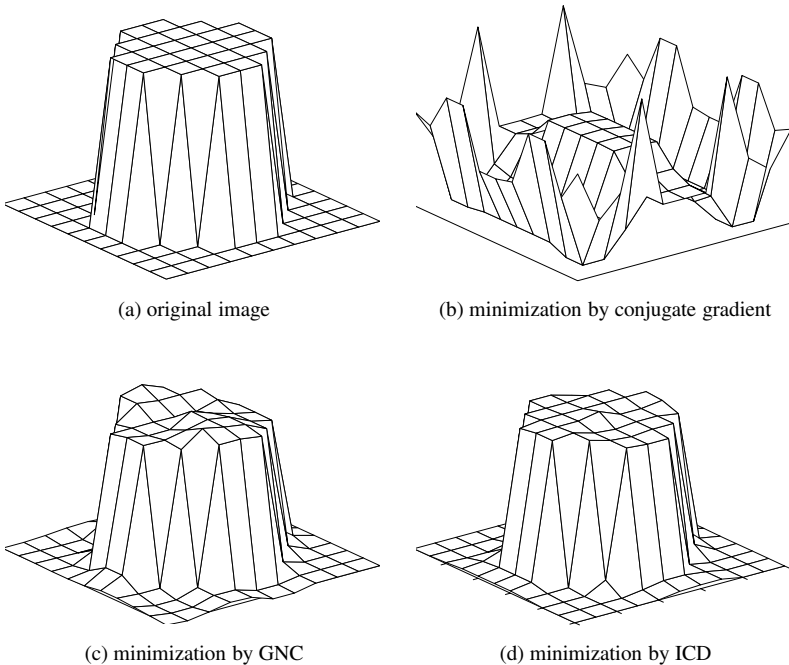


Figure 13.3. Example of reconstruction by minimization of J^{MAP} in a difficult configuration

its reconstruction by minimization of the J^{MAP} criterion by conjugate gradient and by GNC for a potential function $|\cdot|^{1,1}$, then by ICD for a potential function $|\cdot|$. On this example, we can clearly see the presence of a local minimum in which the conjugate gradient algorithm gets stuck. Note that a Tikhonov regularization (Gaussian model: $\Phi(\mathbf{x}) = \|\mathbf{x}\|^2$) would give a far less satisfactory solution than the Markov models considered here.

13.6. Conclusion

This study of a specific nonlinear inverse problem emphasizes some important points for solving other nonlinear problems. In the Bayesian framework, a regularized solution to the problem can be defined as an estimate in the maximum *a posteriori* sense, by explicitly introducing prior information on the object to be reconstructed. Such a solution can be defined as soon as the direct model presents an explicit relationship between the unknowns and the data. If Gaussian perturbations on these data are considered, the solution is a regularized version of the least squares solution. However, it can be interesting to take advantage of formulations of the direct model other than this explicit relationship when defining the solution. These different formulations lead us to solve a non-convex optimization problem, with or without constraints.

As for linear inverse problems, the prior information is introduced in the form of probabilistic models or, equivalently, as a penalization term on the least squares criterion. A large number of image models have been used for linear inverse problems in the past 15 years or so and have proved to be much more efficient than energy penalization as used by Tikhonov.

Calculation of the solution requires optimization algorithms to be implemented, which themselves require the direct problem to be solved, in its explicit form or in another form. Discretizing this model in a simple form that can be calculated at low cost is an important step in solving the problem.

It has been possible to study most of the existing solving methods in this framework, in terms of optimization techniques. Various techniques need to be used according to the difficulty of the problem. For this problem, strong contrasts in the object and a small amount of data entail the presence of local minima in the criterion. In favorable situations, successive linearization of the direct model can be used for local optimization of the criterion. Similarly, the specific form of the model – coupled equations in our case – can be used to calculate the solution by performing a local optimization under constraints. Nevertheless, in situations where local minima are present, global optimization techniques, with much higher computing costs, have to be used. Obviously, the latter can take advantage of the structure of the discretized direct model.

13.7. Bibliography

- [BAR 94] BARKESHLI S., LAUTZENHEISER R. G., “An iterative method for inverse scattering problems based on an exact gradient search”, *Radio Sci.*, vol. 29, num. 4, p. 1119-1130, July-Aug. 1994.
- [BER 95] VAN DEN BERG P. M., KLEINMAN R. E., “A total variation enhanced modified gradient algorithm for profile reconstruction”, *Inverse Problems*, vol. 11, p. L5-L10, 1995.
- [BOL 83] BOLOMEY J. C., PERONNET G., PICHOT C., JOFRE L., GAUTHERIE M., GUERQUIN KERN J. L., “L’imagerie micro-onde active en génie biomédical”, in *L’imagerie du corps humain*, p. 53-76, Les Éditions de physique, Paris, 1983.
- [BOU 93] BOUMAN C. A., SAUER K. D., “A generalized Gaussian image model for edge-preserving MAP estimation”, *IEEE Trans. Image Processing*, vol. 2, num. 3, p. 296-310, July 1993.
- [CAO 95] CAORSI S., GRAGNANI G. L., MEDICINA S., PASTORINO M., PINTO A., “A Gibbs random fields-based active electromagnetic method for noninvasive diagnostics in biomedical applications”, *Radio Sci.*, vol. 30, num. 1, p. 291-301, Jan.-Feb. 1995.
- [CAR 95] CARFANTAN H., MOHAMMAD-DJAFARI A., “A Bayesian Approach for Nonlinear Inverse Scattering Tomographic Imaging”, in *Proc. IEEE ICASSP*, vol. IV, Detroit, MI, p. 2311-2314, May 1995.
- [CAR 96] CARFANTAN H., Approche bayésienne pour un problème inverse non linéaire en imagerie à ondes diffractées, PhD thesis, University of Paris XI, France, Dec. 1996.
- [CAR 97a] CARFANTAN H., MOHAMMAD-DJAFARI A., “An overview of nonlinear diffraction tomography within the Bayesian estimation framework”, in *Inverse Problems of Wave Propagation and Diffraction*, p. 107-124, Lecture Notes in Physics, Springer Verlag, New York, NY, 1997.
- [CAR 97b] CARFANTAN H., MOHAMMAD-DJAFARI A., IDIER J., “A single site update algorithm for nonlinear diffraction tomography”, in *Proc. IEEE ICASSP*, Munich, Germany, p. 2837-2840, Apr. 1997.
- [CHE 90] CHEW W. C., WANG Y. M., “Reconstruction of two-dimensional permittivity distribution using the distorted Born iterative method”, *IEEE Trans. Medical Imaging*, vol. 9, p. 218-225, June 1990.
- [COL 92] COLTON D., KRESS R., *Inverse Acoustic and Electromagnetic Scattering Theory*, Springer Verlag, New York, NY, 1992.
- [ERI 96] ERIKSSON J., Optimization and regularization of nonlinear least squares problems, PhD thesis, Umeå University, Sweden, June 1996.
- [GAR 91] GARNERO L., FRANCHOIS A., HUGONIN J.-P., PICHOT C., JOACHIMOWICZ N., “Microwave imaging – Complex permittivity reconstruction by simulated annealing”, *IEEE Trans. Microwave Theory Tech.*, vol. 39, num. 11, p. 1801-1807, Nov. 1991.
- [HAR 87] HARRINGTON R. F., “The method of moments in electromagnetics”, *J. Electromagnetic Waves Appl.*, vol. 1, num. 3, p. 181-200, 1987.

- [HAR 95] HARADA H., WALL D. J. N., TAKENAKA T., TANAKA M., "Conjugate gradient method applied to inverse scattering problem", *IEEE Trans. Ant. Propag.*, vol. 43, num. 8, p. 784-791, 1995.
- [HOW 86] HOWARD A. Q. J., KRETZSCHMAR J. L., "Synthesis of EM geophysical tomographic data", *Proc. IEEE*, vol. 74, num. 2, p. 353-360, Feb. 1986.
- [JOA 91] JOACHIMOWICZ N., PICHOT C., HUGONIN J.-P., "Inverse scattering: An iterative numerical method for electromagnetic imaging", *IEEE Trans. Ant. Propag.*, vol. AP-39, num. 12, p. 1742-1752, Dec. 1991.
- [KAK 88] KAK A. C., SLANEY M., *Principles of Computerized Tomographic Imaging*, IEEE Press, New York, NY, 1988.
- [KLE 92] KLEINMAN R. E., VAN DEN BERG P. M., "A modified gradient method for two-dimensional problems in tomography", *J. Comput. Appl. Math.*, vol. 42, p. 17-35, 1992.
- [LOB 97] LOBEL P., BLANC-FERAUD L., PICHOT C., BARLAUD M., "A new regularization scheme for inverse scattering", *Inverse Problems*, vol. 13, num. 2, p. 403-410, Apr. 1997.
- [SAB 93] SABBAGH H. A., LAUTZENHEISER R. G., "Inverse problems in electromagnetic nondestructive evaluation", *Int. J. Appl. Electromag. Mat.*, vol. 3, p. 235-261, 1993.
- [TAR 82] TARANTOLA A., VALETTE B., "Inverse problems = quest for information", *J. Geophys.*, vol. 50, p. 159-170, 1982.
- [TAR 87] TARANTOLA A., *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*, Elsevier Science Publishers, Amsterdam, The Netherlands, 1987.
- [WAN 89] WANG Y. M., CHEW W. C., "An iterative solution of the two-dimensional electromagnetic inverse scattering problem", *Int. J. Imag. Syst. Tech.*, vol. 1, p. 100-108, 1989.
- [ZOR 91] ZORGATI R., DUCHENE B., LESSELIER D., PONS F., "Eddy current testing of anomalies in conductive materials, part I: Qualitative imaging via diffraction tomography techniques", *IEEE Trans. Magnetics*, vol. 27, num. 6, p. 4416-4437, 1991.

