The Fundamentals of Density Functional Theory (revised and extended version)

H. Eschrig
Institute for Solid State and Materials Research Dresden
and
University of Technology Dresden

Preface

Density functional methods form the basis of a diversified and very active area of present days computational atomic, molecular, solid state and even nuclear physics. A large number of computational physicists use these methods merely as a recipe, not reflecting too much upon their logical basis. One also observes, despite of their tremendous success, a certain reservation in their acceptance on the part of the more theoretically oriented researchers in the above mentioned fields. On the other hand, in the seventies (Thomas-Fermi theory) and in the eighties (Hohenberg-Kohn theory), density functional concepts became subjects of mathematical physics.

In 1994 a number of activities took place to celebrate the thirtieth anniversary of Hohenberg-Kohn-Sham theory. I took this an occasion to give lectures on density functional theory to senior students and postgraduates in the winter term of 1994, particularly focusing on the logical basis of the theory. Preparing these lectures, the impression grew that, although there is a wealth of monographs and reviews in the literature devoted to density functional theory, the focus is nearly always placed upon extending the practical applications of the theory and on the development of improved approximations. The logical foundation of the theory is found somewhat scattered in the existing literature, and is not always satisfactorily presented. This situation led to the idea to prepare a printed version of the lecture notes, which resulted in the present text.

It is my intention to provide a thorough introduction to the theoretical basis of density functional methods in a form which is both rigorous and yet concise. It is aimed not only for those who are already or are going to be active in the field, but also for those who just want to get a deeper insight into the meaning of the results of practical calculations, and last but not least to provide the interested mathematician with the physicist's view on the logical roots of the theory. High value is put on the self-containment of the text, so that it should be accessible to anybody who has been through the standard course in quantum mechanics.

The shorter Part II of these notes deals with the relativistic theory. This part of the theory is less rigorous logically, due to the unsolved basic problems in quantum field theory, but it is nevertheless physically very important because heavier atoms and many problems of magnetism need a relativistic treatment.

Finally, as is always the case, many things had to be omitted. I decided in particular to omit everything which seemed to me to not yet have its final form or at least be in some stabilized shape. This does not exclude that important points are missing, just because I was unaware of them.

The author's views on the subject have been sharpened by discussions with many colleagues, but particularly during the course of scientific cooperation with M. Richter and P. M. Oppeneer on the applied side of the story.

Dresden, July 1996 Helmut Eschrig

In the present updated and extended version, errors and misprints have been corrected in the original text. Throughout, an effort has been made additionally to improve overall clarity. This includes in particular major changes in Chapter 4. Here, also the non-collinear spin case was added as Section 4.8, a very important issue presently. Chapter 6 and Section 9.2, however, have been completely rewritten, since in the opinion of the author the material gains much in systematics and clarity, if one against tradition starts with the dependence on the particle number and afterwards treats the density-potential interrelation. Although not yet completely developed, the presently most common versions of LDA+U have been included as Section 7.5. The last section of the first edition on the other hand has been omitted as it is not relevant any more: with the drastic increase of computer data storage capacities, radial basis functions are stored numerically now and the direct projection onto the electron sector of the Dirac-Fock space is not a problem any more.

Dresden, August 2003 Helmut Eschrig

Contents

Introduction **9**

Part I: NON-RELATIVISTIC THEORY **12**

1 Many-Body Systems **13**
1.1 The Schrödinger Representation, N Fixed 14
1.2 The Momentum Representation, N Fixed 17
1.3 The Heisenberg Representation, N Fixed 24
1.4 Hartree-Fock Theory 25
1.5 The Occupation Number Representation, N Varying 29
1.6 Field Quantization 33

2 Density Matrices and Density Operators **35**
2.1 Single-Particle Density Matrices 35
2.2 Two-Particle Density Matrices 38
2.3 Density Operators 41
2.4 Expectation Values and Density Matrices 43
2.5 The Exchange and Correlation Hole 47
2.6 The Adiabatic Principle 52
2.7 Coulomb Systems 55

3 Thomas-Fermi Theory **61**
3.1 The Thomas-Fermi Functional and Thomas-Fermi Equation 62
3.2 The Thomas-Fermi Atom 67
3.3 The Thomas-Fermi Screening Length 69
3.4 Scaling Rules 71
3.5 Correction Terms 72

4 Hohenberg-Kohn Theory **76**
4.1 The Basic Theorem by Hohenberg and Kohn 77
4.2 The Kohn-Sham Equation 81
4.3 The Link to the Hartree-Fock-Slater Approximation 86
4.4 Constrained Search Density Functionals 87
4.5 Ensemble State Density Functionals 89
4.6 Dependence on Particle Number N 95
4.7 Spin Polarization 100
4.8 Non-Collinear Spin Configurations 105

5 Legendre Transformation **109**
5.1 Elementary Introduction . . . 110
5.2 Prelude on Topology . . . 116
5.3 Prelude on Lebesgue Integral . . . 121
5.4 Banach Space . . . 124
5.5 Dual Space . . . 129
5.6 Conjugate Functionals . . . 131
5.7 The Functional Derivative . . . 132
5.8 Lagrange Multipliers . . . 135

6 Density Functional Theory by Lieb **138**
6.1 The Ground State Energy . . . 139
6.2 The Hohenberg-Kohn Variational Principle . . . 142
6.3 The Functionals F, G, and H . . . 144
6.4 The Kohn-Sham Equation . . . 151

7 Approximative Variants **155**
7.1 The Homogeneous Electron Liquid . . . 156
7.2 The Local Density Approximation . . . 161
7.3 Generations of Kohn-Sham Type Equations . . . 165
7.4 The Self-Interaction Correction . . . 168
7.5 The LDA+U Approach . . . 171

Part II: RELATIVISTIC THEORY **178**

8 A Brief Introduction to Quantum Electrodynamics **179**
8.1 Classical Electrodynamics . . . 180
8.2 Lorentz Covariance . . . 181
8.3 Lagrange Formalism . . . 183
8.4 Relativistic Kinematics . . . 186
8.5 Relativistic Mechanics . . . 187
8.6 The Principles of Relativistic Quantum Theory . . . 188
8.7 The Dirac Field . . . 191

9 Current Density Functional Theory **196**
9.1 QED Ground State in a Static External Field . . . 197
9.2 Current Density Functionals and Kohn-Sham-Dirac Equation . . . 202
9.3 The Gordon Decomposition and Spin Density . . . 207
9.4 Approximative Variants . . . 210

Bibliography **214**

Index **223**

Introduction

Density functional theory provides a powerful tool for computations of the quantum state of atoms, molecules and solids, and of *ab-initio* molecular dynamics. It was conceived in its initial naïve and approximative version by Thomas and Fermi immediately after the foundation of quantum mechanics, in 1927. In the middle of the sixties, Hohenberg, Kohn and Sham on the one hand established a logically rigorous density functional theory of the quantum ground state on the basis of quantum mechanics, and on the other hand, guided by this construction, introduced an approximative explicit theory called the local-density approximation, which for computations of the quantum ground state of many-particle systems proved to be superior to both Thomas-Fermi and Hartree-Fock theories. From that time on, density functional theory has grown vastly in popularity, and a flood of computational work in molecular and solid state physics has been the result. Motivated by its success, there has been always a tendency to widen the fields of application of density functional theory, and in these developments, some points which were left somewhat obscure in the basic theory, were brought into focus from time to time. This led in the early eighties to a deepening of the logical basis, essentially by Levy and Lieb, and finally Lieb gave the basic theory a form of final mathematical rigor. Since that treatment, however, is based on the tools of modern convex functional analysis, its implications only gradually became known to the many people who apply density functional theory. A thorough treatment of the dependence on particle number on the basis of Lieb's theory is given for the first time in the present text.

While quite a number of high quality and up-to-date surveys and monographs on the variants and applications of density functional theory exist, the aim of these lecture notes is to provide a careful introduction into the safe basis of this type of theory for both beginners and those working in the field who want to deepen their understanding of its logical basis. To this end, references are not only given to further introductory texts, surveys and some key original papers regarding the physics, but also to literature treating the underlying mathematics. Only Chapter 5 of this text provides a more systematic and self-explanatory introduction to a piece of modern

mathematics appropriate in this context, namely convex (non-linear) functional analysis, and in particular duality theory. This chapter was included for those physicists who either are for one reason or another not willing to dive into comprehensive treatments of modern mathematics or who want some guidance into those parts of mathematics necessary for a deeper understanding of the logics of density functional theory. Even the material of this chapter is presented on an introductory level: proofs are only sketched if they are essentially enlightening as regards the logical interrelations.

The material has been divided into two parts. Part I contains the non-relativistic theory, and, because this part has to a larger extent achieved some final logical state, it contains a discussion of the central ideas and constructions of density functional theory, including some relevant mathematical aspects. The shorter Part II is devoted to the relativistic extensions, which have not yet reached the same level of rigor, mainly because realistic quantum field theory is in a much less explicit state than quantum mechanics.

Chapters 1, 2 and at least part of 3 are elementary and provide a physical introduction for students having been through the standard first course in quantum mechanics. Those who are already to some extent familiar with many-body quantum theory may immediately start with Chapter 4 and use the first chapters only for reference (for instance to explain the notation used). Care has been taken to present the material of Part II in such a way that it should be readable also for somebody who is not very familiar with quantum field theory, yet without going smoothly around the reefs. Some remarks scattered throughout the text address the more physics-minded mathematician.

Notation is systematic throughout the text. Bold math ($\boldsymbol{r}$) denotes a usually three-dimensional vector, z^* is the complex conjugate to the complex number z, $\hat{A}$ is the operator corresponding to the quantum observable A (for brevity the same symbol is used to denote an operator in different representation spaces), $\hat{A}^\dagger$ is the Hermitian conjugate to $\hat{A}$, $\det ||A_{ik}||$ means the determinant of the matrix $||A_{ik}||$, while A_{ik} or $\langle i|A|k\rangle$ is the matrix element. In cross-references, (2.1) refers to Formula(e) 1 in Chapter 2. (In some cases the formula number refers to a couple of lines of mathematical expressions not interrupted by text lines; the formula number mostly appears on or below the last of those lines.) Citations are given as the author(s) and year in square brackets. In case the reader is confused by some notion he has no definite idea of, it can most probably be found in the index at the end of the book with a reference to places in the text, where the notion is

explained or put into context.

Density functional theory can be built up in several versions: (i) as a theory with particle densities (summed over spin variables) and spin-independent external potentials only, irrespective whether the quantum state is spin-polarized or not; (ii) as a theory with spin-up and spin-down densities and external potentials which possibly act differently on spin-up and spin-down particles for collinear polarization situations with one global spin quantization direction (generalization to more than two eigenvalues of the z-component of the spin is straightforward); and finally, (iii) as a general theory with (spatially diagonal) spin-density matrices and general doubly indexed spin-dependent potentials. All three versions are treated in the present text, the two former cases are considered in parallel throughout by consequently using a combined variable $x = (\boldsymbol{r}, s)$ of spatial position $\boldsymbol{r}$ and z-component of spin s. Most expressions given refer immediately to the spin-independent cases, if x is read as synonymous with $\boldsymbol{r}$. The reason for presenting the material in this way is that practically all formulae thereby refer simultaneously to both cases, whereas the most general case of spin-density matrices does not lead to essentially new aspects in many questions addressed in large parts of the text.

These are lecture notes, hence neither an attempt in any respect is made to give a complete reference list of existing literature, nor are questions of priority addressed. The cited literature was selected exclusively (and subjectively of course) to facilitate the access to the subject, and to give advice for delving deeper into related subjects. Nevertheless, key original papers which have introduced a new approach or a new theoretical tool are cited, since the author is convinced that in particular students will benefit from reading the key original papers.

Part I:
NON-RELATIVISTIC THEORY

1 Many-Body Systems

The dynamics of a quantum system is governed by the Hamiltonian $\hat{H}$. If $|\Psi\rangle$ is a quantum state of the system in an abstract Hilbert space representation, its time-evolution is given by

$$-\frac{\hbar}{i}\frac{\partial}{\partial t}|\Psi\rangle = \hat{H}|\Psi\rangle. \tag{1.1}$$

Stationary states with definite energies, particularly the ground state of the system, are obtained as solutions of the eigenvalue equation

$$\hat{H}|\Psi\rangle = |\Psi\rangle E, \quad \langle\Psi|\Psi\rangle = 1. \tag{1.2}$$

Alternatively they are obtained as stationary solutions of the variational problem

$$\frac{\langle\Psi|\hat{H}|\Psi\rangle}{\langle\Psi|\Psi\rangle} \Rightarrow \text{stationary}. \tag{1.3}$$

The variation (with respect to $|\Psi\rangle$) of the numerator on the left-hand side with the denominator kept fixed equal to unity leads immediately to (1.2), the energy, E, thereby appearing as a Lagrange multiplier corresponding to the latter constraint.

We are interested in systems of N identical particles (electrons, say) moving in a given external field and interacting with each other with pair forces. The Hamiltonian for this case consists of the kinetic energy operator $\hat{T}$, the potential operator $\hat{U}$ of the interaction of the particles with the external field, and the two-particle interaction operator $\hat{W}$:

$$\hat{H} = \hat{T} + \hat{U} + \hat{W}. \tag{1.4}$$

The case $\hat{W} = 0$, of particles which do not interact with each other, i.e.

$$\hat{H}_0 = \hat{T} + \hat{U}, \tag{1.5}$$

is often considered as a reference system.

In the following we subsequently introduce the most important actual representations of the Hilbert space for the quantum states. (For details see standard textbooks of quantum theory as, e.g., [Sakurai, 1985, Dawydow, 1987, Dirac, 1958, von Neumann, 1955].)

1.1 The Schrödinger Representation, N Fixed

This spatial representation *formally* uses the eigenstates of the coordinates $\boldsymbol{r}$ (and possibly of the spin projection s with respect to some given quantization axis z) of the particles

$$\hat{\boldsymbol{r}}\,|\boldsymbol{r}\rangle = |\boldsymbol{r}\rangle\,\boldsymbol{r}, \quad \hat{\sigma}\,|s\rangle = |s\rangle\,s \tag{1.6}$$

as basis vectors in the Hilbert space of one-particle quantum states. Here, $\hat{\sigma}$ is the z-component of the spin operator. The subscript z will generally be omitted in order not to overload the notation. A combined variable

$$x \stackrel{\text{def}}{=} (\boldsymbol{r}, s), \quad \int dx \stackrel{\text{def}}{=} \sum_s \int d^3r \tag{1.7}$$

will be used throughout for both position and spin of a particle. We will generally have in mind N-electron systems, large parts of what follows apply, however, to a general case of identical particles, having a spin (or some other internal degree of freedom) or not. In expressions applying for both spinless particles and particles with spin, x and $\boldsymbol{r}$ may be considered synonyms in the spinless case.

The N-particle quantum state is now represented by a (spinor-)wavefunction

$$\Psi(x_1 \ldots x_N) = \langle\, x_1 \ldots x_N \,|\Psi\rangle = \langle\, \boldsymbol{r}_1 s_1 \ldots \boldsymbol{r}_N s_N \,|\Psi\rangle. \tag{1.8}$$

The spin variable s_i runs over a finite number of values only ($2S+1$ values for spin-S particles). For one spin-half particle, e.g., the spinor part of the wavefunction (for fixed $\boldsymbol{r}$),

$$\chi(s) = \langle s|\chi\rangle = \begin{pmatrix} \chi_+ \\ \chi_- \end{pmatrix}, \tag{1.9}$$

consists of two complex numbers χ_+ and χ_-, forming the components of a spinor. This latter statement means that a certain linear transformation of those two components is linked to every spatial rotation of the $\boldsymbol{r}$-space (see, e.g., [Landau and Lifshitz, 1977], or any textbook on quantum mechanics). The operator $\hat{\sigma} \equiv \hat{\sigma}_z$ is now represented by a 2×2 matrix. For later use we also give the operators for the x- and y-components of the spin in this representation:

$$\hat{\sigma}_z = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \hat{\sigma}_x = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \hat{\sigma}_y = \frac{1}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}. \tag{1.10}$$

The eigenstates of $\hat{\sigma}_z$,

$$\chi^+(s) = \langle s|\chi^+\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \chi^-(s) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{1.11}$$

form a complete set for the s-dependence at a given space-point $\boldsymbol{r}$:

$$\langle\chi^+|\chi\rangle = \sum_s \langle\chi^+|s\rangle\langle s|\chi\rangle = (1\ 0)\begin{pmatrix} \chi_+ \\ \chi_- \end{pmatrix} = \chi_+, \quad \langle\chi^-|\chi\rangle = \chi_-. \tag{1.12}$$

The full wavefunction of a spin-half particle is given by

$$\phi(x) = \phi(\boldsymbol{r}s) = \begin{pmatrix} \phi_+(\boldsymbol{r}) \\ \phi_-(\boldsymbol{r}) \end{pmatrix}. \tag{1.13}$$

It is called a spin-orbital.