Chapter 14

Imaging from Low-intensity Data

14.1. Introduction

Estimation of digital imagery giving rise to photon-limited data forms an important subset of inverse problems. Its applications are numerous in medical diagnostic imaging, astronomy and industrial inspection. Due to the nature of the data collection process, this class of inverse problems presents a number of challenges specific to low signal-to-noise ratios (SNRs) and Poisson likelihood functions. These problems are discussed in this chapter within a Bayesian estimation framework, which enables numerous modeling and optimization options and leads to improved image estimates.

At the most fundamental physical level, many measurement devices register the superposition of discrete physical events. Emissions from light-emitting diodes, currents from charge-coupled devices (CCDs), or even measurement of simple electrical currents have discrete phenomena as their basis. The data available as input to most inverse problems represents the superposition of a sufficient number of events to regard each datum, after appropriate normalization, as continuously distributed. In some cases as well, underlying Poisson characteristics of a measurement process may be of low enough variance to make secondary system noise effects the dominant noise mechanism. Many image recovery tasks, for example, deal with sufficient quantities of light at each detector that corruption due to shot noise and other disturbances can be quite well approximated as contributing to the variance of Gaussian random variables. This typically holds as well for problems with limited numbers of measurements, as in Chapter 12, where each attenuation measurement may involve enough X-ray photons

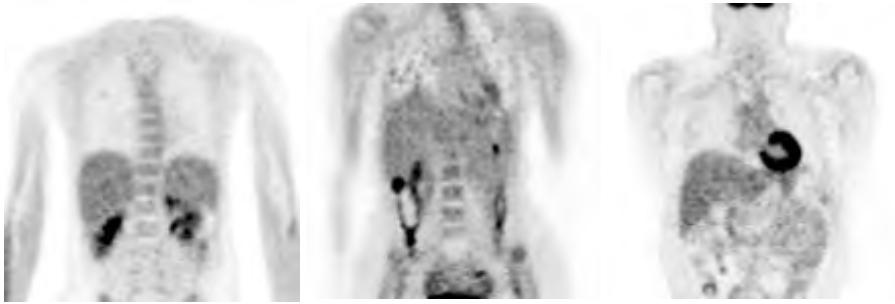


Figure 14.1. *Reconstructions from clinical scans on General Electric's Advance positron emission tomographic diagnostic imaging system (data courtesy of General Electric Imaging Systems, Inc.)*

that the standard Gaussian asymptotic approximation for the distribution of maximum likelihood (ML) attenuation estimates [BIC 77] is sufficiently accurate.

A growing number of estimation settings feature photon-limited measurements, in which the Poisson nature of the distribution is of importance, as we will see below. This situation may arise when newly developed, highly sensitive instruments are available to measure very weak signals, as in image recovery from faint signals emitted far from the earth and received by optical telescopes [SNY 90]. Scintigraphic images, highly useful for medical diagnostics or industrial inspection, are formed from very low intensity planar measurements [IIN 67, NGU 99]. Emission tomographic imaging relies on measurements often consisting of a handful of photons due to patient safety requirements, for both emission rate estimates and attenuation maps derived from low-intensity transmission scans [OLL 97]. All these applications require estimation techniques tailored to the nature of the data in order to produce usable imagery. Figure 14.1 shows the level of image quality achievable with current clinical positron emission tomography (PET) scanners. The quality of these images is significantly lower than is possible with X-ray CT scans, but the images are also of different clinical significance: X-ray offers excellent detail in anatomical diagnostics, whereas PET provides physiological information.

In any of the above settings, it may be worthwhile or even necessary to depart from standard quadratic measures of estimate affinity to data and consequent linear techniques for estimation. Poisson-distributed data has variance equal to its mean, which makes measurement noise properties dependent on the unknown signal. The signal is typically physically constrained to be non-negative, a constraint which can aid estimates but complicate the optimization necessary for calculation of estimators. However, application of models true to the quantum nature of data can yield improvement in both the quality of recovered signals and the inference of system parameters.

In subsequent sections we will explore key elements of commonly occurring weak-signal cases in modern inverse problems. The unknown parameters to be estimated, be they elements of the object of an inverse problem or “hyperparameters” underlying the models, are assumed continuously distributed. After considering important statistical modeling issues, we examine various algorithms designed to efficiently calculate the desired estimates.

14.2. Statistical properties of common low-intensity image data

A number of probability distributions may be of interest for low intensity measurements. Those most unique to this framework involve data which takes the form of counting discrete events and is typically modeled as Poisson-distributed. The following discussion, however, could be applied with small modifications to others, such as binomial.

14.2.1. Likelihood functions and limiting behavior

We shall use the vector \mathbf{X} to represent the unknown object of the inverse problem at hand, and \mathbf{Y} as the random observations. Scalar entries in any of the vectors will be subscripted, as for example X_j for a single image pixel and Y_i for one measurement datum. Should \mathbf{X} not be viewed as random, of course, no distribution will be necessary for it. As in many other problems, the data in weak signal cases are often most accurately modeled as the superposition of the signal and multiple noise sources. Given appropriate independence characteristics, this superposition is frequently modeled as Gaussian through application of the Central Limit Theorem. The primary distinguishing feature here of data in \mathbf{Y} is that the information-bearing component of \mathbf{Y} will be restricted to the non-negative integer values due to small numbers of discrete events in the signal. We will assume \mathbf{Y} to be the sum of perhaps several independent phenomena, with one component purely discretely valued and the other continuous, or

$$\mathbf{Y} = \mathbf{N} + \mathbf{B},$$

where each element in \mathbf{B} will be Gaussian for our discussion, with known mean γ_i and variance σ_i^2 independent of \mathbf{X} . Each N_i will consist of two Poisson components: one has a mean μ_i which is dependent on \mathbf{X} and another has a mean β_i which is independent of \mathbf{X} . The distribution of N_i , conditioned on \mathbf{X} , has the form

$$\Pr(N_i = n_i \mid \mathbf{X} = \mathbf{x}) = \frac{1}{n_i!} \exp \{ -(\mu_i(\mathbf{x}) + \beta_i) \} (\mu_i(\mathbf{x}) + \beta_i)^{n_i} \quad (14.1)$$

and the likelihood function $\Pr(\mathbf{Y} = \mathbf{y} \mid \mathbf{X} = \mathbf{x})$ will be formed by the convolution of the Gaussian probability density of B_i with equation (14.1). The mean of Y_i is

$$\mu_{Y_i}(x) = \mu_i(x) + \beta_i + \gamma_i.$$

Most problems in the conventional Bayesian framework use a linear description of the relationship between \mathbf{x} and $\boldsymbol{\mu}$, which we will express as

$$\boldsymbol{\mu}(\mathbf{x}) = \mathbf{A}\mathbf{x}, \quad (14.2)$$

with matrix \mathbf{A} as a discrete description of, for example, the blurring by camera motion or system optics of image pixel values, or the line integrals in emission tomographic measurements. \mathbf{A} may also be adjusted to include calibrated detector efficiencies and internal attenuation of measured photons. Transmission tomographic problems have a slightly more complicated function as $\boldsymbol{\mu}$, but it is typically modeled with \mathbf{x} entering the distribution in form $\mathbf{A}\mathbf{x}$. This makes the characteristics of the two basic linear tomography systems sufficiently similar for us to discuss primarily the simpler emission case and point to the transmission case only where it is of particular interest.

Emission tomographic data is often corrected before image reconstruction for the effects of attenuation by the patient [MEI 93, OGA 91], scatter [OLL 93, PAN 97] and accidental coincidence accidental coincidences [POL 91]. Such corrections are critical in advance of using deterministic methods such as filtered backprojection. In contrast, statistical methods enable inclusion of models for these degradations in the observation model. Attenuation reduces the probability of detection of each emitted photon; this translates mathematically into a multiplicative factor for each element in \mathbf{A} , similarly to a loss in detector efficiency. Rather than subtracting accidental coincidences and scatter from raw data, we may model them as contributing to random vector \mathbf{N} with expected values in $\boldsymbol{\beta}$. In the following, we assume that these effects are included in \mathbf{A} and $\boldsymbol{\beta}$. A more precise approach would model scatter by a smoothing among the columns of \mathbf{A} , but the spatial extent of these interactions make this computationally costly.

14.2.2. Purely Poisson measurements

The degree of importance of the Poisson nature of \mathbf{N} in formulating and solving the low-intensity estimation problem depends on the relative variances of the components and on the absolute intensity of the signal component expressed in $\mu_i(\mathbf{x})$. Let us first consider the simplest case of the Poisson problem, where $\sigma_i^2, \gamma_i, \beta_i \ll \mu_i(\mathbf{x})$ and the distribution of Y_i is approximately that of N_i . In this case the measure for Y_i conditioned on $\{\mathbf{X} = \mathbf{x}\}$ is a probability mass function $\Pr(Y_i = y_i | \mathbf{X} = \mathbf{x})$, illustrated for a scalar case of both variables in Figure 14.2. Poisson-distributed data possesses a fundamentally important second characteristic in having a single parameter for both their means and variances. Assuming \mathbf{X} controls the rate of photon or particle emission, $\Pr(Y_i = y_i | \mathbf{X} = \mathbf{x})$ will be independent of any other parameters. This may be an advantage in modeling simplicity, but introduces a signal-dependent noise variance which has to some extent impeded the application of least-squares estimation techniques to such problems. The Central Limit Theorem can be used to argue that

as the expected rate of events rises, the distribution of a Poisson variable normalized by the square root of its mean converges to a Gaussian. However, though the discrete form of the distribution may be of decreasing importance for large numbers of events, the dependence of both mean and variance on \mathbf{X} remains. As mean and variance are equal, the SNR increases linearly with variance, opposite the Gaussian.

For estimation or reconstruction of an estimate $\widehat{\mathbf{X}}$ from such data, we are interested in the characteristics of the likelihood functions in terms of the variable representing underlying intensity, reflectivity, or whatever parameter in \mathbf{X} determines the rate of counts at detectors. Therefore, the estimation problem differs from those whose principle corruption is not at the photon counting level only in the shape of the likelihood function as indexed by the continuously-varying intensity parameter. The distribution of variates Y_i approaches a Gaussian for large means, but the Poisson differs significantly for single-digit means. In Figure 14.2 we plot Poisson densities and the corresponding log-likelihood functions with the observed variables taking their mean

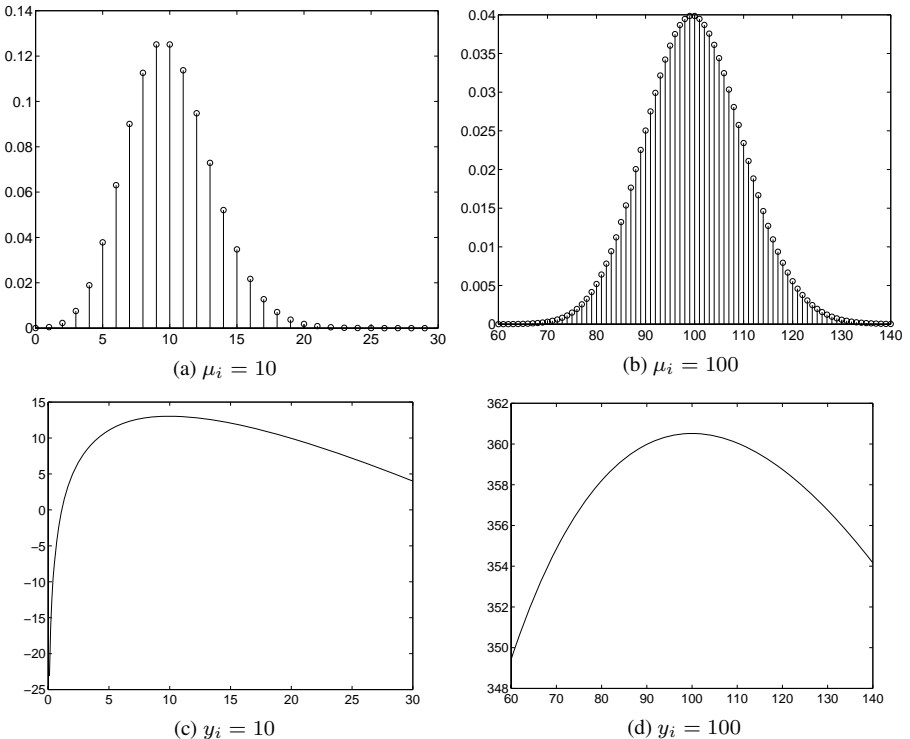


Figure 14.2. Above: probability mass function of Poisson-distributed variable with parameter μ_i ; below, log-likelihood function of the Poisson parameter for observation $Y_i = y_i$

values. In other words, in the upper plots we fix μ_i and consider the distribution of Y_i while on the right we fix the outcome $Y_i = y_i$ and plot the likelihood in $\mu_i(\mathbf{x})$.

Just as asymptotic approximations are often made to describe the distribution of Y_i , the likelihood function in \mathbf{X} may profitably be approximated by simpler functions. As a sort of counterpart to the notion of the similarity of the distribution of high-mean Poisson counts to Gaussian (see Figure 14.2), the log-likelihood in $\mu_i(\mathbf{x})$ is locally well-approximated by an appropriate polynomial. This may be used for both design of simplified estimators and analysis of their behavior, as discussed below.

14.2.3. *Inclusion of background counting noise*

Adding a β_i to the discussion above is relatively simple, as it adds only a shift in $\mu_i(\mathbf{x})$ relative to the plots of Figure 14.2. Such an additive component of the Poisson mean is in practice often due to background noise, such as accidental coincidences in PET, or scattered photons with little or no dependence on local characteristics of \mathbf{X} . In emission tomography, provided equation (14.2) is valid, β_i may be treated as though it were a member of an augmented \mathbf{x} if its estimation is desired. From equation (14.1), it is simple to show that the log-likelihood is also a concave function of the unknown parameters.

In transmission tomography, the addition of the background counts may destroy the concavity of the log-likelihood. This is an important consideration in optimization, as concave functions do not have multiple local minima. Here, the Poisson parameter is, for integral density $l_i(\mathbf{x})$ and input dosage d_i , $d_i \exp(-l_i(\mathbf{x})) + \beta_i$. For $\beta_i = 0$, the log-likelihood is strictly convex, but as demonstrated in the plots of Figure 14.3, it loses this property in the presence of relatively weak background counts. The gravity of the non-convexity in optimization may vary with initial conditions and consistency among separate measurements. Should the background count rate approach the observed count y_i , as in the dashed line, low curvature of the log-likelihood at its maximum indicates potentially high variance in any estimator and serious consequences of the non-convexity phenomenon in estimator performance despite its lesser severity. With $\beta_i \gg y_i$, the concavity is recovered; however, this case is of minimal practical importance.

14.2.4. *Compound noise models with Poisson information*

While the purely Poisson model is considered reasonable for the output of solid-state detectors in most nuclear medicine imaging systems, data from arrays of CCDs in common optical sensors are corrupted by additional noise inherent in reading out the charge at each pixel. This read-out noise is characterized as Gaussian, independent of the underlying image data \mathbf{x} [SNY 93]. These Gaussian parameters (γ_i, σ_i) are generally available via calibration measurements.

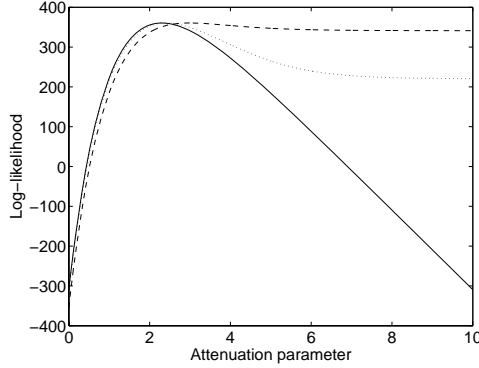


Figure 14.3. Log-likelihoods for integral attenuation parameter (l_i) in transmission tomography with background Poisson noise observed at count y . Total Poisson parameter = $d_i \exp(-l_i) + \beta_i$; $d_i = 1,000$ for all plots. (—) $y_i = 100$, $\beta_i = 0$; (···) $y_i = 100$, $\beta_i = 10$; (---) $y_i = 100$, $\beta_i = 50$

The resulting compound Poisson-Gaussian model for Y_i has a density resulting from the convolution of the two, or

$$p(y_i | \mathbf{X} = \mathbf{x}) = \sum_{n=0}^{\infty} \frac{1}{n!} \exp\{-\mu_i(\mathbf{x}) + \beta_i\} (\mu_i(\mathbf{x}) + \beta_i)^n \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(y_i - n - \gamma_i)^2}{2\sigma_i^2}\right\}. \quad (14.3)$$

The Gaussian component will normally be assumed to have variance significantly greater than one, yielding a density and log-likelihood as illustrated in Figure 14.4. Such compound noise will have distributions depending on the components' parameters, but as is clear from the plots, a Gaussian approximation is quite close in the vicinity of the maximum.