For fermions (half-integer spin), only wavefunctions which are antisymmetric with respect to particle exchange are admissible:

$$\Psi(x_1 \dots x_i \dots x_k \dots x_N) = -\Psi(x_1 \dots x_k \dots x_i \dots x_N). \tag{1.14}$$

In the non-interacting case ($\hat{H}_0$), Slater determinants

$$\Phi_L(x_1 \dots x_N) = \frac{1}{\sqrt{N!}} \det \|\phi_{l_i}(x_k)\| \tag{1.15}$$

of single-particle wavefunctions $\phi_{l_i}(x_k)$ (spin-orbitals) are appropriate. The subscript L denotes an orbital configuration $L \stackrel{\text{def}}{=} (l_1 \dots l_N)$. The determinant (1.15) can be non-zero only if the orbitals ϕ_{l_i} are linear independent, it is normalized if the orbitals are orthonormal. Furthermore, if the ϕ_{l_i} may be written as $\phi_{l_i} = \phi'_{l_i} + \phi''_{l_i}$, where $\langle\phi'_{l_i}|\phi'_{l_k}\rangle \sim \delta_{ik}$, and the ϕ''_{l_i} are linear dependent on the ϕ'_{l_k}, $k \neq i$, then the value of the determinant depends on the orthogonal to each other parts ϕ'_{l_i} only. These statements following from simple determinant rules comprise Pauli's exclusion principle for fermions. For a given fixed complete set of spin-orbitals $\phi_k(x)$, i.e. for a set with the property

$$\sum_k \phi_k(x)\phi_k^*(x') = \delta(x - x'), \tag{1.16}$$

the Slater determinants for all possible orbital configurations span the antisymmetric sector of the N-particle Hilbert space. In particular, the general state (1.14) may be expanded according to

$$\Psi(x_1 \dots x_N) = \sum_L C_L \Phi_L(x_1 \dots x_N) \tag{1.17}$$

('configuration interaction').

Bosonic (integer, in particular zero spin) wavefunctions must be symmetric with respect to particle exchange:

$$\Psi(x_1 \dots x_i \dots x_k \dots x_N) = \Psi(x_1 \dots x_k \dots x_i \dots x_N). \tag{1.18}$$

The corresponding symmetric sector of the N-particle Hilbert space is formed by the product states

$$\Phi_L(x_1 \dots x_N) = \mathcal{N} \sum_{\mathcal{P}} \prod_i \phi_{l_{\mathcal{P}i}}(x_i), \tag{1.19}$$

where $\mathcal{N}$ is a normalization factor, and $\mathcal{P}$ means a permutation of the subscripts $12 \dots N$ into $\mathcal{P}1\mathcal{P}2 \dots \mathcal{P}N$. The subscripts l_i, $i = 1 \dots N$, need not be different from each other in this case. Particularly all ϕ_{l_i} might be equal to each other (Bose condensation).

For both fermionic and bosonic systems, the probability density of a given configuration $(x_1 \dots x_N)$,

$$p(x_1 \dots x_N) = \Psi^*(x_1 \dots x_N)\Psi(x_1 \dots x_N), \tag{1.20}$$

is independent of particle exchange.

The Hamiltonian acting on Schrödinger wavefunctions is now explicitly given as

$$\hat{H} = -\frac{\hbar^2}{2m} \sum_{i=1}^{N} \nabla_i^2 + \sum_{i=1}^{N} v(x_i) + \frac{\lambda}{2} \sum_{i \neq j}^{N} w(|\boldsymbol{r}_i - \boldsymbol{r}_j|), \tag{1.21}$$

where $v(x_i)$ is the potential of the external field acting on the particle with position and spin x_i. For later considerations we allow for a spin dependent field which may be visualized as an external magnetic field $B(\boldsymbol{r})$ in z-direction acting according to a potential term $-2sB(\boldsymbol{r})$ on the spin only while its effect on the orbital motion is neglected. Note that this is even not the most general spin dependence of the external field, because we put v depending on the same z-component of the spin operator (1.10) everywhere in $\boldsymbol{r}$-space. (The most general spin dependent external field would consist of four spatial functions, indexed by two spinor indices (ss') of a general (spatially local) spin-half operator $v_{ss'}(\boldsymbol{r})$; this is used in density functional treatments of ground states with non-collinear spin structure, spiral spin structures for example. The restriction to a fixed z-direction of the external spin-coupled field allows for a treatment completely in parallel to the

spin-independent case and is therefore used throughout in the basic text. It applies for many cases of magnetic structure.) A coupling constant λ of the pair interaction was introduced being equal to e^2 in case of the Coulomb interaction $w(|\boldsymbol{r}_i - \boldsymbol{r}_j|) = |\boldsymbol{r}_i - \boldsymbol{r}_j|^{-1}$ of particles with charge $\pm e$.

Natural units will be used throughout by putting

$$\hbar = m = \lambda = 1. \tag{1.22}$$

This means for electrons, that energies are given in units of Hartree and lengths in units of the Bohr radius a_{Bohr},

$$1\text{Hartree} = 2\text{Rydberg} = 27.212\text{eV}, \quad 1a_{\text{Bohr}} = 0.52918 \cdot 10^{-10}\text{m}. \tag{1.23}$$

The Hamiltonian (1.21) reduces in these units to

$$\hat{H} = -\frac{1}{2}\sum_{i=1}^{N} \nabla_i^2 + \sum_{i=1}^{N} v(x_i) + \frac{1}{2}\sum_{i\neq j}^{N} w(|\boldsymbol{r}_i - \boldsymbol{r}_j|). \tag{1.24}$$

The formal connection with (1.1–1.4) is expressed as $\hat{H}\,\Psi(x_1 \ldots x_N) = \langle x_1 \ldots x_N|\hat{H}|\Psi\rangle$, where for brevity the same symbol $\hat{H}$ was used for the Schrödinger operator (1.24) on the left-hand side and for the abstract Hilbert space operator on the right-hand side.

1.2 The Momentum Representation, N Fixed

The momentum or plane wave representation formally uses the eigenstates of the particle momentum operator

$$\hat{\boldsymbol{p}}\,|\boldsymbol{k}\rangle = |\boldsymbol{k}\rangle\,\hbar\boldsymbol{k} \tag{1.25}$$

instead of the position vector eigenstates from (1.6) as basis vectors in the Hilbert space of one-particle quantum states.

In Schrödinger representation, the particle momentum operator is given by

$$\hat{\boldsymbol{p}} = \frac{\hbar}{i}\,\nabla, \tag{1.26}$$

and the momentum eigenstate (1.25) is represented by a wavefunction

$$\phi_{\boldsymbol{k}}(\boldsymbol{r}) = \langle \boldsymbol{r}|\boldsymbol{k}\rangle = \frac{1}{\sqrt{V}} e^{i\boldsymbol{k}\cdot\boldsymbol{r}}. \tag{1.27}$$

V is the total volume or the normalization volume.

In order to avoid formal mathematical problems with little physical relevance, the infinite position space $\boldsymbol{R}^3$ of the particles is to be replaced by a large torus $\boldsymbol{T}^3$ of volume $V = L^3$ defined by

$$x + L \equiv x, \quad y + L \equiv y, \quad z + L \equiv z, \tag{1.28}$$

where (x, y, z) are the components of the position vector (periodic or Born von Karman boundary conditions). The meaning of (1.28) is that any function of x, y, z must fulfill the periodicity conditions $f(x + L) = f(x)$, and so on. As is immediately seen from (1.27), this restricts the spectrum of eigenvalues $\boldsymbol{k}$ of (1.25) to the values

$$\boldsymbol{k} = \frac{2\pi}{L}\,(n_x, n_y, n_z) \tag{1.29}$$

with integers n_x, n_y, n_z. The $\boldsymbol{k}$-values (1.29) form a simple cubic mesh in the wavenumber space ($\boldsymbol{k}$-space) with a $\boldsymbol{k}$-space density of states (number of $\boldsymbol{k}$-vectors (1.29) within a unit volume of $\boldsymbol{k}$-space)

$$D(\boldsymbol{k}) = \frac{V}{(2\pi)^3}, \quad \text{i.e.} \quad \sum_{\boldsymbol{k}} \to \frac{V}{(2\pi)^3} \int d^3k. \tag{1.30}$$

Again we introduce a combined variable

$$q \stackrel{\text{def}}{=} (\boldsymbol{k}, s), \quad \sum_q \stackrel{\text{def}}{=} \sum_s \sum_{\boldsymbol{k}} = \frac{V}{(2\pi)^3} \sum_s \int d^3k \tag{1.31}$$

for both momentum and spin of a particle.

In analogy to (1.8), the N-particle quantum state is now represented by a (spinor-)wavefunction in momentum space

$$\Psi(q_1 \ldots q_N) = \langle\, q_1 \ldots q_N \,|\Psi\rangle = \langle\, \boldsymbol{k}_1 s_1 \ldots \boldsymbol{k}_N s_N \,|\Psi\rangle \tag{1.32}$$

expressing the probability amplitude of a particle momentum (and possibly spin) configuration $(q_1 \ldots q_N)$ in complete analogy to (1.20). Everything that was said in the preceding section between (1.6) and (1.20) transfers accordingly to the present representation. Particularly, in the case of fermions, Ψ of (1.32) is totally antisymmetric with respect to permutations of the q_i, and it is totally symmetric in the case of bosons. Its spin dependence is in complete analogy to that of the Schrödinger wavefunction (1.8).

Equation (1.2) reads in momentum representation

$$\sum_{(q_1'\ldots q_N')} \langle\, q_1 \ldots q_N \,|\hat{H}|\, q_1' \ldots q_N' \,\rangle \langle\, q_1' \ldots q_N' \,|\Psi\rangle = \langle\, q_1 \ldots q_N \,|\Psi\rangle\, E, \quad (1.33)$$

where summation runs over the physically distinguished states only. This may be achieved by introducing some linear order in the discrete set $\{\boldsymbol{k}s\}$ with $\boldsymbol{k}$'s from (1.29) and only considering ascending sequences $(q_1 \ldots q_N)$ and $(q_1' \ldots q_N')$, having in mind the (anti-)symmetry of (1.32) with respect to particle exchange.

With this rule, the Hamiltonian (1.24) for a fermionic system is represented by[1]

$$\begin{aligned}\langle\, q_1 \ldots q_N \,|\hat{H}|\, q_1' \ldots q_N' \,\rangle \;=\; & \frac{1}{2}\sum_i \boldsymbol{k}_i^2 \prod_j \delta_{q_j q_j'} + \\ & + \sum_i \delta_{s_{\bar{\imath}} s_i'}\, v^{s_{\bar{\imath}}}_{\boldsymbol{k}_{\bar{\imath}} - \boldsymbol{k}_i'} (-1)^{\bar{\mathcal{P}}} \prod_{j(\neq i)} \delta_{q_{\bar{\jmath}} q_j'} + \\ & + \frac{1}{2}\sum_{i\neq j} \frac{\delta_{\boldsymbol{k}_{\bar{\imath}}+\boldsymbol{k}_{\bar{\jmath}}, \boldsymbol{k}_i'+\boldsymbol{k}_j'}}{V} \Big[\delta_{s_{\bar{\imath}} s_i'} \delta_{s_{\bar{\jmath}} s_j'} w_{|\boldsymbol{k}_{\bar{\imath}} - \boldsymbol{k}_i'|} - \\ & \qquad - \delta_{s_{\bar{\imath}} s_j'} \delta_{s_{\bar{\jmath}} s_i'} w_{|\boldsymbol{k}_{\bar{\imath}} - \boldsymbol{k}_j'|}\Big] (-1)^{\bar{\mathcal{P}}} \prod_{k(\neq i,j)} \delta_{q_{\bar{k}} q_k'}, \quad (1.34)\end{aligned}$$

where $i = \bar{\mathcal{P}}\bar{\imath}$, and $\bar{\mathcal{P}}$ is a permutation of the subscripts which puts $\bar{\imath}$ in the position i (and puts $\bar{\jmath}$ in the position j in the last sum) and leaves the order of the remaining subscripts unchanged. There is always at most one permutation $\bar{\mathcal{P}}$ (up to an irrelevant interchange of $\bar{\imath}$ and $\bar{\jmath}$) for which the product of Kronecker δ's can be non-zero. For the diagonal matrix element, that is $q_i' = q_i$ for all i, $\bar{\mathcal{P}}$ is the identity and can be omitted.

The Fourier transforms of $v(x)$ and $w(r)$ are given by

$$v^s_{\boldsymbol{k}} = \frac{1}{V} \int d^3r\, v(\boldsymbol{r}, s) e^{-i\boldsymbol{k}\cdot\boldsymbol{r}} \quad (1.35)$$

[1] The reader not quite familiar with the present considerations may for instance think of five momenta $\boldsymbol{k}_a$, $\boldsymbol{k}_b, \boldsymbol{k}_c, \boldsymbol{k}_d$, $\boldsymbol{k}_e$ combined with spin up and followed by the same momenta combined with spin down in an ascending sequence of the assumed linear order of states, with $\boldsymbol{k}_a + \boldsymbol{k}_d = \boldsymbol{k}_b + \boldsymbol{k}_e$. As an example with $N = 3$ he may take the case $q_1' = (\boldsymbol{k}_a+)$, $q_2' = (\boldsymbol{k}_c+)$, $q_3' = (\boldsymbol{k}_d-)$ and $q_1 = (\boldsymbol{k}_c+)$, $q_2 = (\boldsymbol{k}_e+)$, $q_3 = (\boldsymbol{k}_b-)$ and then play with other combinations of spins.

and

$$w_{|\boldsymbol{k}|} = \frac{1}{V} \int d^3r\, w(r) e^{-i\boldsymbol{k}\cdot\boldsymbol{r}}, \tag{1.36}$$

where the latter integral may readily be further simplified with spherical coordinates.

The first sum in (1.34) is over the kinetic energies of the particles in their momentum eigenstates (1.25, 1.27). Since the momentum operator of single particles commutes with the kinetic energy operator $\hat{T}$ of the system, this part is diagonal in the momentum representation, which is formally expressed by the j-product over the Kronecker symbols $\delta_{q_j q'_j}$. The next sum contains the individual interaction events of particles with the external potential $v(x)$. The amplitude of this interaction process is given by the Fourier transform of the potential, and is *assumed* to be diagonal in the z-component of the particle spin according to the text after (1.21). Since the interaction of each particle with the external field is furthermore assumed to be independent of the other particles ($\hat{U}$ is assumed to be a sum over individual items $v(x_l)$ in (1.21)), all remaining particle states $j \neq i$ are kept unchanged in an interaction event in which one particle makes a transition from the state q'_i to the state $q_{\bar{\imath}}$. This is again expressed by the j-product. Finally, the classical image of an elementary pair-interaction event is that particles in states $\boldsymbol{k}'_i s'_i$ and $\boldsymbol{k}'_j s'_j$ transfer a momentum $\boldsymbol{k}_{\bar{\imath}} - \boldsymbol{k}'_i$ and keep their spin states unchanged, because the interaction potential w is spin-independent. Hence, $\boldsymbol{k}_{\bar{\imath}} + \boldsymbol{k}_{\bar{\jmath}} = \boldsymbol{k}'_i + \boldsymbol{k}'_j$ and $s_{\bar{\imath}} = s'_i,\ s_{\bar{\jmath}} = s'_j$. Quantum-mechanically, one cannot decide which of the particles, formerly in states $q'_i,\ q'_j$, is afterwards in the state $q_{\bar{\imath}}$ and which one is in the state $q_{\bar{\jmath}}$. This leads to the second term in (1.34), the exchange term with $q_{\bar{\imath}}$ and $q_{\bar{\jmath}}$ reversed. For a bosonic system, the exchange term would appear with a positive sign. The elementary processes corresponding to the terms of (1.34) are depicted in Fig.1.