14.3. Quantum-limited measurements in inverse problems

14.3.1. Maximum likelihood properties

The ML estimate of scalar means from the single quantum-limited observations above is easily seen as $\hat{\mu}_i = y_i$. However, $\mu_i(\mathbf{x})$ for each measurement is typically parameterized by a non-trivial function of vector \mathbf{x} , and \mathbf{Y} is of high dimension, which classifies the estimation as a nontrivial inverse problem. The dependence is often modeled as linear, with the means $\boldsymbol{\mu}(\mathbf{x})$ satisfying (14.2). \mathbf{A} may represent such factors as local blurring due to limited instrument resolution, integrals along typical

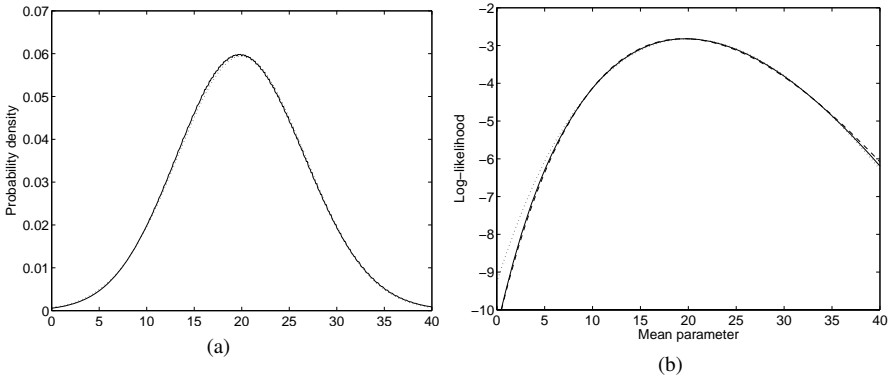


Figure 14.4. Compound noise model $p(\mathbf{y} | \mathbf{X} = \mathbf{x})$ with observations equal to the sum of independent Poissons and Gaussians. (a) (—) Probability density for $\mu_i(\mathbf{x}) = 20$, $\beta_i = \gamma_i = 0$, $\sigma_i = 5$; (···) Gaussian pdf of mean 20 and variance 45; (b) (—) log-likelihood as function of $\mu_i(\mathbf{x})$ with $y_i = 20$, $\beta_i = \gamma_i = 0$, $\sigma_i = 5$; (···) Poisson log-likelihood as a function of $\mu_i(\mathbf{x})$ with mean of $\mu_i(\mathbf{x}) + 25$ and $y_i = 45$; (---) Gaussian log-likelihood as function of $\mu_i(\mathbf{x})$ with mean of $\mu_i(\mathbf{x})$, variance $\mu_i(\mathbf{x}) + 25$, and $y_i = 20$

paths of photons in tomography or any expected fractions of detections of photons from various points of origin. Quantum-limited estimation problems are therefore more complicated than the simple scalar plots of Figures 14.2-14.4 might indicate. Fortunately for the problems' tractability, the individual data can usually be accurately modeled as independent conditioned on \mathbf{X} , making the total log-likelihood the sum of those of scalar measurements. A standard formulation of an emission tomographic data distribution, for example, is

$$\Pr(\mathbf{Y} = \mathbf{y} | \mathbf{X} = \mathbf{x}) = \prod_{i=1}^M \frac{1}{y_i!} \exp \{ -(\mathbf{a}_{i*}^T \mathbf{x} + \beta_i) \} (\mathbf{a}_{i*}^T \mathbf{x} + \beta_i)^{y_i}$$

with $\mathbf{a}_{i*} \mathbf{x}$ the product of the i th row of \mathbf{A} with \mathbf{x} .

The ML estimator

$$\widehat{\mathbf{X}}^{\text{ML}} = \arg \max_{\mathbf{x}} \log \Pr(\mathbf{Y} = \mathbf{y} | \mathbf{X} = \mathbf{x})$$

may be roughly approximated by solution of the equation

$$\mathbf{E}(\mathbf{Y} | \mathbf{X} = \mathbf{x}) \simeq \mathbf{y}$$

or, in the linear case

$$\mathbf{A}\mathbf{x} + \boldsymbol{\beta} + \boldsymbol{\gamma} \simeq \mathbf{y}.$$

Thus, the low-intensity inverse problems share the principle features of more conventional ones. A good number of Poisson data problems can, in fact, be solved accurately with the actual log-likelihood replaced by

$$\log \Pr(\mathbf{Y} = \mathbf{y} \mid \mathbf{X} = \mathbf{x}) \simeq -\frac{1}{2}(\mathbf{y} - \mu_{\mathbf{Y}}(\mathbf{x}))^T \mathbf{D}(\mathbf{y} - \mu_{\mathbf{Y}}(\mathbf{x})) + \text{const.}, \quad (14.4)$$

with \mathbf{D} diagonal and a simple function of the data [BOU 96]. Such weighted least-squares formulations are more widely understood than the exact ML and may provide acceptable estimates in many applications [TSU 91, KOU 96]. Higher degree polynomials may extend the applicability of such approximations [FES 95].

As explained in section 14.2.2, a principal difference in the low-intensity case is that lower SNR mean lower curvature in the log-likelihood functions above. This results in greater inconsistencies in measurements due to noise, and values of $\mu_i(\widehat{\mathbf{X}}^{\text{ML}})$ which are distant from the maxima of the scalar log-likelihoods. Thus, while the polynomial approximations to these functions may be quite effective in high-count data, estimator accuracy may demand the precise likelihood functions in low SNR [THI 00]. Our discussions thus center on solving the problems posed by exact discrete-data likelihoods.

Performance bounds such as Cramèr-Rao may be used to analyze ML performance, similarly to canonical inverse problems in estimation. These measures depend on the characteristics of the expectation of the log-likelihood Hessian, with the approximate form $\mathbf{A}^T \mathbf{D} \mathbf{A}$ for common Poisson problems [BOU 96]. The diagonal matrix \mathbf{D} weights according to values in \mathbf{y} , with values inversely proportional to data for $\mu(\mathbf{x}) = \mathbf{A}\mathbf{x}$. While this suggests poorer performance for higher counts, we recall that for the Poisson in general, rising data variance is accompanied by increasing means and increasing SNR and estimator performance improves in higher counts.

The character of the Hessian varies among applications, but the primary feature of interest in inverse problems is that eigenvalues corresponding to high frequency information are small, indicating its suppression in the forward problem. This translates into very high variances in these components of $\widehat{\mathbf{X}}^{\text{ML}}$. The ML estimator typically has unacceptably poor performance in low intensity data [SNY 85]. A popular remedy to this problem is iterative approximation of the ML solution, commencing with a smooth image and stopping iterations of algorithms such as expectation-maximization (EM) long before convergence [LLA 89]. Alternatively, the ML problem is attacked by block-iterative techniques such as ordered subsets EM (OS-EM) [HUD 94]. OS-EM achieves rapid initial convergence toward the ML estimate but requires modification to guarantee complete convergence.

14.3.2. Bayesian estimation

The inadequacy of least-squares and ML methods in low-intensity measurements is well known and has given rise to alternative formulations in a variety of low-intensity application areas. Bayesian methods, in particular, have found frequent application and success in these problems. Among often-cited early developments in Bayesian image analysis were applications to emission tomography [GEM 85, SNY 85, LEV 87]. Our particular set of reconstruction problems differs little in fundamentals from many others in this volume. The maximum *a posteriori* probability (MAP) estimator has the usual form

$$\widehat{\mathbf{X}}^{\text{MAP}} = \arg \max_{\mathbf{x}} (\log \Pr(\mathbf{Y} = \mathbf{y} \mid \mathbf{X} = \mathbf{x}) + \log p(\mathbf{x})). \quad (14.5)$$

The attributes of the MAP estimate depends on the balance between the two terms in equation (14.5). For a fixed *a priori* density $p(\mathbf{x})$, low SNR results in relatively light weighting of the log-likelihood term in (14.5). This makes Bayesian MAP estimation in this problem potentially sensitive to the choice of the *a priori* model expressed in density $p(\mathbf{x})$. The log *a priori* may, in fact, dominate the cost function if its parameters are not heuristically adjusted to data levels.

Any of the image models discussed in previous chapters may in principle be applied to our problem. The most commonly found are variations of Markov random fields (MRFs), which provide relative simplicity in calculation and hyperparameter estimation. The examples in this chapter all include the generalized Gaussian MRF (GGMRF) [BOU 93], whose potential function has the form

$$U(\mathbf{x}) = \sum_{\{i,j\} \in C} \frac{|x_i - x_j|^q}{q\alpha^q}.$$

For $q > 1$, the GGMRF is among the convex models in \mathbf{x} , which maintain continuity of MAP solutions as functions of data, and relative simplicity of optimization. A host of other models are available with a similar performance given appropriate selection of hyperparameters (see Chapter 7). As is true of many others, the GGMRF includes the Gaussian MRF as a limiting case of $q = 2$. Much has been written concerning the advantages of edge-preserving prior models such as the GGMRF in Bayesian image reconstruction [BLA 87, GEM 84, GRE 90]. The Gaussian MRF applies heavy penalties to large differences among neighboring pixels, thereby discouraging the sharp discontinuities typical of many real images. Those models with a potential increasing less rapidly, such as GGMRF with $q \simeq 1$ or the logcosh model near its edge-preserving limit, allow these discontinuities to form naturally in estimates and can greatly improve reconstructions, particularly for piece-wise homogeneous objects. Non-convex priors produce even more dramatic edge rendering [BLA 87, CHA 97], often at the cost of desirable convergence properties. In more general, low SNR inverse problems the formation of sharp, low-contrast edges may not be a universal benefit. Contouring and unnatural texture may be visually disturbing to human interpreters of image

content, an important consideration in most medical imaging applications. The two estimates in Figure 14.5 demonstrate strong qualitative difference in limited-count data. The Gaussian MRF has two strong advantages: it helps form a simpler, more rapidly converging iterative estimate, and the solution's degradation in falling SNR is graceful and of a nature understood by experienced observers. We expect that the Gaussian will remain a popular option in low-intensity data.

In all the applications discussed here, unknown image \mathbf{X} may be assumed non-negative due to physical considerations. Should high quality measurements be available, such as in conventional X-ray computed tomography, a Bayesian image estimate may be little affected by enforcement of a non-negativity constraint. Higher-variance estimates resulting from low-intensity data violate this constraint much more

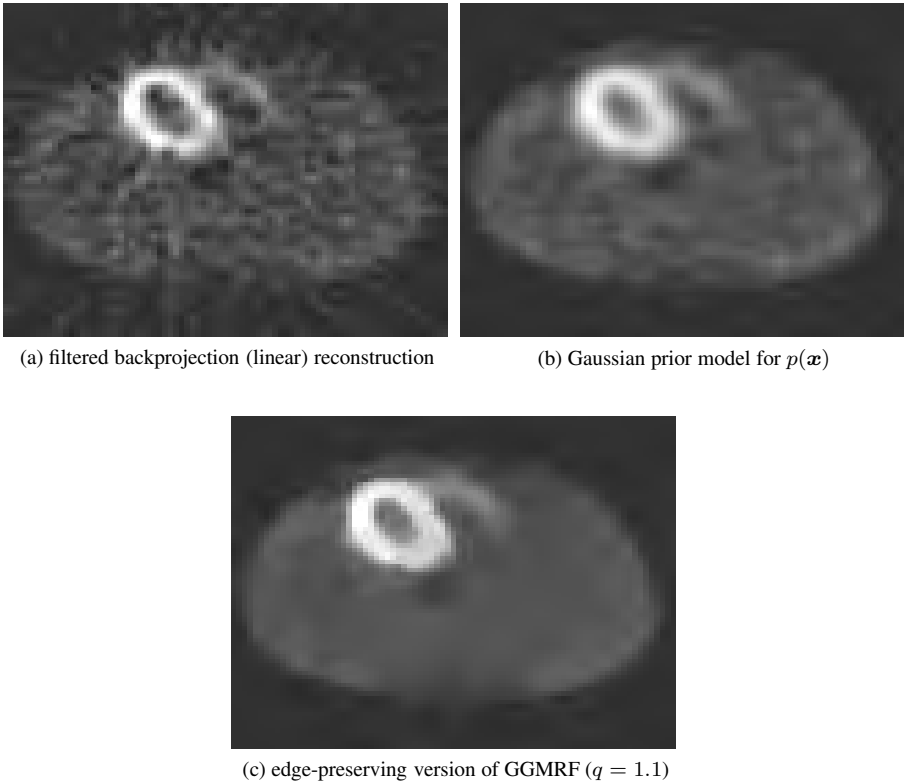


Figure 14.5. MAP reconstructions from single photon emission computed tomography (SPECT) heart perfusion data. Both (b) and (c) use ML values of scale parameter α . (Data courtesy of T.-S. Pan, M. King and University of Massachusetts)

frequently at their unconstrained optimum and are thus often appreciably improved by limiting estimates to non-negative values. In the SPECT heart imaging example of Figure 14.6, we see the constraint has a marked effect on the quality of reconstruction, particularly in less densely sampled portions of the image. This data is uncorrected for attenuation and scatter and may be improved by accurate inclusion of these effects in the model.

Characteristics related to the low-count Poisson data model in imaging have given rise to special models for \mathbf{X} as well. For example, the known positivity of \mathbf{X} also serves as motivation for the I-divergence prior model, which has been shown to be useful for low-intensity data [O'S 94]. The additivity of Poisson variates also facilitates multiscale representation of emission densities and specialized prior models for relations among nodes [TIM 99]. Paucity of data as well as low intensity has also led to more global, geometric approaches to reconstruction [CUN 98].

14.4. Implementation and calculation of Bayesian estimates

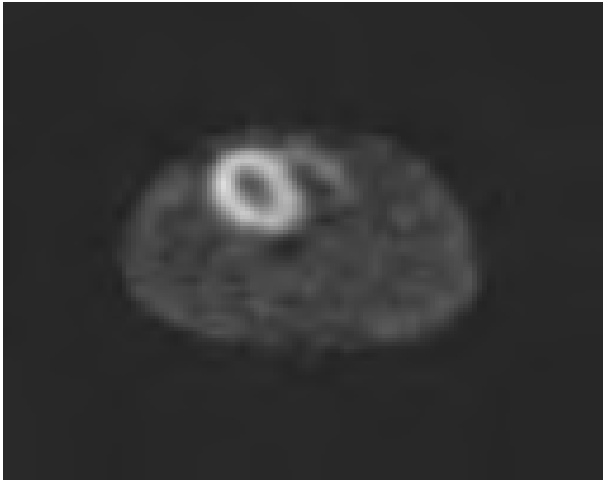
Any of the inverse problem formulations in the previous sections poses questions of implementation. Some aspects are very problem-dependent, such as attenuation and scatter correction in emission tomographic imaging. These issues arise nearly independently of the overall rates of emission and are thus not particularly a low-intensity imaging issue. For present purposes, we assume appropriate calibrations and corrections have been made such that the data used as input to the Bayesian inverse match the compound model of (14.3). The principal remaining question of implementation is thus the computational aspects of determining the Bayesian image estimate. Our discussion centers on the MAP reconstruction, which involves relatively direct optimization. More computationally demanding estimators such as the *a posteriori* mean [TIE 94] are applied relatively infrequently. Only rarely do such inverse problems allow the tractable direct calculation of $\widehat{\mathbf{X}}$, and iterative optimization is the norm.

14.4.1. Implementation for pure Poisson model

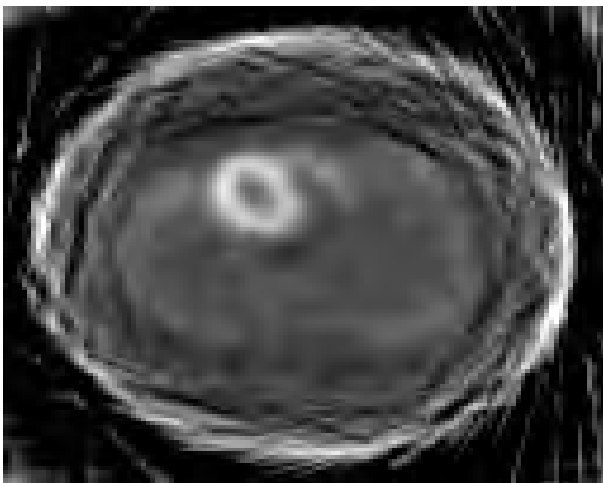
It is physically realistic and potentially beneficial to constrain $\widehat{\mathbf{X}}$ to be non-negative. This property appears to have been a major influence in the ascendance of EM type algorithms [KAU 87, SHE 82] as the most common option for optimizing over the Poisson log-likelihood. Though EM iterations for this problem are often viewed as a type of gradient descent, they yield a multiplicative correction at each iteration which guarantees preservation of positivity under a strictly positive initial condition. The k th update at pixel x_j has the form

$$x_j^{k+1} = x_j^k \sum_{i=1}^M \frac{A_{ij} y_i}{\mathbf{a}_{i*}^T \mathbf{x}^k + \beta_i}. \quad (14.6)$$

The EM algorithm is quite general and powerful for ML estimation, though the update step is not always so simple. The emission reconstruction update of (14.6) is as easy as a gradient step, but no practical version exists for the transmission case. Unmodified



(a) reconstruction under non-negativity constraint



(b) reconstruction without positivity constraint

Figure 14.6. MAP reconstruction with Gaussian prior model, with or without positivity constraint. In the latter case, maximally negative values have greater magnitude than positive maxima

EM is generally rather slow to converge, and has widely been replaced by OS-EM, which cycles among subsets of Y , updating pixels according to equation (14.6) for each subset. This optimization technique has also been proposed under the name “block iterative” [BYR 96]. This sacrifices the guaranteed convergence of EM, but since convergence to the ML image is seldom desired, heuristic termination times have proved useful. Generalizations of EM (and OS-EM) to handle the Bayesian estimate have been developed as well [DEP 95, HEB 89], but convergence limitations remain.