(As a formal exercise the reader may cast the stationary Schrödinger equation (1.2) into the Schrödinger representation and the Hamiltonian (1.24) into a form analogous to (1.33, 1.34), as an integral operator with a δ-like kernel.)

Both Hamiltonians (1.24) and (1.34) may be considered as deduced from experiment. However, they are also formally equivalent. This is obtained by a derivation encountered frequently, and although it requires some effort it should be performed at least once in life. We consider for example the term $\langle q_1 \dots q_N | w_{mn} | q'_1 \dots q'_N \rangle$, where w_{mn} means the interaction term of the

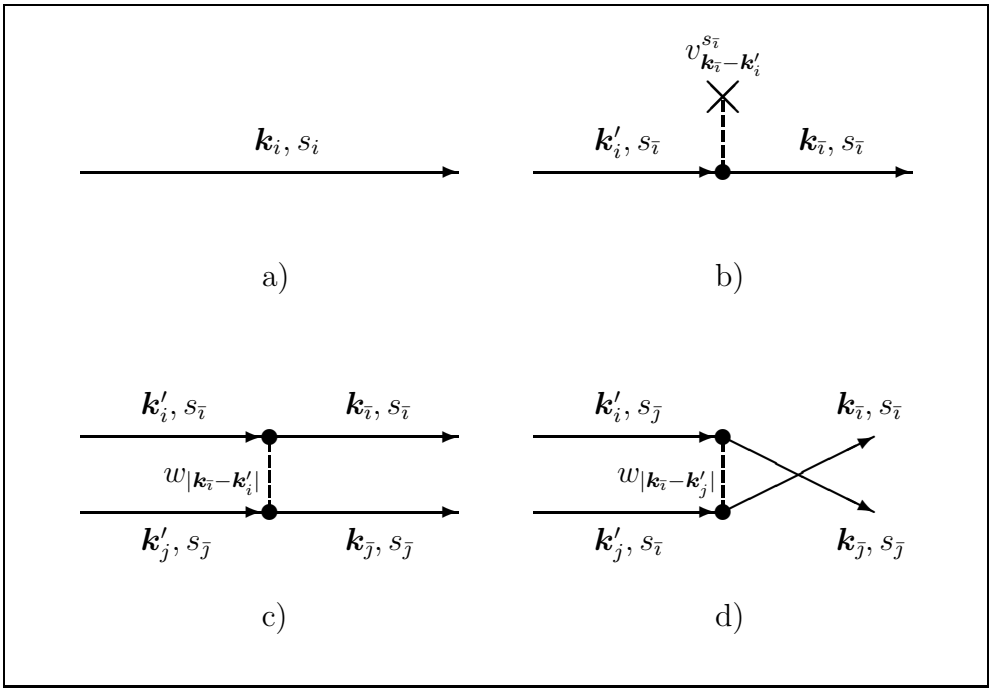

Figure 1: Elementary events corresponding to the terms contained in (1.34):
a) Propagation of a particle with (conserved) momentum $\boldsymbol{k}_i$ and spin s_i, and having kinetic energy $\boldsymbol{k}_i^2/2$.
b) Scattering event of a particle having initially momentum $\boldsymbol{k}'_i$ and (conserved) spin $s_{\bar{\imath}}$, on the potential $v(x)$.
c) Direct interaction event of a pair of particles having initially momenta $\boldsymbol{k}'_i$ and $\boldsymbol{k}'_j$, and transferring a momentum $\boldsymbol{k}_{\bar{\imath}} - \boldsymbol{k}'_i$ from one particle to the other, the spins remaining unchanged.
d) Exchange interaction event of a pair of particles.

Hamiltonian which in Schrödinger representation is equal to $w(|\boldsymbol{r}_m - \boldsymbol{r}_n|)$. To compute it we need the Schrödinger representation of the states $|q_1 \dots q_N\rangle$, which is composed of plane-wave spin-orbitals

$$\phi_q(x) = \phi_{\boldsymbol{k}s}(\boldsymbol{r}\bar{s}) = \frac{1}{\sqrt{V}} e^{i\boldsymbol{k}\cdot\boldsymbol{r}} \chi^s(\bar{s}) = \frac{1}{\sqrt{V}} e^{i\boldsymbol{k}\cdot\boldsymbol{r}} \delta_{\bar{s}s} = \langle \boldsymbol{r}\bar{s} | \boldsymbol{k}s \rangle. \tag{1.37}$$

(We have denoted the spin variable of the x-state by $\bar{s}$ in order to distinguish it from the spin variable s of the q-state. $\chi^s(\bar{s})$ is one of the two spinors (1.11).) According to (1.15),

$$\langle x_1 \dots x_N | q_1 \dots q_N \rangle = \frac{1}{\sqrt{N!}} \det \|\phi_{q_k}(x_l)\| \tag{1.38}$$

and

$$\begin{aligned}
&\langle q_1 \dots q_N | w_{mn} | q'_1 \dots q'_N \rangle = \\
&= \frac{1}{N!} \int dx_1 \dots dx_N \, (\det \|\phi_{q_k}(x_l)\|)^* \, w(|\boldsymbol{r}_m - \boldsymbol{r}_n|) \, \det \|\phi_{q'_k}(x_l)\| \; = \\
&= \frac{1}{N!} \int dx_1 \dots dx_N \; * \\
&\qquad * \sum_{\mathcal{P}} (-1)^{\mathcal{P}} \prod_l \phi^*_{q_{\mathcal{P}l}}(x_l) w(|\boldsymbol{r}_m - \boldsymbol{r}_n|) \sum_{\mathcal{P}'} (-1)^{\mathcal{P}'} \prod_l \phi_{q'_{\mathcal{P}'l}}(x_l).
\end{aligned} \tag{1.39}$$

The two determinants have been expanded into sums of products. $\mathcal{P}$ means a permutation of the subscripts $12 \dots N$ into $\mathcal{P}1\mathcal{P}2 \dots \mathcal{P}N$ with the indicated sign factor being ± 1 for even and odd orders of permutations, respectively. Due to the orthonormality of the plane-wave spin-orbitals (1.37) the integral over dx_l for $l \neq m, n$ is equal to unity if and only if $q_{\mathcal{P}l} = q'_{\mathcal{P}'l}$. With our convention on the sequences $(q_1 \dots q_N)$ and $(q'_1 \dots q'_N)$ introduced after (1.33), this is only possible, if after removing the items with $l = m, n$, the sequences $(\mathcal{P}l)$ and $(\mathcal{P}'l)$ become identical, and furthermore $q'_{\mathcal{P}'l} = q_{\mathcal{P}l}$, $l \neq m, n$. Hence, if the sets $\{q_i\}$ and $\{q'_i\}$ differ in more than two q's, the matrix element is zero. Otherwise, for a given $\mathcal{P}$, only two items of the $\mathcal{P}'$-sum are to be retained: with $\mathcal{P}'l = \bar{\mathcal{P}}\mathcal{P}l$ for $l \neq m, n$ and $\bar{\mathcal{P}}$ as defined in the text below (1.34), and with $\mathcal{P}'m = \bar{\mathcal{P}}\mathcal{P}m \stackrel{\text{def}}{=} i$, $\mathcal{P}'n = \bar{\mathcal{P}}\mathcal{P}n \stackrel{\text{def}}{=} j$ and $\mathcal{P}'m = \bar{\mathcal{P}}\mathcal{P}n = j$, $\mathcal{P}'n = \bar{\mathcal{P}}\mathcal{P}m = i$, respectively. (There are two possible choices to fix $\bar{\mathcal{P}}$, both giving the same result.) With fixed i and j, $(N-2)!$ permutations $\mathcal{P}$ of the remaining integration variables yield identical results. If we finally denote $\mathcal{P}l$ by $\bar{k}$ and $\mathcal{P}'l$ by $k = \bar{\mathcal{P}}\bar{k}$, we are left with

$$\begin{aligned}
&\frac{1}{N(N-1)} \sum_{i \neq j} \int dx_m dx_n \, \phi^*_{q_{\bar{i}}}(x_m) \phi^*_{q_{\bar{j}}}(x_n) \, w(|\boldsymbol{r}_m - \boldsymbol{r}_n|) * \\
&\qquad * (\phi_{q'_i}(x_m)\phi_{q'_j}(x_n) - \phi_{q'_j}(x_m)\phi_{q'_i}(x_n)) \, (-1)^{\bar{\mathcal{P}}} \prod_{k(\neq i,j)} \delta_{q_{\bar{k}} q'_k}.
\end{aligned} \tag{1.40}$$

This result is independent of m and n, because the denotation of integration variables is irrelevant. This fact is a direct consequence of the total symmetry of $\Psi^*(x_1 \dots x_N)\Psi'(x_1 \dots x_N)$ with respect to permutations of subscripts, whereupon each term of the sums of (1.24) yields the same expectation value; in our case, the last sum of (1.24) yields $N(N-1)$ times the result (1.40). The rest is easy: introducing instead of $\boldsymbol{r}_m$, $\boldsymbol{r}_n$ coordinates of the center of

gravity and distance of the pair, the result contained in (1.34, 1.36) is immediately obtained. The single-particle part of (1.34) is obtained along the same lines, but this time both permutations $\mathcal{P}'$ and $\bar{\mathcal{P}}\mathcal{P}$ must be completely identical for a non-zero contribution.

An important reference system is formed by interaction-free fermions ($w = 0$) in a constant external potential $v(x) = 0$: the homogeneous interaction-free fermion gas with the Hamiltonian $\hat{H}_f$:

$$\langle\, q_1 \dots q_N \,|\hat{H}_f|\, q'_1 \dots q'_N \,\rangle = \frac{1}{2} \sum_i \boldsymbol{k}_i^2 \prod_j \delta_{q_j q'_j}. \tag{1.41}$$

Its ground state is a Slater determinant of plane waves so that the sum $\sum_{i=1}^N \boldsymbol{k}_i^2$ is minimum. This is obviously the case if all $\boldsymbol{k}_i$ lie inside a sphere of radius k_f determined by

$$N = \sum_{q_i}^{k_i \le k_f} 1 = 2\, \frac{V}{(2\pi)^3} \int_{k \le k_f} d^3k, \tag{1.42}$$

where the factor 2 in front of the last expression comes from summation over the two spin values for each $\boldsymbol{k}$. Hence

$$\frac{N}{V} = n = \frac{k_f^3}{3\pi^2} \tag{1.43}$$

with n denoting the constant particle density in position space of this ground state, related to the Fermi radius k_f. The Fermi sphere of radius k_f separates in $\boldsymbol{k}$-space the occupied orbitals $\langle \boldsymbol{r} | \boldsymbol{k} \rangle$ from the unoccupied ones. The ground state energy is

$$E = 2 \sum_{k \le k_f} \frac{k^2}{2} = \frac{3}{10}\, k_f^2 N \tag{1.44}$$

implying an average energy per particle

$$\varepsilon = \frac{E}{N} = \frac{3}{10}\, k_f^2. \tag{1.45}$$

Energies are given in natural units (1.23) in both cases.

1.3 The Heisenberg Representation, N Fixed

The above matrix notation (1.33, 1.34) of the momentum representation may by considered as a special case of a more general scheme.

Let $\{|L\rangle\}$ be any given complete orthonormal set of N-particle states labeled by some multi-index L. Any state may then be expanded according to

$$|\Psi\rangle = \sum_L |L\rangle\, C_L = \sum_L |L\rangle\langle L|\Psi\rangle, \tag{1.46}$$

and the stationary Schrödinger equation (1.2) takes on the form of a matrix problem:

$$\sum_{L'} [H_{LL'} - E\delta_{LL'}]\, C_{L'} = 0, \quad H_{LL'} = \langle L|\hat{H}|L'\rangle. \tag{1.47}$$

with an infinite matrix $\|\boldsymbol{H}\|$ and the eigenstate represented by a column vector $\boldsymbol{C}$. This is Heisenberg's matrix mechanics.

To be a bit more specific let $\{\phi_l(x)\}$ be a complete orthonormal set of single-particle (spin-)orbitals, and let $\Phi_L, L = (l_1 \dots l_N)$ run over the N-particle Slater determinants (1.15) of all possible orbital configurations (again using some linear order of the l-labels). In analogy to (1.34–1.36) one now gets

$$\begin{aligned} H_{LL'} &= \sum_i \left\langle l_{\bar{\imath}} \left| -\frac{1}{2}\nabla^2 + v(x) \right| l'_i \right\rangle (-1)^{\bar{\mathcal{P}}} \prod_{j(\neq i)} \delta_{l_{\bar{\jmath}} l'_j} + \\ &\quad + \frac{1}{2} \sum_{i\neq j} \left[\langle l_{\bar{\imath}} l_{\bar{\jmath}} | w | l'_i l'_j \rangle - \langle l_{\bar{\imath}} l_{\bar{\jmath}} | w | l'_j l'_i \rangle \right] (-1)^{\bar{\mathcal{P}}} \prod_{k(\neq i,j)} \delta_{l_{\bar{k}} l'_k} \end{aligned} \tag{1.48}$$

with $i = \bar{\mathcal{P}}\bar{\imath}$, and $\bar{\mathcal{P}}$ is defined in the same way as in (1.34), particularly again $\bar{\mathcal{P}}$ = identity for $L' = L$. The orbital matrix elements are

$$\begin{aligned} \langle l|\hat{h}|m\rangle &\stackrel{\text{def}}{=} \left\langle l \left| -\frac{1}{2}\nabla^2 + v(x) \right| m \right\rangle = \\ &= \sum_s \int d^3r\, \phi_l^*(\boldsymbol{r}, s) \left[-\frac{1}{2}\nabla^2 + v(\boldsymbol{r}, s) \right] \phi_m(\boldsymbol{r}, s) \end{aligned} \tag{1.49}$$

and

$$\begin{aligned} &\langle lm|w|pq\rangle = \\ &\quad \sum_{ss'} \int d^3r d^3r'\, \phi_l^*(\boldsymbol{r}, s)\phi_m^*(\boldsymbol{r}', s')\, w(|\boldsymbol{r} - \boldsymbol{r}'|)\, \phi_q(\boldsymbol{r}', s')\phi_p(\boldsymbol{r}, s). \end{aligned} \tag{1.50}$$

Clearly, $\langle lm|w|pq\rangle = \langle ml|w|qp\rangle$.

1.4 Hartree-Fock Theory

For an interacting N-fermion system, a single Slater determinant (1.15) can of course not be a solution of the stationary Schrödinger equation (1.2). However, referring to (1.3), one can ask for the best Slater determinant approximating the true N-particle ground state as that one which minimizes the expectation value of the Hamiltonian $\hat{H}$ among Slater determinants. The corresponding minimum value will estimate the true ground state energy from above.

This would, however, in general be a too restrictive search. The point is that in most cases of interest the Hamiltonian $\hat{H}$ does not depend on the spins of the particles: $v(\boldsymbol{r}, s) = v(\boldsymbol{r})$. Consequently, the true ground state has a definite total spin $S^2 = \langle \sum_\alpha^{x,y,z} (\sum_i \hat{\sigma}_{i\alpha})^2 \rangle$, whereas a Slater determinant of spin-orbitals in general does not have a definite total spin; rather such a spin eigenstate can be build as a linear combination of Slater determinants with the same spatial orbitals but different single-particle spin states occupied. Depending on whether the total spin of the ground state is zero or non-zero, the approach is called the closed-shell and open-shell Hartree-Fock method, respectively.

We restrict our considerations to the simpler case of closed shells and will see in a minute that in this special case a determinant of spin-orbitals would do. In this case, the number N of spin-half particles must be even because otherwise the total spin would again be half-integer and could not be zero. A spin-zero state of two spin-half particles is obtained as the antisymmetric combination of a spin-up and a spin-down state:

$$\langle s_1 s_2 | S = 0 \rangle = \frac{1}{\sqrt{2}} \left(\chi^+(s_1)\chi^-(s_2) - \chi^-(s_1)\chi^+(s_2) \right). \tag{1.51}$$

This is easily seen by successively operating with $\hat{\sigma}_{1\alpha} + \hat{\sigma}_{2\alpha}, \alpha = x, y, z$ (see (1.10, 1.11)) on it, giving a zero result in all cases. Hence, a simple product of $N/2$ spin pairs in states (1.51) provides a normalized N-particle $S = 0$ spin state, which is antisymmetric with respect to particle exchange within the pair and symmetric with respect to exchange of pairs. (It cannot in general be symmetric or antisymmetric with respect to exchange between different pairs.)