Viewing the Bayesian problem as in section 14.2, a broader view of optimization choices seems appropriate. Though we remain with the exact log-likelihood component in our problem, its proximity to a quadratic makes conjugate gradient (CG) methods [BEC 60] natural options. CG avoids oscillatory behavior often observed in such ill-conditioned problems. Preconditioned conjugate gradient (PCG) may further accelerate convergence by defining a form of conjugacy closer to optimal for the system at hand [LUE 73]. A difficulty for CG methods lies in the application of non-negativity constraints on \widehat{X} , for which several alternatives have been proposed [BIE 91, MUM 94]. These constraints necessarily alter the directions of updates at the expense of some convergence speed. Particularly in cases where large numbers of pixels encounter the constraint, CG falters somewhat. Figure 14.7 shows PCG in comparison to De Pierro’s adaptation of EM [DEP 95] and Lange’s Convex method [LAN 90], which is a similar algorithm for transmission tomography. The irregular path for PCG is due to the solution’s encountering the constraint.

Such optimization problems may also be solved via updates of single pixels or small subsets of the image, just as subsets of projection data or even single measurements (as in the algebraic reconstruction technique (ART) [HER 80]) are used for OS-EM. The advantage in this “column-action” approach is that guaranteed convergence can be maintained, and the reconstruction can easily be made independent of the starting condition with convex models. Two such methods are included in Figure 14.7: space-alternating expectation-maximization (SAGE), which develops pixel updates from an EM formulation [FES 93] and iterative coordinate descent (ICD), which optimizes the log-likelihood directly via local quadratic approximations in one dimension [BOU 96]. As updates are made in one dimension, enforcement of positivity is trivial for these techniques.

14.4.2. Bayesian implementation for a compound data model

Should the Gaussian component of equation (14.3) have large variance, we must deal with a more complex likelihood function under lower SNR, and without simple, explicit update expressions. A multiplicative, iterative algorithm for this case, similar to EM, has been presented [SNY 93], in which each update requires evaluation of a summation similar to (14.3). As in the preceding section, in which we considered the approximation of a Poisson log-likelihood with a low-order polynomial, we

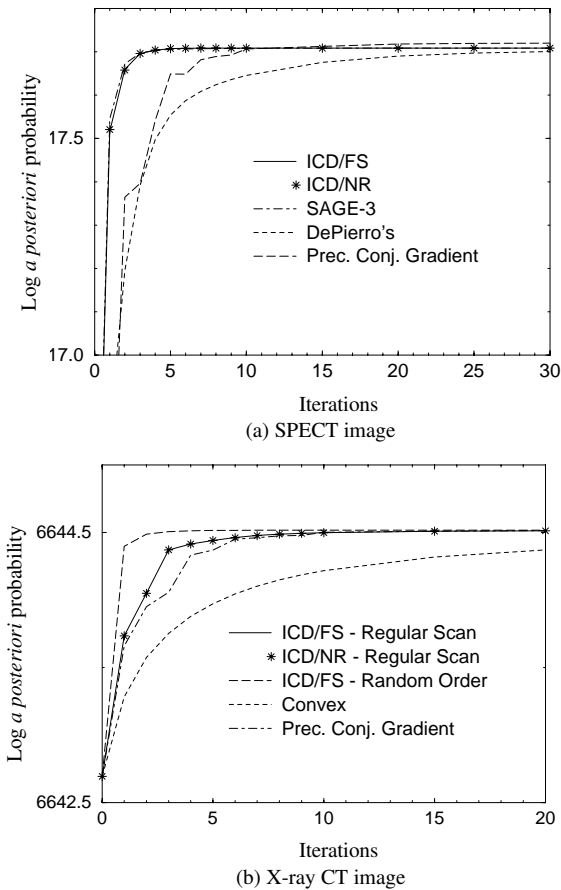


Figure 14.7. Convergence of various optimization methods for MAP reconstruction with a Gaussian prior image model

may make an approximation of the Poisson-Gaussian in order to simplify calculation, with little effect on the quality of the estimate. Alternatively, we may apply common iterative approaches after replacing each piece of data y_i with the approximation $y_i + \sigma_i - \gamma_i$ [SNY 93]. Using this transformation, the data variance becomes equal to the mean, which invites approximation by a Poisson distribution. Figure 14.4b shows that, within the range of parameters illustrated, this Poisson approximation yields a similar log-likelihood function to a conditional Gaussian whose mean and variance both depend on $\mu_i(x)$. The Poisson, though, appears to offer a simpler implementation than the Gaussian.

14.5. Conclusion

This chapter is dedicated to the reconstruction of images from limited intensity data, whose discrete nature is inherent in the counting of individual photons. Though we may view the inverse problem quite similarly to the standard Gaussian data case, dealing with the Poisson likelihood directly is necessary to gain the full advantages of Bayesian methods when event counts become few. We have provided a brief overview of unique features of these problems and effective approaches for the formulation and solution of their Bayesian estimators. The most appropriate formulation will depend on the specifics of the problem at hand. An understanding of the physical systems generating the observed phenomena must guide the choice of models for the most robust, accurate reconstructions.

14.6. Bibliography

- [BEC 60] BECKMAN F., “The solution of linear equations by the conjugate gradient method”, in RALSTON A., WILF H., ENSLEIN K. (Eds.), *Mathematical Methods for Digital Computers*, Wiley, 1960.
- [BIC 77] BICKEL P. J., DOKSUM K. A., *Mathematical Statistics: Basic Ideas and Selected Topics*, Holden-Day, Oakland, CA, 1977.
- [BIE 91] BIERLAIRE M., TOINT P. L., TUYTTENS D., “On iterative algorithms for linear least squares problems with bound constraints”, *Linear Alg. Appl.*, vol. 143, p. 111-143, 1991.
- [BLA 87] BLAKE A., ZISSERMAN A., *Visual Reconstruction*, The MIT Press, Cambridge, MA, 1987.
- [BOU 93] BOUMAN C. A., SAUER K. D., “A generalized Gaussian image model for edge-preserving MAP estimation”, *IEEE Trans. Image Processing*, vol. 2, num. 3, p. 296-310, July 1993.
- [BOU 96] BOUMAN C. A., SAUER K. D., “A unified approach to statistical tomography using coordinate descent optimization”, *IEEE Trans. Image Processing*, vol. 5, num. 3, p. 480-492, Mar. 1996.
- [BYR 96] BYRNE C. L., “Block-iterative methods for image reconstruction from projections”, *IEEE Trans. Image Processing*, vol. 5, p. 792-794, May 1996.
- [CHA 97] CHARBONNIER P., BLANC-FERAUD L., AUBERT G., BARLAUD M., “Deterministic edge-preserving regularization in computed imaging”, *IEEE Trans. Image Processing*, vol. 6, num. 2, p. 298-311, Feb. 1997.
- [CUN 98] CUNNINGHAM G. S., HANSON K. M., BATTLE X. L., “Three-dimensional reconstructions from low-count SPECT data using deformable models”, *Optics Express*, vol. 2, p. 227-236, 1998.
- [DEP 95] DE PIERRO A. R., “A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography”, *IEEE Trans. Medical Imaging*, vol. 14, num. 1, p. 132-137, 1995.

- [FES 93] FESSLER J. A., HERO A. O., "Complete data spaces and generalized EM algorithms", in *Proc. IEEE ICASSP*, Minneapolis, MN, p. IV 1-4, 1993.
- [FES 95] FESSLER J., "Hybrid Poisson/polynomial objective functions for tomographic image reconstruction from transmission scans", *IEEE Trans. Image Processing*, vol. 4, num. 10, p. 1439-1450, Oct. 1995.
- [GEM 84] GEMAN S., GEMAN D., "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, num. 6, p. 721-741, Nov. 1984.
- [GEM 85] GEMAN S., MCCLURE D., "Bayesian images analysis: An application to single photon emission tomography", in *Proc. Statist. Comput. Sect. Amer. Stat. Assoc.*, Washington, DC, p. 12-18, 1985.
- [GRE 90] GREEN P. J., "Bayesian reconstructions from emission tomography data using a modified EM algorithm", *IEEE Trans. Medical Imaging*, vol. 9, num. 1, p. 84-93, Mar. 1990.
- [HEB 89] HEBERT T., LEAHY R., "A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors", *IEEE Trans. Medical Imaging*, vol. 8, num. 2, p. 194-202, June 1989.
- [HER 80] HERMAN G. T., *Image Reconstruction from Projections. The Fundamentals of Computerized Tomography*, Academic Press, New York, NY, 1980.
- [HUD 94] HUDSON H., LARKIN R., "Accelerated image reconstruction using ordered subsets of projection data", *IEEE Trans. Medical Imaging*, vol. 13, num. 4, p. 601-609, Dec. 1994.
- [IIN 67] IINUMA T. A., NAGAI T., "Image restoration in radioisotope imaging system", *Phys. Med. Biol.*, vol. 12, num. 4, p. 501-509, Mar. 1967.
- [KAU 87] KAUFMAN L., "Implementing and accelerating the EM algorithm for positron emission tomography", *IEEE Trans. Medical Imaging*, vol. MI-6, num. 1, p. 37-51, 1987.
- [KOU 96] KOULIBALY P., Régularisation et corrections physiques en tomographie d'émission, PhD thesis, University of Nice-Sophia Antipolis, Nice, France, Oct. 1996.
- [LAN 90] LANGE K., "An overview of Bayesian methods in image reconstruction", in *Proc. SPIE Conf. on Digital Image Synth. and Inv. Optics*, vol. 1351, San Diego, CA, p. 270-287, July 1990.
- [LEV 87] LEVITAN E., HERMAN G., "A maximum *a posteriori* probability expectation maximization algorithm for image reconstruction in emission tomography", *IEEE Trans. Medical Imaging*, vol. MI-6, p. 185-192, Sep. 1987.
- [LLA 89] LLACER J., VEKLEROV E., "Feasible images and practical stopping rules for iterative algorithms in emission tomography", *IEEE Trans. Medical Imaging*, vol. 8, p. 186-193, 1989.
- [LUE 73] LUENBERGER D. G., *Introduction to Linear and Nonlinear Programming*, Addison-Wesley, New York, NY, 1st edition, 1973.
- [MEI 93] MEIKLE S. R., DAHLBOM M., CHERRY S. R., "Attenuation correction using count-limited transmission data in positron emission tomography", *J. Nuclear Med.*, vol. 34, num. 1, p. 143-144, 1993.

- [MUM 94] MUMCUOGLU E. U., LEAHY R., CHERRY S. R., ZHOU Z., "Fast gradient-based methods for Bayesian reconstruction of transmission and emission PET images", *IEEE Trans. Medical Imaging*, vol. 13, num. 4, p. 687-701, Dec. 1994.
- [NGU 99] NGUYEN M. K., GUILLEMIN H., FAYE C., "Regularized restoration of scintigraphic images in Bayesian frameworks", in *Proc. IEEE ICIP*, Kobe, Japan, p. 194-197, Oct. 1999.
- [OGA 91] OGAWA K., HARATA Y., ICHIHARA T., KUBO A., HASHIMOTO S., "A practical method for position-dependent Compton scatter correction in single photon emission CT", *IEEE Trans. Medical Imaging*, vol. 10, p. 408-412, Sep. 1991.
- [OLL 93] OLLINGER J. M., "Model-based scatter correction for fully 3D PET", in *IEEE Nuclear Science Symp. & Medical Imaging Conf.*, San Francisco, CA, p. 1264-1268, 1993.
- [OLL 97] OLLINGER J. M., FESSLER J. A., "Positron-emission tomography", *IEEE Signal Processing Mag.*, vol. 14, num. 1, p. 43-55, Jan. 1997.
- [O'S 94] O'SULLIVAN J. A., "Divergence penalty for image regularization", in *Proc. IEEE ICASSP*, vol. V, Adelaide, Australia, p. 541-544, Apr. 1994.
- [PAN 97] PAN T.-S., KING M. A., DE VRIES D. J., DAHLBERG S. T., VILLEGAS B. J., "Estimation of attenuation maps from single photon emission computed tomographic images of technetium 99m-labeled sestamibi", *J. Nucl. Cardiol.*, vol. 4, num. 1, p. 42-51, 1997.
- [POL 91] POLITTE D., SNYDER D. L., "Corrections for accidental coincidences and attenuation in maximum-likelihood image reconstruction for positron-emission tomography", *IEEE Trans. Medical Imaging*, vol. 10, p. 82-89, 1991.
- [SHE 82] SHEPP L. A., VARDI Y., "Maximum likelihood reconstruction for emission tomography", *IEEE Trans. Medical Imaging*, vol. MI-1, p. 113-122, 1982.
- [SNY 85] SNYDER D., MILLER M., "The use of sieves to stabilize images produced with the EM algorithm for emission tomography", *IEEE Trans. Nuclear Sciences*, vol. NS-32, num. 5, p. 3864-3871, Oct. 1985.
- [SNY 90] SNYDER D. L., SCHULZ T. J., "High-resolution imaging at Low Light Levels through Weak Turbulence", *J. Opt. Soc. Am. (A)*, vol. 7, p. 1251-1265, 1990.
- [SNY 93] SNYDER D. L., HAMMOUD A. M., WHITE R. L., "Image recovery from data acquired with a charge-coupled-device camera", *J. Opt. Soc. Am. (A)*, vol. 10, p. 1014-1023, 1993.
- [THI 00] THIBAUT J.-B., SAUER K., BOUMAN C., "Newton-style optimization for emission tomographic estimation", *J. Electr. Imag.*, vol. 9, num. 3, p. 269-282, 2000.
- [TIE 94] TIERNEY L., "Markov chain for exploring posterior distribution", *Annals Statist.*, vol. 22, num. 4, p. 1701-1762, Dec. 1994.
- [TIM 99] TIMMERMAN K. E., NOWAK R. D., "Multiscale modeling and estimation of Poisson processes with application to photon-limited imaging", *IEEE Trans. Inf. Theory*, vol. 45, num. 3, p. 846-862, Apr. 1999.
- [TSU 91] TSUI B., FREY E., GULLBERG G., "Comparison between ML-EM and WLS-CG algorithms for SPECT image reconstruction", *IEEE Trans. Nuclear Sciences*, vol. 38, num. 6, p. 1766-1772, Dec. 1991.

List of Authors

Laure BLANC-FÉRAUD
CNRS
Nice-Sophia Antipolis
France

Hervé CARFANTAN
Observatoire Midi-Pyrénées
Toulouse
France

Frédéric CHAMPAGNAT
ONERA
Châtillon
France

Guy DEMOMENT
L2S
University of Paris XI
France

Xavier DESCOMBES
INRIA
Nice-Sophia Antipolis
France

Jean-Marc DINTEN
LETI
CEA
Grenoble
France

Stéphane GAUTIER
EDF R&D
Chatou
France

Jean-François GIOVANNELLI
L2S
University of Paris XI
France

Yves GOUSSARD
Ecole Polytechnique Montreal
Quebec, Canada

Alain HERMENT
INSERM
Paris
France

Jérôme IDIER
IRCCyN
CNRS
Nantes
France

Guy LE BESNERAIS
ONERA
Châtillon
France

Serge MEIMON
ONERA
Châtillon
France

Ali MOHAMMAD-DJAFARI
L2S
CNRS
Gif-sur-Yvette
France

Laurent MUGNIER
ONERA
Châtillon
France

Ken SAUER
University of Notre Dame, IN
USA

Jean-Baptiste THIBAUT
GE Medical Systems
Milwaukee, WI
USA

Index

A

A-scan 223
accidental coincidence 360, 362
adaptive optics (AO) 251, 263-267
affine approximation 348
algorithm
 ART 50, 325, 329
 CLEAN 118
 column-action 50, 370
 EM 206, 208, 264, 303, 365, 368
 forward-backward 302
 Gerchberg-Saxton-Papoulis-
 Van Cittert 43, 264
 Gibbs sampling 190, 327
 golden section 295
 gradient 50, 145, 208, 303, 324
 conjugate (CG) 50, 94, 133, 370
 preconditioned conjugate (PCG)
 50, 96, 370
 ICD 158, 160, 295, 326, 352, 370
 ICM 129, 329
 Levinson 49, 94
 limited-memory 51
 Metropolis sampling 190
 Newton, quasi-Newton 51
 reconstruction 323
 Richardson-Lucy 265, 326
 row-action 50
 SAGE 370
 SIRT 325
 SMLR 126, 129