The two particles in the spin state (1.51) may occupy the same spatial orbital $\phi(\boldsymbol{r})$, maintaining the antisymmetric character of the pair wavefunction

$$\Phi(x_1 x_2) = \phi(\boldsymbol{r}_1)\phi(\boldsymbol{r}_2)\langle s_1 s_2 | S = 0 \rangle. \tag{1.52}$$

If, for even N, we consider a Slater determinant of spin orbitals, where each spatial orbital is occupied twice with spin up and down, and expand the determinant into a sum over permutations of products, then a permutation within a doubly occupied pair of states does not change the spatial part of those terms. For the spin part, those permutations just combine to a product of $N/2$ spin-zero states (1.51). The total Slater determinant is thus a linear combination of products of spin-zero states, hence it is itself a spin-zero state in this special case. Moreover, as a Slater determinant it has the correct antisymmetry with respect to all particle exchange operations. Therefore, such a single Slater determinant can provide a spin-zero approximation to a closed-shell ground state.

Now, take (for even N) $N/2$ spatial orbitals ϕ_i and build a Slater determinant with spin-orbitals $\phi_{i+}(x_k)$ in the first $N/2$ rows and with spin-orbitals $\phi_{i-}(x_k)$ in the lower $N/2$ rows. Using the Laplace expansion, this Slater determinant may be written as

$$\Phi_{\mathrm{HF}}(x_1 \dots x_N) = \frac{(N/2)!}{\sqrt{N!}} \sum_{\{k|k'\}} \frac{(-1)^{\{k|k'\}}}{(N/2)!} \det \|\phi_{i+}(x_k)\| \det \|\phi_{i-}(x_{k'})\|, \tag{1.53}$$

where $\{k|k'\}$ means a selection of $N/2$ numbers k among the numbers $1,\dots,N$, the remaining unselected numbers being denoted by k'. There are $N!/(N/2)!^2$ different selections to be summed up with an appropriately chosen sign for each item. The items of the sum are normalized, and they are orthogonal to each other with respect to their spin dependence, because they differ in the selection of the variables of spin-up particles. Hence there are no crossing matrix elements for any spin independent operator, and its expectation value may be calculated just with one of the terms in the sum of (1.53), all terms giving the same result.

With the help of (1.48–1.50) the expectation value of the Hamiltonian (with s-independent external potential v) in the state (1.53) is easily obtained to be

$$E_{\mathrm{HF}} = \langle \Phi_{\mathrm{HF}} | \hat{H} | \Phi_{\mathrm{HF}} \rangle = 2 \sum_{i=1}^{N/2} \langle i | \hat{h} | i \rangle + 2 \sum_{i,j=1}^{N/2} \langle ij | w | ij \rangle - \sum_{i,j=1}^{N/2} \langle ij | w | ji \rangle. \tag{1.54}$$

The three terms are called in turn one-particle energy, Hartree energy, and exchange energy. Summation over both spin directions for each orbital

ϕ_i results in factors 2 for the one-particle term, 4 for the Hartree term, but only 2 for the exchange term because the contained matrix element is only nonzero if both interacting particles have the same spin direction. (Recall that the interaction part of the Hamiltonian comes with a prefactor 1/2.) Note that both the Hartree and exchange terms for $i = j$ contain (seemingly erroneously) the self-interaction of a particle in the orbital ϕ_i with itself. Actually those terms of the Hartree and exchange parts mutually cancel in (1.54) thus not posing any problem. (There is of course one term with $i = j$ but opposite spin directions remaining correctly in the Hartree part.)

In order to find the minimum of this type of expression one must vary the orbitals keeping them orthonormal. However, as we already know, the determinant remains unchanged upon an orthogonalization of the orbitals, and hence it suffices to keep the orbitals normalized while varying them. Adding the normalization integral for ϕ_k, multiplied with a Lagrange multiplier $2\varepsilon_k$, to (1.54) and then varying ϕ_k^* leads to the minimum condition

$$\hat{h}\,\phi_k(\boldsymbol{r}) + v_{\mathrm{H}}(\boldsymbol{r})\,\phi_k(\boldsymbol{r}) + (\hat{v}_{\mathrm{X}}\,\phi_k)(\boldsymbol{r}) = \phi_k(\boldsymbol{r})\varepsilon_k \tag{1.55}$$

with the Hartree potential

$$v_{\mathrm{H}}(\boldsymbol{r}) = 2\sum_{j=1}^{N/2}\int d^3r'\,\phi_j^*(\boldsymbol{r}')\,w(|\boldsymbol{r}-\boldsymbol{r}'|)\,\phi_j(\boldsymbol{r}') \tag{1.56}$$

and the exchange potential operator

$$(\hat{v}_{\mathrm{X}}\,\phi_k)(\boldsymbol{r}) = -\sum_{j=1}^{N/2}\int d^3r'\,\phi_j^*(\boldsymbol{r}')\,w(|\boldsymbol{r}-\boldsymbol{r}'|)\,\phi_k(\boldsymbol{r}')\phi_j(\boldsymbol{r}). \tag{1.57}$$

(ϕ_k^* is varied independently of ϕ_k which is equivalent to independently varying the real and imaginary parts of ϕ_k; the variation then is carried out by using the simple rule $\delta/\delta\phi_k^*(x)\int dx'\,\phi_k^*(x')F(x') = F(x)$ for any expression $F(x)$ independent of $\phi_k^*(x)$.)

The Hartree-Fock equations (1.55) have the form of effective single-particle Schrödinger equations

$$\hat{F}\,\phi_k = \phi_k\,\varepsilon_k, \tag{1.58}$$

where the Fock operator $\hat{F} = -(1/2)\nabla^2 + \hat{v}_{\mathrm{eff}}$ consists of the kinetic energy operator and an effective potential operator

$$\hat{v}_{\mathrm{eff}} = v + v_{\mathrm{H}} + \hat{v}_{\mathrm{X}} \tag{1.59}$$

called the mean field or molecular field operator.

For a given set of $N/2$ occupied orbitals ϕ_i the Fock operator $\hat{F}$ as an integral operator is the same for all orbitals. Hence, from (1.58), the Hartree-Fock orbitals may be obtained orthogonal to each other. From (1.55) it then follows that

$$\sum_{i=1}^{N/2} \varepsilon_i = \sum_{i=1}^{N/2} \langle i|\hat{h}|i\rangle + 2 \sum_{i,j=1}^{N/2} \langle ij|w|ij\rangle - \sum_{i,j=1}^{N/2} \langle ij|w|ji\rangle. \tag{1.60}$$

Comparison with (1.54) yields

$$E_{\mathrm{HF}} = \sum_{i=1}^{N/2} (\varepsilon_i + \langle i|\hat{h}|i\rangle) = 2 \sum_{i=1}^{N/2} \varepsilon_i - \langle \hat{W}\rangle \tag{1.61}$$

for the total Hartree-Fock energy. The sum over all occupied ε_i (including the spin sum) double-counts the interaction energy.

Coming back to the expression (1.54), one can ask for its change, if one removes one particle in the Hartree-Fock orbital ϕ_k (of one spin direction) *while keeping all orbitals ϕ_j unrelaxed.* This change is easily obtained to be $-\langle k|\hat{h}|k\rangle - 2\sum_j \langle kj|w|kj\rangle + \sum_j \langle kj|w|jk\rangle$, which is just $-\varepsilon_k$ as seen from (1.55–1.57). For a given set of occupied ϕ_i, (1.58) yields also unoccupied orbitals as solutions. The change of (1.54), if one additionally occupies one of those latter orbitals ϕ_k, is analogously found to be $+\varepsilon_k$. This result, which may be written as

$$\left(\frac{\partial E_{\mathrm{HF}}}{\partial n_k}\right)_{\phi_j} = \varepsilon_k \tag{1.62}$$

with n_k denoting the occupation number of the Hartree-Fock orbital ϕ_k and the subscript ϕ_j indicating the constancy of the orbitals, goes under the name Koopmans' theorem [Koopmans, 1934]. It guarantees in most cases that the minimum of E_{HF} is obtained if one occupies the orbitals with the lowest ε_i, because removing a particle from ϕ_i and occupying instead a state ϕ_j yields a change of E_{HF} equal to $\varepsilon_j - \varepsilon_i$ plus the orbital relaxation energy, which is usually smaller than $\varepsilon_j - \varepsilon_i$ in closed shell situations. There may otherwise, however, be situations where levels ε_i cross each other when the orbitals are allowed to relax after a re-occupation. If this happens for the highest occupied and lowest unoccupied level (called HOMO and LUMO, where MO stands for molecular orbital), then a convergence problem may appear in the solution process of the closed shell Hartree-Fock equations.

Note that the similarity of the Hartree-Fock equations (1.55) with a single-electron Schrödinger equation is rather formal and not very far-reaching: the Fock operator is not a linear operator in contrast to any Hamiltonian. It depends on the $N/2$ Hartree-Fock orbitals ϕ_i, lowest in energy, which appear as solutions to the Hartree-Fock equations. Thus, those equations are highly nonlinear and must be solved iteratively by starting with some guessed effective potential, solving (linearly) for the ϕ_i, recalculating the effective potential, and iterating until self-consistency is reached. The molecular field is therefore also called the self-consistent field.

More details can for instance be found in [Slater, 1960, Chapters 12ff].

1.5 The Occupation Number Representation, N Varying

Up to here we considered representations of quantum mechanics with the particle number N of the system fixed. If this number is macroscopically large, it cannot be fixed at a single definite number in experiment. Zero mass bosons as e.g. photons may be emitted or absorbed in systems of any scale. (In a relativistic description any particle may be created or annihilated, possibly together with its antiparticle, in a vacuum region just by applying energy.) From a mere technical point of view, quantum statistics of identical particles is much simpler to formulate with the grand canonical ensemble with varying particle number, than with the canonical one. Hence there are many good reasons to consider quantum dynamics with changes in particle number.

In order to do so, we start with building the Hilbert space of quantum states of this wider frame. The considered up to now Hilbert space of all N-particle states having the appropriate symmetry with respect to particle exchange will be denoted by $\mathcal{H}_N$. In Section 1.3 an orthonormal basis $\{|L\rangle\}$ of (anti-)symmetrized products of single-particle states out of a given fixed complete and orthonormalized set $\{\phi_i\}$ of such single-particle states was introduced. The set $\{\phi_i\}$ with some fixed linear order $(\phi_1, \phi_2, \ldots)$ of the orbitals will play a central role in the present section. The states $|L\rangle$ will alternatively be denoted by

$$|n_1 \ldots n_i \ldots\rangle, \quad \sum_i n_i = N, \tag{1.63}$$

where n_i denotes the occupation number of the i-th single-particle orbital in the given state. For fermions, $n_i = 0, 1$, for bosons $n_i = 0, 1, 2, \ldots$. Two

states (1.63) not coinciding in all occupation numbers n_i are orthogonal. $\mathcal{H}_N$ is the complete linear space spanned by the basis vectors (1.63), i.e. the states of $\mathcal{H}_N$ are either linear combinations $\sum |L\rangle C_L$ of states (1.63) or Cauchy sequences of such linear combinations. (A Cauchy sequence is a sequence $\{|\Psi_n\rangle\}$ with $\lim_{m,n\to\infty} \langle \Psi_m - \Psi_n | \Psi_m - \Psi_n \rangle = 0$. The inclusion of such sequences into $\mathcal{H}_N$ means realizing the completeness property of the Hilbert space, being extremely important in all considerations of limits; cf. Section 5.2. This completeness is not to be confused with the completeness of a basis set $\{\phi_i\}$.) The extended Hilbert space $\mathcal{F}$ (Fock space) of all states with the particle number N not fixed is now defined as the completed direct sum of all $\mathcal{H}_N$. It is spanned by all state vectors (1.63) for *all* N with the above given definition of orthogonality retained, and is completed by corresponding Cauchy sequences. Its vectors are given by all *series* $\sum |L\rangle C_L$ with $|L\rangle$ running over all states (1.63) for all N, for which the series $\sum C_L^2$ converges to a finite number. (A mathematical rigorous treatment can for instance be found in [Cook, 1953, Berezin, 1965].)

Note that now $\mathcal{F}$ contains not only quantum states which are linear combinations with varying n_i so that n_i does not have a definite value in the quantum state (occupation number fluctuations), but also linear combinations with varying N so that now quantum fluctuations of the total particle number are allowed too. For bosonic fields (as e.g. laser light) those quantum fluctuations can become important experimentally even for macroscopic N.

In order to introduce the possibility of a dynamical change of N, operators must be introduced providing such a change. For bosons those operators are introduced as

$$\hat{b}_i |\ldots n_i \ldots\rangle = |\ldots n_i - 1 \ldots\rangle \sqrt{n_i}, \tag{1.64}$$

$$\hat{b}_i^\dagger |\ldots n_i \ldots\rangle = |\ldots n_i + 1 \ldots\rangle \sqrt{n_i + 1}. \tag{1.65}$$

These operators annihilate and create, respectively, a particle in the orbital ϕ_i and multiply by a factor chosen for the sake of convenience. Particularly, in (1.64) it prevents producing states with negative occupation numbers. (Recall that the n_i are integers; application of $\hat{b}_i$ to a state with $n_i = 0$ gives zero instead of a state with $n_i = -1$.) Considering all possible matrix elements with the basis states (1.63) of $\mathcal{F}$, one easily proves that $\hat{b}$ and $\hat{b}^\dagger$ are Hermitian conjugate to each other. In the same way the key relations

$$\hat{n}_i |\ldots n_i \ldots\rangle \stackrel{\text{def}}{=} \hat{b}_i^\dagger \hat{b}_i |\ldots n_i \ldots\rangle = |\ldots n_i \ldots\rangle\, n_i, \tag{1.66}$$

and

$$[\hat{b}_i, \hat{b}_j^\dagger] = \delta_{ij}, \quad [\hat{b}_i, \hat{b}_j] = 0 = [\hat{b}_i^\dagger, \hat{b}_j^\dagger] \tag{1.67}$$

are proven, where the brackets in standard manner denote the commutator $[\hat{b}_i, \hat{b}_j^\dagger] \stackrel{\text{def}}{=} [\hat{b}_i, \hat{b}_j^\dagger]_- = \hat{b}_i\hat{b}_j^\dagger - \hat{b}_j^\dagger\hat{b}_i$. The occupation number operator $\hat{n}_i$ is Hermitian and can be used to define the particle number operator

$$\hat{N} = \sum_i \hat{n}_i \tag{1.68}$$

having arbitrarily large but *always finite* expectation values in the basis states (1.63) of the Fock space $\mathcal{F}$. The Fock space itself is the complete hull (in the above described sense) of the linear space spanned by all possible states obtained from the vacuum state

$$|\rangle \stackrel{\text{def}}{=} |0\ldots0\ldots\rangle, \quad \hat{b}_i|\rangle = 0 \text{ for all } i \tag{1.69}$$

by applying polynomials of the $\hat{b}_i^\dagger$ to it. This situation is expressed by saying that the vacuum state is a cyclic vector of $\mathcal{F}$ with respect to the algebra of the $\hat{b}_i$ and $\hat{b}_i^\dagger$. Obviously, any operator in $\mathcal{F}$, that is any operation transforming vectors of $\mathcal{F}$ linearly into new ones, can be expressed as a power series of operators $\hat{b}_i^\dagger$ and $\hat{b}_i$. This all together means that the Fock space provides an irreducible representation space for the algebra of operators $\hat{b}_i^\dagger$ and $\hat{b}_i$, defined by (1.67).

For fermions, the definition of creation and annihilation operators must have regard for the antisymmetry of the quantum states and for Pauli's exclusion principle following from this antisymmetry. They are defined by

$$\hat{c}_i|\ldots n_i\ldots\rangle = |\ldots n_i - 1\ldots\rangle\, n_i\, (-1)^{\sum_{j<i} n_j}, \tag{1.70}$$

$$\hat{c}_i^\dagger|\ldots n_i\ldots\rangle = |\ldots n_i + 1\ldots\rangle\, (1 - n_i)\, (-1)^{\sum_{j<i} n_j}. \tag{1.71}$$

Again by considering the matrix elements with all possible occupation number eigenstates (1.63), it is easily seen that these operators have all the needed properties, do particularly not create non-fermionic states (that is, states with occupation numbers n_i different from 0 or 1 do not appear: application of $\hat{c}_i$ to a state with $n_i = 0$ gives zero, and application of $\hat{c}_i^\dagger$ to a state with $n_i = 1$ gives zero as well). The $\hat{c}_i$ and $\hat{c}_i^\dagger$ are mutually Hermitian conjugate, obey the key relations

$$\hat{n}_i|\ldots n_i\ldots\rangle \stackrel{\text{def}}{=} \hat{c}_i^\dagger\hat{c}_i|\ldots n_i\ldots\rangle = |\ldots n_i\ldots\rangle\, n_i \tag{1.72}$$

and

$$[\hat{c}_i, \hat{c}_j^\dagger]_+ = \delta_{ij}, \quad [\hat{c}_i, \hat{c}_j]_+ = 0 = [\hat{c}_i^\dagger, \hat{c}_j^\dagger]_+ \tag{1.73}$$

with the anticommutator $[\hat{c}_i, \hat{c}_j^\dagger]_+ = \hat{c}_i\hat{c}_j^\dagger + \hat{c}_j^\dagger\hat{c}_i$ defined in standard way. Their role in the fermionic Fock space $\mathcal{F}$ is completely analogous to the bosonic case. (The $\hat{c}^\dagger$- and $\hat{c}$-operators of the fermionic case form a normed complete algebra provided with a norm-conserving adjugation †, called a c^*-algebra in mathematics. Such a (normed) c^*-algebra can be formed out of the bosonic operators $\hat{b}^\dagger$ and $\hat{b}$, which themselves are not bounded in $\mathcal{F}$ and hence have no norm, by complex exponentiation.)

As an example, the Hamiltonian (1.24) is expressed in terms of creation and annihilation operators and orbital matrix elements (1.49, 1.50) as

$$\hat{H} = \sum_{ij} \hat{c}_i^\dagger \langle i|\hat{h}|j\rangle \hat{c}_j + \frac{1}{2} \sum_{ijkl} \hat{c}_i^\dagger \hat{c}_j^\dagger \langle ij|w|kl\rangle \hat{c}_l \hat{c}_k \tag{1.74}$$

Observe the order of operators being important in expressions of that type. This form is easily verified by considering the matrix element $\langle L|\hat{H}|L'\rangle$ with $|L\rangle$ and $|L'\rangle$ represented in notation (1.63), and comparing the result with (1.48).

In order to write down some useful relations holding accordingly in both the bosonic and fermionic cases, we use operator notations $\hat{a}_i$ and $\hat{a}_i^\dagger$ denoting either a bosonic or a fermionic operator. One easily obtains

$$[\hat{n}_i, \hat{a}_i] = -\hat{a}_i, \quad [\hat{n}_i, \hat{a}_i^\dagger] = \hat{a}_i^\dagger \tag{1.75}$$

with the *commutator* in both the bosonic and fermionic cases.

Sometimes it is useful (or simply hard to be avoided) to use a non-orthogonal basis $\{\phi_i\}$ of single-particle orbitals. The whole apparatus may be generalized to this case by merely generalizing the first relations (1.67) and (1.73) to

$$[\hat{a}_i, \hat{a}_j^\dagger]_\pm = \langle \phi_i|\phi_j\rangle, \tag{1.76}$$

which generalization of course comprises the previous relations of the orthogonal cases. Even with a non-orthogonal basis $\{\phi_i\}$ the form of the original relations (1.67) and (1.73) may be retained, if one defines the operators $\hat{a}_i$ with respect to the ϕ_i and replaces the operators $\hat{a}_i^\dagger$ by modified creation operators $\hat{a}_i^+$ with respect to a contragredient basis $\{\chi_i\}$, $\langle \phi_i|\chi_j\rangle = \delta_{ij}$. Of course, this way the $\hat{a}_i^+$ are no longer Hermitian conjugate to the $\hat{a}_i$.

1.6 Field Quantization

Finally, a spatial representation may be introduced by defining field operators

$$\hat{\psi}(x) = \sum_i \phi_i(x)\hat{a}_i, \quad \hat{\psi}^\dagger(x) = \sum_i \phi_i^*(x)\hat{a}_i^\dagger \tag{1.77}$$

providing a spatial particle density operator

$$\hat{n}(x) = \hat{\psi}^\dagger(x)\hat{\psi}(x) \tag{1.78}$$

and obeying the relations

$$\begin{aligned} [\hat{\psi}(x), \hat{\psi}^\dagger(x')]_\pm &= \delta(x - x'), \\ [\hat{\psi}(x),\ \hat{\psi}(x')]_\pm &= 0 \ = \ [\hat{\psi}^\dagger(x), \hat{\psi}^\dagger(x')]_\pm. \end{aligned} \tag{1.79}$$

These relations are readily obtained from those of the creation and annihilation operators, and by taking into account the completeness relation

$$\sum_i \phi_i(x)\phi_i^*(x') = \delta(x - x') \tag{1.80}$$

of the basis orbitals.

In terms of field operators, the Hamiltonian (1.24) reads

$$\begin{aligned} \hat{H} \ = \ & \int dx\, \hat{\psi}^\dagger(x) \left[-\frac{1}{2}\nabla^2 + v(x) \right] \hat{\psi}(x) + \\ & + \frac{1}{2} \int dx dx'\, \hat{\psi}^\dagger(x)\hat{\psi}^\dagger(x')\, w(|\boldsymbol{r} - \boldsymbol{r}'|)\, \hat{\psi}(x')\hat{\psi}(x), \end{aligned} \tag{1.81}$$

which is easily obtained by combining (1.74) and (1.77).

Field-quantized interaction terms contain higher-order than quadratic expressions in the field operators and hence yield operator forms of equations of motion (in Heisenberg picture) which are nonlinear. Note, however, that, contrary to the Fock operator of (1.55), the Hamiltonians (1.74, 1.81) are *linear* operators in the Fock space of states $|\Psi\rangle$ as demanded by the superposition principle of quantum theory. In this respect, the Fock operator rather compares to those operator equations of motion than to a Hamiltonian. (See also the comment at the end of Section 1.4.)

We do not dive here into the subtle mathematical problems connected with the thermodynamic limit $N \to \infty$, $N/V = \text{const.}$ (see, e.g. [Sewell, 1986] and references given therein). In the simplest possible case

of noninteracting with each other fermions, where the ground state may be a single determinant, this ground state may be redefined as a new vacuum state $|0\rangle$ on which the former creation operators $\hat{c}_i^\dagger$ of *occupied* states i now act as annihilation operators $\hat{h}_i = \hat{c}_i^\dagger$ (yielding as previously the result zero), and the former annihilation operators of those states act now as *hole state* creation operators $\hat{h}_i^\dagger = \hat{c}_i$. For unoccupied states the previous operators are retained. This redefinition, if done in the Hamiltonian also, puts the ground state energy to zero and introduces instead of the former fermions two types of particle excitations: former particles and holes. Both types of excitation have now positive excitation energies equal to the absolute value of the energy distance of the former states from the Fermi level. This sign reversion for hole state energies in the single-particle part of the Hamiltonian (1.74) comes about by bringing the operator product $\hat{h}_i\hat{h}_i^\dagger$ appearing after the redefinition back to the normal order $\hat{h}_i^\dagger\hat{h}_i$ and observing the anticommutation rule. By this elementary renormalization, for zero temperature the Fock space is saved in the thermodynamic limit. Recall that by definition of the Fock space, every state vector of the Fock space can be norm-approximated arbitrarily close by a state vector which holds a finite number of particles. For finite temperatures or for interacting particles much more complex constructs are necessary for the thermodynamic limit. These remarks conclude our short survey over the representations of quantum mechanics.

2 Density Matrices and Density Operators

As was repeatedly seen in the last chapter, for computing most of the usual matrix elements only a few of the many coordinates of an N-particle wavefunction are used, where due to the symmetry properties it is irrelevant which ones. Density matrices are a tool to extract the relevant information out of such monstrous constructs as are N-particle wavefunctions. (For reviews, see [Erdahl and Smith, 1987, Coleman, 1963, McWeeny, 1960]; in the text below we try to keep the notation as canonical as possible.)

In this chapter, we consider exclusively *reduced* density matrices related to an N-particle wavefunction. In a wider sense of quantum states, N-particle wavefunction states are called pure states as distinct from ensemble states described by *N-particle* density matrices. The latter are considered in Sections 4.5 and 4.6.

The first four sections of the present chapter deal directly with reduced density matrices to the extent to which it is needed in the context of our program. The last three sections of this chapter also contain material indispensable in this context, and they are placed in the present chapter because they are in one or the other way connected to the notion of density matrices or densities, and logically they must precede the following chapters.

2.1 Single-Particle Density Matrices

The spin-dependent single-particle density matrix of a state $|\Psi\rangle$ is defined as (recall that $\Psi\Psi^*$ is always symmetric in its arguments)

$$\gamma_1(x;x') = N \int dx_2 \dots dx_N \, \Psi(xx_2 \dots x_N)\Psi^*(x'x_2 \dots x_N). \tag{2.1}$$

Its spin-independent version is

$$\gamma_1(\boldsymbol{r};\boldsymbol{r}') = \sum_s \gamma_1(\boldsymbol{r}s;\boldsymbol{r}'s). \tag{2.2}$$

The pre-factor N, the particle number, was chosen in order that the diagonal of the latter density matrix gives the spatial density of particles, $n(\boldsymbol{r})$. By

definition, $n(\boldsymbol{r})$ is the probability density of measuring one of the particle coordinates $\boldsymbol{r}_1 \ldots \boldsymbol{r}_N$ at point $\boldsymbol{r}$. This is just N times the probability density to measure the coordinate $\boldsymbol{r}_1$ at $\boldsymbol{r}$, hence

$$n(\boldsymbol{r}) = \gamma_1(\boldsymbol{r};\boldsymbol{r}), \quad \operatorname{tr}\gamma_1 = \int dx\, \gamma_1(x;x) = \int d^3r\, \gamma_1(\boldsymbol{r};\boldsymbol{r}) = N. \tag{2.3}$$

Treating the arguments of γ_1 as (continuous) matrix indices, the trace tr of the matrix is to be understood as the integral over its diagonal.

The spatial diagonal of the spin-dependent single-particle density matrix is the spin-density matrix

$$n_{ss'}(\boldsymbol{r}) = \gamma_1(\boldsymbol{r}s;\boldsymbol{r}s') \tag{2.4}$$

containing the information on the direction and spatial density of spin polarization (cf. Section 2.4).

With a fixed basis of single-particle orbitals $\{\phi_i\}$, the Heisenberg representation of these density matrices is

$$\langle i|\gamma_1|j\rangle = \int dx dx'\, \phi_i^*(x)\gamma_1(x;x')\phi_j(x') \tag{2.5}$$

for spin-orbitals $\phi_i(x)$ and

$$\langle i|\gamma_1|j\rangle = \int d^3r d^3r'\, \phi_i^*(\boldsymbol{r})\gamma_1(\boldsymbol{r};\boldsymbol{r}')\phi_j(\boldsymbol{r}') \tag{2.6}$$

for spatial orbitals $\phi_i(\boldsymbol{r})$. Note that the index sets for both cases are different: In (2.5) the subscript i distinguishes spin-orbitals, which may have the same spatial parts but different spin parts of the orbitals, in (2.6) the equally denoted subscript only differs between spatial orbitals. The trace over the spin variables has already been performed in (2.6) according to (2.2). Expressions like $\langle i|\gamma_1|j\rangle$ are to be understood always in the context of the orbital sets considered. Due to this difference the first expression is a spin-dependent density matrix whereas the second is a spatial density matrix. Their diagonals yield the occupation numbers of spin-orbitals and spatial orbitals, respectively.

If Ψ of (2.1) is an (anti-)symmetrized product of those orbitals, these density matrices (2.5) and (2.6) are diagonal with integer diagonal elements, restricted to 1 or 0 in the case of fermions. In general the eigenvalues of the single-particle density matrix of fermions are real numbers (because, as is

easily seen, the density matrix is Hermitian) between 0 and 1. Furthermore, for any (anti-)symmetrized product Ψ of single-particle orbitals $\phi_i(x)$,

$$\gamma_1^2 = \int dx''\, \gamma_1(x;x'')\gamma_1(x'';x') = \sum_k \gamma_1|k\rangle\langle k|\gamma_1 = \gamma_1, \tag{2.7}$$

and this property is decisive for a state to be an (anti-)symmetrized product of single-particle orbitals. These latter statements are easily obtained with the help of the definition (2.1), yielding

$$\gamma_1(x;x') = \sum_{i=1}^{N} \phi_i(x)\phi_i^*(x') \tag{2.8}$$

in the considered case.

As an example, the density matrix in momentum representation

$$\langle \boldsymbol{k}|\gamma_1|\boldsymbol{k}'\rangle = \delta_{\boldsymbol{k}\boldsymbol{k}'}\, 2\Theta(k_f - |\boldsymbol{k}|) \tag{2.9}$$

corresponding to the ground state of the Hamiltonian $\hat{H}_f$ of (1.41)—the homogeneous interaction-free fermion gas—is considered. It is diagonal because the particles occupy $\boldsymbol{k}$-eigenstates, one per spin for $\boldsymbol{k}$-vectors inside the Fermi sphere. Fourier back-transformation yields it in spatial representation:

$$\begin{aligned}
\gamma_1(\boldsymbol{r};\boldsymbol{r}') &= \frac{1}{V}\sum_{\boldsymbol{k}\boldsymbol{k}'} e^{i\boldsymbol{k}\cdot\boldsymbol{r}}\langle \boldsymbol{k}|\gamma_1|\boldsymbol{k}'\rangle e^{-i\boldsymbol{k}'\cdot\boldsymbol{r}'} = \\
&= \frac{2}{8\pi^3}\int_{k<k_f} d^3k\, e^{i\boldsymbol{k}\cdot(\boldsymbol{r}-\boldsymbol{r}')} = \\
&= \frac{k_f^3}{\pi^2}\frac{\sin(k_f|\boldsymbol{r}-\boldsymbol{r}'|) - (k_f|\boldsymbol{r}-\boldsymbol{r}'|)\cos(k_f|\boldsymbol{r}-\boldsymbol{r}'|)}{(k_f|\boldsymbol{r}-\boldsymbol{r}'|)^3}.
\end{aligned} \tag{2.10}$$

In agreement with (1.43) one finds for $\boldsymbol{r}' \to \boldsymbol{r}$

$$\gamma_1(\boldsymbol{r};\boldsymbol{r}) = \frac{k_f^3}{3\pi^2}, \tag{2.11}$$

the previous connection between the Fermi radius and the particle density.

In the language of the algebra of representations, the first line of (2.10) may be understood as $\gamma_1(\boldsymbol{r};\boldsymbol{r}') = \langle \boldsymbol{r}|\gamma_1|\boldsymbol{r}'\rangle = \sum_{\boldsymbol{k}\boldsymbol{k}'}\langle \boldsymbol{r}|\boldsymbol{k}\rangle\langle \boldsymbol{k}|\gamma_1|\boldsymbol{k}'\rangle\langle \boldsymbol{k}'|\boldsymbol{r}'\rangle$ on the basis of completeness of the $\boldsymbol{k}$-states (1.27).

2.2 Two-Particle Density Matrices

The spin-dependent two-particle density matrix is defined as (some authors omit the factor 2! in the denominator; our definitions pursue the idea that $\gamma_N = \Psi_N \Psi_N^*$ for any closed N-particle system)

$$\begin{aligned} &\gamma_2(x_1 x_2; x_1' x_2') = \\ &= \frac{N(N-1)}{2!} \int dx_3 \ldots dx_N \, \Psi(x_1 x_2 x_3 \ldots x_N) \Psi^*(x_1' x_2' x_3 \ldots x_N). \end{aligned} \tag{2.12}$$

It is related to the single-particle density matrix by the integral

$$\gamma_1(x; x') = \frac{2}{N-1} \int dx_2 \, \gamma_2(x x_2; x' x_2). \tag{2.13}$$

The spin-independent relations are obtained accordingly by taking the trace over spin variables. Analogously to the single-particle case, the diagonal of the two-particle density matrix gives the pair density $n_2(x_1, x_2)$, i.e. the probability density to find one particle at x_1 and another at x_2. This is the probability density to measure one of the particle coordinates at x_1 and a second one at x_2, being just $N(N-1)$ times the probability density that the original coordinates x_1 and x_2 of the wavefunction are measured at those points:

$$n_2(x_1, x_2) = 2 \, \gamma_2(x_1 x_2; x_1 x_2). \tag{2.14}$$

Matrices with respect to single-particle orbital sets like (2.5, 2.6) may be defined in an analogous manner.