Viterbi 302

AO *see* adaptive optics
astronomy 243-276
atmospheric turbulence 243-276
attenuation 360

B

B-scan 223
backprojection (BP) 315
Bayes' rule 64, 197
bias 98
bias-variance compromise 98
boundary effect 87, 96
BP *see* backprojection

C

Chandrasekhar factorization 110
clique 174
complete conditional 214
condition
 boundary 88, 89, 107
 Hadamard 30, 149, 154
 number,conditioning 38, 89, 291
 positivity 180
conductivity 338
covariance 72, 75
 a posteriori 186
criterion
 augmented 148-158
 convex 51, 123, 231, 323, 366
 half-quadratic (HQ) 149-158, 183

Picard 33
 quadratic 323
 Rayleigh 83
 regularized 29, 92
 cross-validation 54

D

DCT *see* transform (discrete cosine)
 deconvolution 16-18
 blind 134, 226, 232
 conventional 253
 DBG 230
 DL2Hy 230, 231
 from wavefront sensing (DWFS) 250,
 259-263
 L2Hy 123
 L2LP 123
 minimum entropy 228
 multipulse 118, 228
 myopic
 long-exposure images 264-267
 short-exposure images 260-263
 predictive 106, 226
 detection-estimation 119, 148, 154
 DFT *see* transform (discrete Fourier)
 diffraction 244
 diffusion
 anisotropic 155
 isotropic 145
 discontinuity 147, 148
 discretization 340
 distribution *see* law
 Doppler 285, 286
 DWFS *see* deconvolution from wavefront
 sensing

E

electrical permittivity 338
 equation
 coupled 340, 343
 diffusion 144
 Euler-Lagrange 35, 143, 145
 Fredholm 31
 observation 293
 partial derivative (PDE) 144
 propagation 338

Riccati 105
 state *see* model (state)
 ergodicity 189
 estimation
 Bayesian 64-75, 172-174
 joint 258, 261, 345
 linear 81
 marginal 258
 robust 159
 exponential family 70, 198

F

FFT *see* transform (fast Fourier)
 field
 Gibbs-Markov (GMRF) 70, 173-185,
 202-204, 207, 210, 366
 compound 183-185
 Gaussian 182
 non-Gaussian 185
 mean 204
 filter
 inverse 82, 84
 matched 117, 119
 minimum phase 227, 229
 stabilizing 84
 Wiener 75, 84, 95, 111, 121
 small kernel 111

Fourier
 series 27, 93
 synthesis 245, 271, 315
 frequency
 mean 286, 297
 pure 290, 297
 frequentist interpretation 62
 Fried's diameter 247, 255
 FT *see* transform (Fourier)
 full Bayesian approach 213
 function
 L_2L_1 47, 151, 254-276
 L_2L_0 47, 152
 basis 320, 341
 Huber 161, 189
 partition 174, 183

G

generalized inverse 35, 91

geophysics 338

Gibbs

phenomenon 102

potential 174

GML *see* maximum likelihood
(generalized)

GMRF *see* field (Gibbs-Markov)

GNC *see* gradual non-convexity

graduated non-convexity 123, 149, 155,
159, 352

H

Hessian 200, 365

HT *see* transform (Hilbert)

hyperparameter 53, 67, 147, 162, 186, 294,
302, 366

I

image registration 181-182

imaging

magnetic resonance 285

medical 285

microwave 337

optical 243-276

scintigraphic 358

ultrasound 221-241, 285, 286

importance sampling 200

interferometry 27, 46

inverse crime 18

K

Kalman

filter 102, 121, 293

asymptotic 110

fast 108

non-standard 111

smoother 293

fixed-interval 103, 107

fixed-lag 103

Kullback divergence, Kullback
pseudo-distance 47, 69, 206

L

L-curve 53

law

Beer-Lambert 312

compound 363

direct 61

Gaussian 62

improper 69, 176, 183

Kolmogorov 246

Poisson 359, 360

proper 66

least squares 18, 290, 365

adaptive 290

recursive 105

reweighted (RLS) 159

likelihood 63-64, 186, 199, 295, 298, 362

generalized 68, 212

joint 124

marginal 67, 73, 124

local minimum 51, 52, 347

M

MAP *see* maximum *a posteriori*

marginalization 298

Markov chain 176, 298

Monte carlo *see* method (MCMC)

snake 106

matrix

circulant 49, 88

circulant-block-circulant 88

Hankel 89

ill-conditioned 37

sparse 49

Toeplitz 49, 87, 161

Toeplitz-block-Toeplitz 49, 88, 160,
161

Toeplitz-plus-Hankel 89

maximum

a posteriori (MAP) 65, 173, 301, 322,
366

likelihood (ML) 63, 186, 200, 302, 364
generalized (GML) 68, 212, 213

mean

a posteriori 186

a priori 93

mean square error (MSE) 54, 72, 73

measure

Hausdorff 148

roughness 46

method

- analytical 317
- Bialy's 43
- coding 202
- Gauss-Seidel 50
- Hunt 94
- Landweber 43, 50
- Lucy's 44
- maximum entropy (MEM) 69
- MCMC 130, 192-193
- Monte Carlo 189, 200
- Newton 200
- of moments 341
- quadrature 85

ML *see* maximum likelihood

model

- AR 105, 290
 - long 290
- ARMA 290
- Bernoulli-Gaussian (BG) 119, 124
- binary 328
- geometrical 328
- in companion form 106
- Ising 181
- nonlinear 335
- Potts 182
- random walk 106
- state 85, 102, 293
 - degenerate 104, 110
- white noise 106

MSE *see* mean square error

N

- NDE *see* non-destructive evaluation
- neighborhood relationship 177
- non-destructive evaluation (NDE) 337
 - ultrasonic 221-241
- non-negativity constraint 367

O

- optical interferometry (OI) 251, 268
- optimization
 - constrained 351
 - global 351
 - joint 350
 - local 352

P

- PDE *see* partial derivative equations
- periodogram 28, 36, 46, 287, 299
- PET *see* tomography (positron emission)
- phase indetermination 252, 271
- polygon, polyhedron 328
- preconditioning 96, 160
- probability density function (pdf) 65
- problem
 - direct 15-19, 25
 - ill-posed 16-19, 30-34
 - instrumentation 25
 - inverse 15-19, 25
 - well-posed 30
- pseudo-likelihood 203
- pseudo-solution 35

R

- random search 53, 191
- reflectivity 118, 225
- regularization
 - L_2L_1 *see* function (L_2L_1)
 - L_2L_0 *see* function (L_2L_0)
 - quadratic *see* Tikhonov regularization
- regularizer 41
- relaxation *see* algorithm (ICD)
- RLS *see* least squares (reweighted)
- robust norm 151
- RT *see* transform (Radon)

S

- scattered photon 360, 362
- simulated annealing 53, 193, 327, 351
- single most likely replacement *see* SMLR
 - algorithm
- singular value 33, 72
 - decomposition
 - truncated (TSVD) 42, 101
 - decomposition (SVD) 33
- Sobolev distance 292
- speckle 247, 250
- spectral
 - aliasing 286, 297
 - analysis 286
 - characterization 285
 - density

- energy (ESD) 265, 267
- power (PSD) 84, 102, 246, 247, 257, 263, 274, 286-297
- equalization 81, 85, 112, 121
- moment 286, 297
- regularization 292

spectrogram 287

state representation *see* model (state)

successive linearization 348, 351

SVD *see* singular value decomposition

T

theorem

- Hammersley-Clifford 179

- Riesz 32

- Szegő's 90

- Van Cittert-Zernike 246

Tikhonov regularization 17, 45, 92, 93, 142, 158, 164, 292, 300

time-frequency analysis 286

tomography 312

- diffraction 335

- discrete 319

- emission 358, 367

- positron (PET) 358

- transmission 362

transform

- discrete cosine (DCT) 89, 96

- Fourier (FT) 26-29, 315

- discrete (DFT) 88-89, 94-96, 98

- fast (FFT) 88, 97, 302

- Hilbert (HT) 230, 315

- J-orthogonal 109

- Radon (RT) 313

transformation group 69

V

variable

- complex 344

- edge 147

- hidden 68, 149

- interactive 149, 184-185

- uncoupled 157

variance 99

W

wavefront sensor (WFS) 251, 260, 261, 264

- Hartmann-Shack 251, 255-257, 260, 262, 265

- phase

- diversity 255, 257

- retrieval 257

WFS *see* wavefront sensor

windowing 291

Z

Zernike polynomial 256, 257, 261