Consider as an example the two-particle density matrix of the determinant state (1.15) of non-interacting fermions with the first N spin-orbitals

occupied ($l_i = i$). It is explicitly given by

$$\begin{aligned}
&\gamma_2(x_1x_2;x_1'x_2') = \\
&= \frac{N(N-1)}{2}\int dx_3\dots dx_N * \\
&\qquad *\frac{1}{N!}\sum_{\mathcal{P}}(-1)^{\mathcal{P}}\left[\prod_i \phi_{\mathcal{P}i}(x_i)\right]\sum_{\mathcal{P}'}(-1)^{\mathcal{P}'}\left[\prod_i \phi^*_{\mathcal{P}'i}(x_i')\right] = \\
&= \frac{N(N-1)}{2N!}\sum_{ij=1}^{N}\big[\phi_i(x_1)\phi_j(x_2)\phi_i^*(x_1')\phi_j^*(x_2') - \\
&\qquad -\phi_i(x_1)\phi_j(x_2)\phi_j^*(x_1')\phi_i^*(x_2')\big]\,(N-2)! = \\
&= \frac{1}{2}\sum_{ij=1}^{N}\left[\phi_i(x_1)\phi_j(x_2)\phi_i^*(x_1')\phi_j^*(x_2') - \phi_i(x_1)\phi_j(x_2)\phi_j^*(x_1')\phi_i^*(x_2')\right].
\end{aligned} \tag{2.15}$$

This follows from just a somewhat simplified variant of the algebra of (1.39, 1.40). In the first expression the two determinants of spin-orbitals are expanded. $\mathcal{P}$ means a permutation of the subscripts $12\dots N$ into $\mathcal{P}1\mathcal{P}2\dots\mathcal{P}N$ with the indicated sign factor being ± 1 for even and odd orders of permutations, respectively. For $i > 2$, $x_i' = x_i$. Due to the orthonormality of the spin-orbitals the integral is equal to unity if and only if $\mathcal{P}i = \mathcal{P}'i$ for $i = 3\dots N$ ($(N-2)!$ cases). In each of those cases either $\mathcal{P}1 = \mathcal{P}'1, \mathcal{P}2 = \mathcal{P}'2$ or $\mathcal{P}1 = \mathcal{P}'2, \mathcal{P}2 = \mathcal{P}'1$, $\mathcal{P}$ and $\mathcal{P}'$ having the same order in the first possibility and differing by one order in the second. Summation over all $\mathcal{P}1 = i, \mathcal{P}2 = j$ yields the final result. (Although, of course, $\mathcal{P}1 \neq \mathcal{P}2$, retaining the $i = j$ terms in the double sum is harmless, because the corresponding square bracket expressions are zero. See the corresponding discussion after (1.54).)

Suppose now that $N/2$ spatial orbitals are occupied for both spin directions. From the last expression it immediately follows that

$$\begin{aligned}
&2\gamma_2(\boldsymbol{r}_1\boldsymbol{r}_2;\boldsymbol{r}_1'\boldsymbol{r}_2') = \\
&= \sum_{ij=1}^{N/2}\left[4\phi_i(\boldsymbol{r}_1)\phi_j(\boldsymbol{r}_2)\phi_i^*(\boldsymbol{r}_1')\phi_j^*(\boldsymbol{r}_2') - 2\phi_i(\boldsymbol{r}_1)\phi_j(\boldsymbol{r}_2)\phi_j^*(\boldsymbol{r}_1')\phi_i^*(\boldsymbol{r}_2')\right].
\end{aligned} \tag{2.16}$$

According to the rules, summation was taken over $s_1' = s_1$ and $s_2' = s_2$ giving a factor 4 in the first term, but only a factor 2 in the second because this is

nonzero only if the spins of $\phi_i(x)$ and $\phi_j(x)$ were equal. Considering

$$2\sum_{i=1}^{N/2} \phi_i(\boldsymbol{r})\phi_i^*(\boldsymbol{r}) = n(\boldsymbol{r}) \tag{2.17}$$

and a spin-summed variant of (2.14) yields

$$n_2(\boldsymbol{r}_1, \boldsymbol{r}_2) = n(\boldsymbol{r}_1)n(\boldsymbol{r}_2) - \frac{1}{2}\,|\gamma_1(\boldsymbol{r}_1; \boldsymbol{r}_2)|^2. \tag{2.18}$$

As is seen, even for non-interacting fermions the pair density does not reduce to the product of single-particle densities as it would be for non-correlated particles. The correlation expressed by the last term of (2.18) is called exchange.[2]

Two different pair correlation functions are introduced in the general case:

$$g(\boldsymbol{r}_1, \boldsymbol{r}_2) = \frac{n_2(\boldsymbol{r}_1, \boldsymbol{r}_2)}{n(\boldsymbol{r}_1)n(\boldsymbol{r}_2)}, \quad h(\boldsymbol{r}_1, \boldsymbol{r}_2) = n_2(\boldsymbol{r}_1, \boldsymbol{r}_2) - n(\boldsymbol{r}_1)n(\boldsymbol{r}_2). \tag{2.19}$$

For large spatial distances $|\boldsymbol{r}_1 - \boldsymbol{r}_2|$, g usually tends to unity and h to zero.

For non-interacting fermions, h is just given by the last term of (2.18). For the homogeneous non-interacting fermion gas it is given by half the square of (2.10), and

$$g(r) = 1 - \frac{9}{2}\left[\frac{\sin k_f r - k_f r \cos k_f r}{(k_f r)^3}\right]^2 \tag{2.20}$$

with $r = |\boldsymbol{r}_1 - \boldsymbol{r}_2|$, $g(0) = 1/2$. An *exchange hole* around a given fermion is dug out of the distribution of all the other fermions, half of the average density in depth and oscillating with the wavelength π/k_f at large distances (cf. Fig.2). Its depth has to do with only the particles of equal spin direction taking part in the exchange.

[2]In the probability-theoretical sense of particle distributions, exchange in the pair density is of course a correlation. It is a particular type of correlation which has its origin solely in quantum kinematics (symmetry of the many-particle wavefunction), therefore it appears even in non-interacting systems. Of course, it does not change the energy of a system of non-interacting particles, as this energy depends only on the single-particle density matrix (cf. Section 2.5). In the context of many-particle physics, however, the word 'correlation' is used in a narrower meaning and is reserved for particle correlation due to interaction and beyond exchange. For interacting systems, both exchange and correlation contribute to the energy.

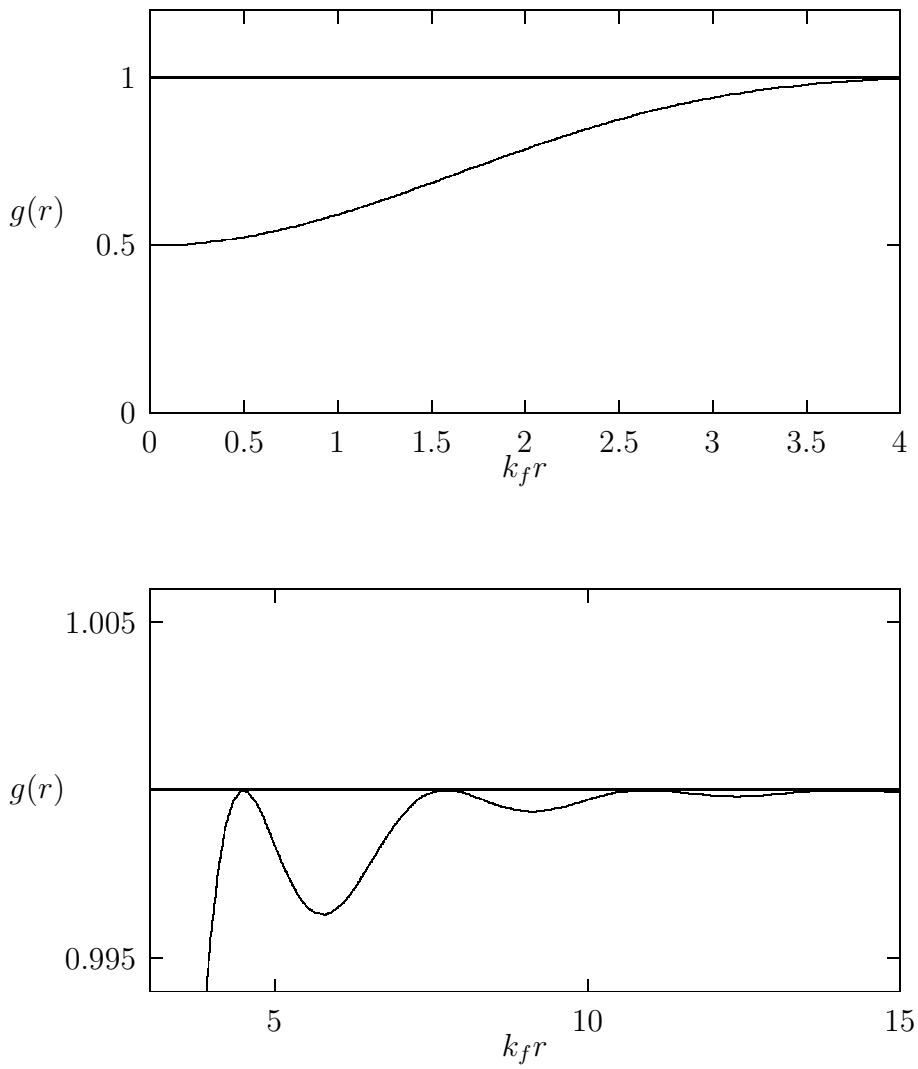

Figure 2: Pair correlation function $g(r)$ of the homogeneous non-interacting electron gas. In this case, $h(r) = n^2\,[g(r) - 1]$. If the volume per electron is represented by a ball of radius r_s, i.e. $4\pi r_s^3/3 = n^{-1}$, then, from (1.43), $k_f r_s = (9\pi/4)^{1/3} = 1.919$.

2.3 Density Operators

The particle density may be represented as the expectation value of a particle density operator, which, for an N-particle system and with the spin

dependence retained, is formally defined as

$$\hat{n}(x) = \sum_{i=1}^{N} \delta(\boldsymbol{r} - \hat{\boldsymbol{r}}_i)\, \delta_{s\hat{\sigma}_i}. \tag{2.21}$$

Here, $\hat{\boldsymbol{r}}_i$ is the position operator and $\hat{\sigma}_i$ is the spin operator of the i-th particle. In the Schrödinger representation, $\hat{\boldsymbol{r}}_i$ reduces simply to the position vector $\boldsymbol{r}_i$. As $\int d^3r\, \delta(\boldsymbol{r} - \hat{\boldsymbol{r}}_i) = 1$ and $\sum_s \delta_{s\hat{\sigma}_i} = 1$, one has

$$\int dx\, \hat{n}(x) = N. \tag{2.22}$$

To be precise, 1 and N, respectively, have to be understood as the real number multiplied with the identity operator in the corresponding representation space. The spin-dependent number density in the many-body quantum state Ψ is

$$n(x) = \langle \Psi | \hat{n}(x) | \Psi \rangle = \langle \hat{n}(x) \rangle. \tag{2.23}$$

The equivalence of this expression with the definition of $n(x) = n_{ss}(\boldsymbol{r})$ as the diagonal of (2.4) is easily seen from (2.21) and the (anti-)symmetry of the wavefunction Ψ with respect to an interchange of particle coordinates and spin variables. Both definitions of $n(x)$ yield immediately

$$\int dx\, n(x) = N, \tag{2.24}$$

and hence

$$\int dx\, (\hat{n}(x) - n(x)) = 0, \tag{2.25}$$

i.e., not only the quantum average $\langle \hat{n}(x) - n(x) \rangle$ of density fluctuations is zero but also the spatial average.

Summing $\hat{n}(x)$ over the spin variable s, one obtains a spin-independent particle density operator $\hat{n}(\boldsymbol{r})$ with the relations

$$\hat{n}(\boldsymbol{r}) = \sum_{i=1}^{N} \delta(\boldsymbol{r} - \hat{\boldsymbol{r}}_i), \qquad \int d^3r\, \hat{n}(\boldsymbol{r}) = N, \tag{2.26}$$

$$n(\boldsymbol{r}) = \langle \Psi | \hat{n}(\boldsymbol{r}) | \Psi \rangle = \langle \hat{n}(\boldsymbol{r}) \rangle, \tag{2.27}$$

$$\int d^3r\, n(\boldsymbol{r}) = N, \quad \int d^3r\, (\hat{n}(\boldsymbol{r}) - n(\boldsymbol{r})) = 0 \tag{2.28}$$

in analogy to (2.21-2.25).

The same operators (2.21, 2.26) in a Fock-space representation may be expressed as

$$\hat{n}(x) = \hat{\psi}^\dagger(x)\hat{\psi}(x), \quad \hat{n}(\boldsymbol{r}) = \sum_s \hat{\psi}^\dagger(x)\hat{\psi}(x) \tag{2.29}$$

through field operators (cf. (1.78)).

2.4 Expectation Values and Density Matrices

We start these considerations with the Schrödinger representation. Given some single-particle operator

$$\hat{T}_1 = \sum_{i=1}^{N} \hat{t}_1(\boldsymbol{r}_i), \tag{2.30}$$

its expectation value in the many-body state Ψ is

$$\begin{aligned}
\langle \hat{T}_1 \rangle &= \int dx_1 \ldots dx_N\, \Psi^*(x_1 \ldots x_N) \sum_{i=1}^{N} \hat{t}_1(\boldsymbol{r}_i)\, \Psi(x_1 \ldots x_N) = \\
&= N \int dx \left[\hat{t}_1(\boldsymbol{r}) \int dx_2 \ldots dx_N * \right. \\
&\qquad\qquad \left. * \, \Psi(x x_2 \ldots x_N) \Psi^*(x' x_2 \ldots x_N) \right]_{x'=x} = \\
&= \int d^3r \, \left[\hat{t}_1(\boldsymbol{r})\, \gamma_1(\boldsymbol{r}; \boldsymbol{r}') \right]_{\boldsymbol{r}'=\boldsymbol{r}} = \mathrm{tr}\, \left(\hat{t}_1 \gamma_1 \right).
\end{aligned} \tag{2.31}$$

In the second line, the symmetry of $\Psi\Psi^*$ with respect to an interchange of particle variables x_i and x_1 was used. After summing over $s = s'$ from which $\hat{t}_1$ is assumed independent, the spin-independent single-particle density matrix (2.2) appears. To ensure that $\hat{t}_1$ only operates on the variable $\boldsymbol{r}$ coming from Ψ, $\boldsymbol{r}'$ is to be put equal to $\boldsymbol{r}$ at the very end. Hence, the knowledge of this single-particle density matrix suffices to calculate $\langle \hat{T}_1 \rangle$. The kinetic energy, for instance, is obtained as

$$\langle \hat{T} \rangle = -\frac{1}{2} \int d^3r \, \left[\nabla^2\, \gamma_1(\boldsymbol{r}; \boldsymbol{r}') \right]_{\boldsymbol{r}'=\boldsymbol{r}}. \tag{2.32}$$

Similarly, expectation values of a spin operator are expressed by means of the spin dependent single-particle density matrix. For instance, from (1.10), a vector spin operator $\hat{\boldsymbol{\sigma}} = \sum_\alpha \boldsymbol{e}_\alpha \hat{\sigma}_\alpha$ may be defined with Cartesian coordinate unit vectors $\boldsymbol{e}_\alpha$. (It is this construct which connects the SU(2), the transformation group of the two-dimensional unitary space of spinors, with the SO(3), the transformation group of the orthogonal space of three-dimensional Euclidean vectors.) We define the operator of the vector spin density in Schrödinger representation as

$$\hat{\boldsymbol{S}}(\boldsymbol{r}) = \sum_{i=1}^{N} \hat{\boldsymbol{\sigma}}_i \, \delta(\boldsymbol{r} - \boldsymbol{r}_i). \tag{2.33}$$

Twice its expectation value, which is equal to the spin magnetization density in units of μ_{Bohr}—the Bohr magneton; the factor two is the gyromagnetic factor—is obtained from the spin-density matrix (2.4):

$$\boldsymbol{m}(\boldsymbol{r}) \stackrel{\text{def}}{=} 2\langle \hat{\boldsymbol{S}}(\boldsymbol{r}) \rangle = 2 \sum_{ss'} \boldsymbol{\sigma}_{s's} \, n_{ss'}(\boldsymbol{r}) = 2 \, \mathrm{tr}_{\,s}(\boldsymbol{\sigma} n(\boldsymbol{r})). \tag{2.34}$$

With the help of (1.10), the components of $\boldsymbol{m}$ are expressed through the elements of the Hermitian spin-density matrix

$$n_{ss'}(\boldsymbol{r}) = \begin{pmatrix} n_{++}(\boldsymbol{r}) & n_{+-}(\boldsymbol{r}) \\ n_{-+}(\boldsymbol{r}) & n_{--}(\boldsymbol{r}) \end{pmatrix} \tag{2.35}$$

as

$$m_x(\boldsymbol{r}) = 2\mathrm{Re}\, n_{-+}(\boldsymbol{r}), \qquad m_y(\boldsymbol{r}) = 2\mathrm{Im}\, n_{-+}(\boldsymbol{r}),$$

$$m_z(\boldsymbol{r}) = n_{++}(\boldsymbol{r}) - n_{--}(\boldsymbol{r}). \tag{2.36}$$

(Formally, the spin-density matrix is a symmetric second rank spinor; every symmetric second rank spinor can be related to a vector as every symmetric even rank spinor can be related to a symmetric irreducible tensor.) It is easy to verify that $\boldsymbol{m}^2 = (\,\mathrm{tr}\, n_{ss'})^2 - 4 \det n_{ss'}$. The trace (being equal to the particle density n) and the determinant of the spin-density matrix are the two invariants with respect to spatial rotations. Likewise, two independent invariants are n and $|\boldsymbol{m}|$. The *degree of spin polarization* ζ and its direction $\boldsymbol{e}_m$ are defined as

$$\zeta \stackrel{\text{def}}{=} \frac{|\boldsymbol{m}|}{n}, \qquad 0 \le \zeta \le 1, \qquad \boldsymbol{e}_m \stackrel{\text{def}}{=} \frac{\boldsymbol{m}}{|\boldsymbol{m}|}. \tag{2.37}$$

Particle density $n(\boldsymbol{r})$ and degree of spin polarization $\zeta(\boldsymbol{r})$ form another pair of invariants of the spin-density matrix. Together with the magnetization direction $\boldsymbol{e}_m(\boldsymbol{r})$, which contains two further independent real functions of $\boldsymbol{r}$, they comprise the four independent real functions of $\boldsymbol{r}$ contained in the Hermitian spin-density matrix (2.35). Sometimes it makes sense to fix only an undirected magnetization axis, and to let ζ vary between -1 and 1.

A simpler situation appears, if the spin quantization axis (magnetization axis) is fixed in the whole space (collinear spin situation). In this case, given some single-particle operator

$$\hat{B}_1 = \sum_{i=1}^{N} B_1(x_i), \tag{2.38}$$

which is merely an ordinary function of the particle variables x_i, its expectation value is readily obtained from the number density $n(x)$ as

$$\langle \hat{B}_1 \rangle = \int dx\, B_1(x) n(x). \tag{2.39}$$

An important example is

$$B_1(x) = v(\boldsymbol{r}) - 2sB(\boldsymbol{r}) \tag{2.40}$$

with an external potential v and an external magnetic field B in z-direction coupled to the spin only (z being the quantization axis of the spin).

A prototype of a two-particle operator is a pair-potential operator, which in the case of a spin-independent interaction, the Coulomb interaction say, has the form

$$\hat{A}_2 = \frac{1}{2} \sum_{i \neq j} A_2(\boldsymbol{r}_i, \boldsymbol{r}_j). \tag{2.41}$$

(In the Coulomb case $A_2(\boldsymbol{r}_i, \boldsymbol{r}_j) = 1/|\boldsymbol{r}_i - \boldsymbol{r}_j|$.) In analogy to (2.31) one finds

$$\begin{aligned} \langle \hat{A}_2 \rangle &= \int d^3r d^3r'\, A_2(\boldsymbol{r}, \boldsymbol{r}') \gamma_2(\boldsymbol{r}, \boldsymbol{r}'; \boldsymbol{r}, \boldsymbol{r}') = \\ &= \frac{1}{2} \int d^3r d^3r'\, A_2(\boldsymbol{r}, \boldsymbol{r}') n_2(\boldsymbol{r}, \boldsymbol{r}') \end{aligned} \tag{2.42}$$

which expresses the expectation value of a two-particle operator by means of the two-particle density matrix γ_2 (more specifically its diagonal part—in

correspondence with the expectation value (2.39) being expressed through the density, i.e. the diagonal part of the single-particle density matrix). The factor 1/2 does not appear in the first expression because it was already absorbed into the definition (2.12) of the two-particle density matrix.

An alternative way of deriving (2.39) uses the particle density operator (2.21):

$$\begin{aligned}
\langle \hat{B}_1 \rangle &= \left\langle \int dx\, B_1(x) \sum_{i=1}^{N} \delta(\boldsymbol{r} - \boldsymbol{r}_i)\delta_{ss_i} \right\rangle = \\
&= \int dx\, B_1(x) \left\langle \sum_{i=1}^{N} \delta(\boldsymbol{r} - \boldsymbol{r}_i)\delta_{ss_i} \right\rangle = \\
&= \int dx\, B_1(x) n(x). \qquad (2.43)
\end{aligned}$$

The first equality just rewrites the operator (2.38) with the help of an integral over a sum of δ-functions. Taking out of the brackets terms not depending on the particle variables $\boldsymbol{r}_i, s_i$ leaves inside the brackets just the particle density operator (in the Schrödinger representation used for (2.38)), the expectation value of which is the particle density.

The expectation value (2.42) may also be expressed via the particle density operator (2.21). This needs only an additional little trick consisting in adding to and subtracting from the double sum of (2.41) 'self-interaction' terms:

$$\begin{aligned}
\langle \hat{A}_2 \rangle &= \left\langle \frac{1}{2} \int d^3r d^3r'\, A_2(\boldsymbol{r}, \boldsymbol{r}') * \right. \\
&\quad \left. * \left[\sum_{ij} \delta(\boldsymbol{r} - \boldsymbol{r}_i)\delta(\boldsymbol{r}' - \boldsymbol{r}_j) - \sum_{i} \delta(\boldsymbol{r} - \boldsymbol{r}_i)\delta(\boldsymbol{r}' - \boldsymbol{r}_i) \right] \right\rangle = \\
&= \frac{1}{2} \int d^3r d^3r'\, A_2(\boldsymbol{r}, \boldsymbol{r}') * \\
&\quad * \left\langle \sum_{ij} \delta(\boldsymbol{r} - \boldsymbol{r}_i)\delta(\boldsymbol{r}' - \boldsymbol{r}_j) - \sum_{i} \delta(\boldsymbol{r} - \boldsymbol{r}_i)\delta(\boldsymbol{r}' - \boldsymbol{r}) \right\rangle = \\
&= \frac{1}{2} \int d^3r d^3r'\, A_2(\boldsymbol{r}, \boldsymbol{r}') \left[\langle \hat{n}(\boldsymbol{r})\hat{n}(\boldsymbol{r}') \rangle - n(\boldsymbol{r})\delta(\boldsymbol{r} - \boldsymbol{r}') \right]. \qquad (2.44)
\end{aligned}$$

(Those usually infinite 'self-interaction terms' of a point particle 'on place' are not to be confused with the finite self-interaction terms of a particle in a

spatial orbital appearing in (1.54) of the Hartree-Fock theory.) In the second sum of the second line, the argument of the second δ-function was changed in accordance with the argument of the first one. After this change, the second δ-function can be taken out of the brackets. Note that for a singular interaction potential as the Coulomb one the separate $\boldsymbol{r}$-integral over the last item containing the δ-function would be infinite. This shows that the splitting of the expression for $n_2(\boldsymbol{r}, \boldsymbol{r}')$ in square brackets into two items is formal: the expectation value of the first item must contain the same δ-function contribution canceling the second since the Coulomb interaction energy of an N-electron quantum system is finite. The formal splitting allows, however, an explicit expression of $\langle \hat{A}_2 \rangle$ in terms of the particle density *operator* only. (Of course it cannot be expressed in terms of the particle *density* $n(\boldsymbol{r})$ only.) Integration of the expression in square brackets over $\boldsymbol{r}$ and $\boldsymbol{r}'$ yields with the help of (2.26)

$$\begin{aligned}
\int d^3r d^3r' \, n_2(\boldsymbol{r}, \boldsymbol{r}') &= \int d^3r d^3r' \, [\langle \hat{n}(\boldsymbol{r}) \hat{n}(\boldsymbol{r}') \rangle - n(\boldsymbol{r}) \delta(\boldsymbol{r} - \boldsymbol{r}')] = \\
&= \int d^3r \, [\langle \hat{n}(\boldsymbol{r}) N \rangle - n(\boldsymbol{r})] = \\
&= \int d^3r \, n(\boldsymbol{r})(N-1) = N(N-1) \qquad (2.45)
\end{aligned}$$

in accordance with (2.14) and (2.12).

Since the density operators (2.21) and (2.26) were defined representation independent, the result (2.44) likewise holds in every representation.

2.5 The Exchange and Correlation Hole

The results of the last section show that knowledge of γ_1 (comprising the knowledge of n as its diagonal) and of n_2 suffices to calculate the total energy of the system as the expectation value of the Hamiltonian (1.24).

Considering additionally (2.19) yields

$$\begin{aligned} E = \langle \hat{H} \rangle &= -\frac{1}{2} \int d^3r \, \left[\nabla^2 \gamma_1(\boldsymbol{r}; \boldsymbol{r}') \right]_{\boldsymbol{r}'=\boldsymbol{r}} + \\ &\quad + \int dx \, v(x) n(x) + \\ &\quad + \frac{1}{2} \int d^3r d^3r' \, n(\boldsymbol{r}') w(|\boldsymbol{r}' - \boldsymbol{r}|) n(\boldsymbol{r}) + \\ &\quad + \frac{1}{2} \int d^3r d^3r' \, w(|\boldsymbol{r}' - \boldsymbol{r}|) h(\boldsymbol{r}', \boldsymbol{r}) = \\ &= E_{\text{kin}} + E_{\text{pot}} + E_{\text{H}} + W_{\text{XC}}. \end{aligned} \tag{2.46}$$

In a natural way the energy is decomposed into the kinetic energy E_{kin}, the interaction energy E_{pot} with the external potential $v(x)$, the so-called Hartree energy E_{H} being the classical interaction energy of a density $n(\boldsymbol{r})$ with itself, and an exchange and correlation energy W_{XC}, which may be visualized as appearing from the interaction energy $\omega_{\text{XC}}(\boldsymbol{r})$ of a particle at $\boldsymbol{r}$ with an exchange and correlation hole (as used in Quantum Chemistry, cf. however (2.68))

$$h_{\text{XC}}(\boldsymbol{r}', \boldsymbol{r}) = h(\boldsymbol{r}', \boldsymbol{r}) / n(\boldsymbol{r}) \tag{2.47}$$

surrounding the particle:

$$\omega_{\text{XC}}(\boldsymbol{r}) = \frac{1}{2} \int d^3r' \, w(|\boldsymbol{r}' - \boldsymbol{r}|) h_{\text{XC}}(\boldsymbol{r}', \boldsymbol{r}). \tag{2.48}$$

The exchange and correlation energy

$$W_{\text{XC}} = \int d^3r \, n(\boldsymbol{r}) \omega_{\text{XC}}(\boldsymbol{r}) \tag{2.49}$$

is obtained by summing $\omega_{\text{XC}}(\boldsymbol{r})$ over all particles, i.e. integrating $n(\boldsymbol{r})\omega_{\text{XC}}(\boldsymbol{r})$ over the $\boldsymbol{r}$-space. The factor 1/2 in ω_{XC} appears because each particle is considered twice: once as the particle interacting with the exchange and correlation hole and once as taking part in the composition of this hole.

The number of particles missing in the exchange and correlation hole (2.47) can easily be obtained from (2.14) and (2.13) implying

$$\int d^3r' \, n_2(\boldsymbol{r}', \boldsymbol{r}) = (N-1) n(\boldsymbol{r}), \tag{2.50}$$

which immediately leads to the general sum rule (cf. (2.19))

$$\int d^3r'\, h_{\mathrm{XC}}(\boldsymbol{r}',\boldsymbol{r}) = \int d^3r' \left[\frac{n_2(\boldsymbol{r}',\boldsymbol{r})}{n(\boldsymbol{r})} - n(\boldsymbol{r}')\right] = -1, \tag{2.51}$$

that is, exactly one particle is missed in any exchange and correlation hole, independent of the system and of the interaction, and independent of the position $\boldsymbol{r}$ of the particle which sees the hole.

In the case of an uncorrelated N-fermion wavefunction (1.15) only an exchange hole appears. From (2.16–2.18) it is obtained as

$$h_{\mathrm{X}}(\boldsymbol{r}',\boldsymbol{r}) = -\frac{\sum_{ij=1}^{N/2} \phi_i(\boldsymbol{r}')\phi_j(\boldsymbol{r})\phi_j^*(\boldsymbol{r}')\phi_i^*(\boldsymbol{r})}{\sum_{i=1}^{N/2} \phi_i(\boldsymbol{r})\phi_i^*(\boldsymbol{r})}. \tag{2.52}$$

For the homogeneous non-interacting fermion gas one finds from (2.19, 2.20) and (1.43)

$$h_{\mathrm{X}}(r) = -\frac{3k_f^3}{2\pi^2}\left[\frac{\sin k_f r - k_f r\cos k_f r}{(k_f r)^3}\right]^2, \quad h_{\mathrm{X}}(0) = -\frac{k_f^3}{6\pi^2} = -\frac{n}{2} \tag{2.53}$$

for the exchange hole (cf. the end of Section 2.2). Of course, in this hole also just one particle is missed, because this result (2.51) was obtained on completely general grounds. The additional amount of density, by particle repulsion expelled from the region of small r-values as compared to the pure exchange hole of Fig.2, is piled up at larger r-values, where in case of particle repulsion $g(r)$ may exceed unity. Compared to Fig.2, the pair correlation function of an electron liquid with Coulomb interaction at a typical valence density starts out at $r = 0$ at a value well below 1/2 and approaches unity at $k_f r$ well before 4, then slightly overshooting. The relevance of this issue will be discussed in Section 7.3.

The physical idea behind the notion of an exchange and correlation hole is that due to Pauli's exclusion principle electrons of parallel spin cannot come arbitrarily close to each other in space, and that the repulsive interaction leads to a further reduction of the pair density for small distances of the particles in the pair. In fact (2.53) results due to the first of those two mechanisms. (The exchange and correlation hole h_{XC} of an interacting homogeneous electron liquid at metallic densities is close to $-n$ for $r = 0$.) Formally however, the exchange and correlation hole is defined by (2.47) on the basis of (2.46–2.49), and those expressions have no refer to the actual total particle number N, or to the actual particle number contributing to

the particle density in a particular spatial region. This gives rise to situations, where the above physical idea becomes irrelevant, and the exchange and correlation hole takes on a different, more formal meaning. (A detailed discussion may be found in [Perdew and Zunger, 1981]). The extreme case is a single particle bound in an external potential well. In this case, of course,

$$E_{\mathrm{H}} + W_{\mathrm{XC}} = 0, \tag{2.54}$$

since there is no second particle to interact with. Nevertheless there is a nonzero particle density $n(\boldsymbol{r}) = \phi^*(\boldsymbol{r})\phi(\boldsymbol{r})$ and hence, formally, a Hartree energy which must exactly be compensated by the W_{XC}-term. This is achieved by putting

$$h_{\mathrm{XC}}(\boldsymbol{r}', \boldsymbol{r}) = -n(\boldsymbol{r}') \tag{2.55}$$

in this case. (Note that h_{XC} is by definition not symmetric in its arguments. It describes a hole distribution over vectors $\boldsymbol{r}'$ around a given position $\boldsymbol{r}$.) The above result can be brought in accordance with the relations of Section 2.2 by putting

$$\gamma_2 \stackrel{\mathrm{def}}{=} 0 \text{ for } N < 2 \tag{2.56}$$

and hence $h(\boldsymbol{r}', \boldsymbol{r}) = -n(\boldsymbol{r}')n(\boldsymbol{r})$ in this case. The sum rule (2.51) is obviously obeyed also by this pathological case $N = 1$. The 'exchange and correlation hole' of that single particle is independent of the particle's position $\boldsymbol{r}$, and is just equal to minus its own density.

Consider next an H_2 molecule in its singlet ground state. This ground state is reasonably well described by a Slater determinant in which both electrons occupy the lowest molecular orbital $\phi(\boldsymbol{r})$. From (2.52) we find in this case

$$h_{\mathrm{XC}}(\boldsymbol{r}', \boldsymbol{r}) = -\phi(\boldsymbol{r}')\phi^*(\boldsymbol{r}') = -\frac{1}{2}n(\boldsymbol{r}'), \tag{2.57}$$

again in agreement with the sum rule. This result is easily understood. No exchange repulsion takes place, because the two electrons have opposite spin. Correlation due to interaction was neglected by the determinant ansatz. As $-h_{\mathrm{XC}}$ equals the density of one of the electrons, $n + h_{\mathrm{XC}}$ is the density of the second one. Thus, with the help of this h_{XC}, W_{XC} just subtracts the self-interaction of one electron in its orbital from the Hartree energy, containing by definition those self-interaction contributions. Again, in the considered

approximation of the ground state, the 'exchange and correlation hole' is independent of the position $\boldsymbol{r}$ of the particle it surrounds. It provides again only a correction term and not really a correlation term.

Let now the H_2 molecule slowly dissociate. As is well known, for larger distances between the two protons the Slater determinant of molecular orbitals loses its value as an approximation for the electronic ground state, rather the latter is now well approximated by a Heitler-London ansatz

$$\Psi(\boldsymbol{r}'\boldsymbol{r}) = \frac{1}{\sqrt{2}}(\phi_A(\boldsymbol{r}')\phi_B(\boldsymbol{r}) + \phi_B(\boldsymbol{r}')\phi_A(\boldsymbol{r})), \tag{2.58}$$

where ϕ_A and ϕ_B are now atomic orbitals around the centers A and B of the two protons. The two-particle density matrix is just the product of this two-particle wavefunction with its complex conjugate. The exchange and correlation hole can be written down in a straightforward way from this wavefunction although it is rather involved. For large distances, however, for which the two orbitals do no longer overlap, $\phi_A(\boldsymbol{r})\phi_B(\boldsymbol{r}) = 0$, it simplifies to

$$\begin{aligned} & h_{\mathrm{XC}}(\boldsymbol{r}', \boldsymbol{r}) = \\ & \quad = -\frac{\phi_A(\boldsymbol{r}')\phi_A^*(\boldsymbol{r}')\phi_A(\boldsymbol{r})\phi_A^*(\boldsymbol{r}) + \phi_B(\boldsymbol{r}')\phi_B^*(\boldsymbol{r}')\phi_B(\boldsymbol{r})\phi_B^*(\boldsymbol{r})}{\phi_A(\boldsymbol{r})\phi_A^*(\boldsymbol{r}) + \phi_B(\boldsymbol{r})\phi_B^*(\boldsymbol{r})}. \end{aligned} \tag{2.59}$$

If now both position vectors $\boldsymbol{r}'$ and $\boldsymbol{r}$ are close to A, then only the first terms of the numerator and denominator are nonzero, and we obtain a result around A similar to (2.55), likewise for B. If one of the positions is close to A and the other close to B, the expression (2.59) vanishes. Given a position vector $\boldsymbol{r}$ close to one atomic site, the exchange and correlation hole extends only over that site. The Heitler-London ansatz contains only configurations with one electron at each atomic site. Hence there is only locally on each site a self-interaction correction needed to correct for the Hartree energy.

As we have seen, for few-particle systems a large part if not all of the exchange and correlation hole has formally to be introduced in order to self-interaction correct the Hartree term. For large systems of essentially delocalized particles the self-interaction correction of the Hartree term tends to zero, and the exchange and correlation hole takes on its physically intuitive meaning discussed above. However, even in large systems, if the particles remain essentially in localized orbitals as in a Heitler-London situation, self-interaction corrections keep up at a non-zero level.

2.6 The Adiabatic Principle

This section introduces an important tool for the formal development of many-body theory yielding a basis for many proofs and representations of issues and being called integration over the coupling constant. To this end, in the present section we will write explicitly down the coupling constant λ of (1.21) which is otherwise put to unity in our context. In the Hamiltonian (1.4) we replace accordingly $\hat{W}$ by $\lambda\hat{W}$ and denote the Hamiltonian itself by $\hat{H}_\lambda$. This notation is in accordance with (1.5).

Consider now the ground state solution of (1.2) at some given value of λ, denoted by $|\Psi_\lambda\rangle$ and corresponding to the ground state energy

$$E_\lambda = \langle\Psi_\lambda|\,\hat{H}_\lambda\,|\Psi_\lambda\rangle. \tag{2.60}$$

Differentiating the second equation (1.2) yields

$$\left\langle \frac{\partial\Psi_\lambda}{\partial\lambda}\middle|\,\Psi_\lambda\right\rangle + \left\langle \Psi_\lambda\,\middle|\frac{\partial\Psi_\lambda}{\partial\lambda}\right\rangle = 0, \tag{2.61}$$

hence

$$\begin{aligned}&\left\langle \frac{\partial\Psi_\lambda}{\partial\lambda}\middle|\,\hat{H}_\lambda\,\middle|\,\Psi_\lambda\right\rangle + \left\langle \Psi_\lambda\,\middle|\,\hat{H}_\lambda\,\middle|\frac{\partial\Psi_\lambda}{\partial\lambda}\right\rangle = \\ &= \left\langle \frac{\partial\Psi_\lambda}{\partial\lambda}\middle|\,\Psi_\lambda\right\rangle E_\lambda + E_\lambda\left\langle \Psi_\lambda\,\middle|\frac{\partial\Psi_\lambda}{\partial\lambda}\right\rangle = 0.\end{aligned} \tag{2.62}$$

By merely applying a phase factor, the matrix elements of these relations can be made real. Then, due to its normalization, a parametric change of the ground state will be orthogonal to that ground state.

We now assume that the derivative $\partial\Psi_\lambda/\partial\lambda$ indeed exists. This can generally be assumed to hold true if Ψ_λ is non-degenerate. In case of level crossing on varying λ, at least a special choice among the degenerate states is needed as in perturbation theory, implying that $|\Psi_\lambda\rangle$ can generally no longer be the ground state for all values of λ considered. Under the assumption for $\partial\Psi_\lambda/\partial\lambda$ to exist, from (2.60, 2.62),

$$\frac{dE_\lambda}{d\lambda} = \left\langle \Psi_\lambda\,\middle|\,\frac{d\hat{H}_\lambda}{d\lambda}\,\middle|\,\Psi_\lambda\right\rangle = \langle\Psi_\lambda|\,\hat{W}\,|\Psi_\lambda\rangle \stackrel{\text{def}}{=} \langle\hat{W}\rangle_\lambda. \tag{2.63}$$

This very elementarily derivable relation is of enormous importance in quantum physics wherefore it carries the name of two, sometimes even three

famous people having been among the first who put focus on it: It is called the (Pauli-)Hellmann-Feynman theorem[3] [Pauli, 1933, Section A11], [Hellmann, 1937, Feynman, 1939].

The next important assumption is that there is a unique differentiable path of ground states $|\Psi_\lambda\rangle$ in the state space for the whole λ-interval between zero and unity. This is called the adiabaticity assumption because in important applications one has in mind a time-dependent λ and considers the case $d\lambda/dt \to 0$.

In our context, if we call E_0 the ground state energy of the interaction-free reference system (1.5) and denote by $E_{\rm int}$ the total energy change due to particle interaction, we find

$$E = E_0 + E_{\rm int}, \quad E_{\rm int} = \int_0^1 d\lambda \, \langle \hat{W} \rangle_\lambda. \tag{2.64}$$

Note that $E_{\rm int}$ is different from $E_{\rm H} + W_{\rm XC}$: in contrast to the latter $E_{\rm int}$ contains kinetic energy changes due to correlation and also eventually changes in the interaction energy with an external field due to a particle-interaction dependence of the ground state density.

Besides (2.46), there is another decomposition of the total ground state energy of the system due to Kohn and Sham [Kohn and Sham, 1965], namely

$$E = T + E_{\rm pot} + E_{\rm H} + E_{\rm XC}, \tag{2.65}$$

where T is the kinetic energy of an interaction-free reference system in such an external potential $v_0(x)$ that it has *the same ground state density* $n(x)$ *as the considered interacting system in the external potential* $v(x)$. The contributions $E_{\rm pot}$ and $E_{\rm H}$ are defined as previously with the actual ground state density $n(x)$. The last term $E_{\rm XC}$ is *defined* as the difference between the left-hand side and the sum of the preceding terms, and is called the Kohn-Sham exchange and correlation energy. It differs from $W_{\rm XC}$ by the difference between T and $E_{\rm kin}$, the exchange and correlation contribution to the kinetic energy:

$$T + E_{\rm XC} = E_{\rm kin} + W_{\rm XC} \tag{2.66}$$

At this point one further assumption is needed: There should be such a path $v_\lambda(x)$, $0 \le \lambda \le 1$, in the functional space of potentials, that the

[3] In fact, the relation was considered even earlier, probably for the first time by P. Güttinger, Z. Phys. **73**, 169 (1931).

ground state density $n_\lambda(x) \equiv n(x)$ is kept constant for $0 \leq \lambda \leq 1$. Then, $d\hat{H}_\lambda/d\lambda = d\hat{V}_\lambda/d\lambda + \hat{W}$, and hence, from the first relation (2.63),

$$\frac{dE_\lambda}{d\lambda} = \int dx\, n(x) \frac{\partial v_\lambda(x)}{\partial \lambda} + E_{\mathrm{H}} + \frac{1}{2} \int d^3r d^3r'\, n(\boldsymbol{r}) w(|\boldsymbol{r} - \boldsymbol{r}'|) h_{\mathrm{XC},\lambda}(\boldsymbol{r}', \boldsymbol{r}).$$

The essential point in obtaining the second term of this relation, E_{H}, was that $n_\lambda(x)$ is kept constant as a function of λ. Integration over λ yields

$$\begin{aligned} E - E_0 &= \int_0^1 d\lambda \frac{dE_\lambda}{d\lambda} = \\ &= \int dx\, n(x)(v(x) - v_0(x)) + E_{\mathrm{H}} + \\ &\quad + \frac{1}{2} \int d^3r d^3r'\, n(\boldsymbol{r}) w(|\boldsymbol{r} - \boldsymbol{r}'|) \int_0^1 d\lambda\, h_{\mathrm{XC},\lambda}(\boldsymbol{r}', \boldsymbol{r}), \\ E_0 &= T + \int dx\, n(x) v_0(x). \end{aligned}$$

Comparison with (2.65) results in

$$E_{\mathrm{XC}} = \frac{1}{2} \int d^3r d^3r'\, n(\boldsymbol{r}) w(|\boldsymbol{r} - \boldsymbol{r}'|) h_{\mathrm{KS}}(\boldsymbol{r}', \boldsymbol{r}) \tag{2.67}$$

with the Kohn-Sham exchange and correlation hole

$$h_{\mathrm{KS}}(\boldsymbol{r}', \boldsymbol{r}) = \int_0^1 d\lambda\, h_{\mathrm{XC},\lambda}(\boldsymbol{r}', \boldsymbol{r}) \tag{2.68}$$

[Perdew and Zunger, 1981, Gunnarsson and Lundqvist, 1976].

Coming back to theorem (2.63), it, of course, holds for any parameter λ the Hamiltonian depends on. A great variety of applications considers a change in time of the external field $v(x)$. For instance, think of a sufficiently slow motion of nuclei, with electrons quickly moving in the Coulomb field of those nuclei. The adiabatic principle says that a system in the ground state $|\Psi_\lambda\rangle$ will remain all the time in its ground state, if $d\lambda/dt \to 0$, the adiabaticity assumption mentioned above is valid, and that ground state is separated by a gap from the excitation spectrum of the system [Gell-Mann and Low, 1951]. Among many other applications this adiabatic principle forms a basis for a quantum mechanical treatment of adiabatic (in this sense) forces for the ionic motion in molecules and solids.

2.7 Coulomb Systems

In this section we further specialize the Hamiltonian (1.24) to

$$\begin{aligned}\hat{H}_{\rm Coul} &= \hat{H}_{\rm Coul}(\boldsymbol{r}_i; Z_\mu, \boldsymbol{R}_\mu) = \hat{T} + \hat{V}_{\rm Coul} = \\ &= -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - \sum_{i=1}^{N}\sum_{\mu=1}^{M}\frac{Z_\mu}{|\boldsymbol{r}_i - \boldsymbol{R}_\mu|} + \frac{1}{2}\sum_{i\neq j}^{N}\frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} + \\ &\quad + \frac{1}{2}\sum_{\mu\neq\nu}^{M}\frac{Z_\mu Z_\nu}{|\boldsymbol{R}_\mu - \boldsymbol{R}_\nu|}. \end{aligned} \tag{2.69}$$

The external potential v is thought to be created by the Coulomb field of M nuclei fixed at positions $\boldsymbol{R}_\mu$ and having electric charges Z_μ. The N electrons mutually interact with Coulomb forces too, and, for fixed $\boldsymbol{R}_\mu$, the last term is just a constant energy—the Coulomb interaction energy of the nuclei. Natural units are used as in (1.24). The nuclear positions $\boldsymbol{R}_\mu$ and the nuclear charges Z_μ are treated as parameters upon which the Hamiltonian (2.69), the total ground state energy

$$E = E(Z_\mu, \boldsymbol{R}_\mu) = E_{\rm kin} + E_{\rm Coul} \tag{2.70}$$

and the N-electron state

$$\Psi(x_i; Z_\mu, \boldsymbol{R}_\mu) = \langle x_1 \ldots x_N | \Psi(Z_\mu, \boldsymbol{R}_\mu)\rangle \tag{2.71}$$

depend. ($E_{\rm kin} = \langle\hat{T}\rangle$ is as previously defined, and $E_{\rm Coul} = \langle\hat{V}_{\rm Coul}\rangle$ is the total Coulomb energy; the notation in (2.71) is as in (1.8).)

We now consider various re-scalings of those parameters. Replace first the charges Z_μ by λZ_μ, $0 \leq \lambda \leq 1$. For $\lambda = 0$ the N electrons move in a constant (zero) external potential and repel each other by Coulomb forces. They move to infinite mutual distances with the infimum of the total energy being $E(\lambda = 0) = 0$. Simply by integrating the first equation of (2.63) over λ from 0 to 1 and taking into account that this time $\hat{W}$ does *not* depend on λ, one immediately gets [Wilson, 1962]

$$\begin{aligned} E &= E(\lambda = 1) = \\ &= -\sum_{\mu=1}^{M} Z_\mu \int_0^1 d\lambda \int d^3r\, \frac{n(\boldsymbol{r};\lambda)}{|\boldsymbol{r} - \boldsymbol{R}_\mu|} + \frac{1}{2}\sum_{\mu\neq\nu}^{M}\frac{Z_\mu Z_\nu}{|\boldsymbol{R}_\mu - \boldsymbol{R}_\nu|} \end{aligned} \tag{2.72}$$

for the total energy of the Coulomb system (2.69). Here, $n(\boldsymbol{r};\lambda)$ is the density of the N interacting electrons in the Coulomb field of scaled nuclear charges.

In order to derive virial theorems we next scale all coordinate vectors according to

$$\begin{aligned}\hat{H}_{\text{Coul},\,\lambda} &\stackrel{\text{def}}{=} \hat{H}_{\text{Coul}}(\lambda \boldsymbol{r}_i; Z_\mu, \lambda \boldsymbol{R}_\mu) = \\ &= \lambda^{-2}\,\hat{T}(\boldsymbol{r}_i) + \lambda^{-1}\,\hat{V}_{\text{Coul}}(\boldsymbol{r}_i; Z_\mu, \boldsymbol{R}_\mu)\end{aligned} \tag{2.73}$$

and

$$\Psi_\lambda = \Psi(\lambda \boldsymbol{r}_i s_i; Z_\mu, \lambda \boldsymbol{R}_\mu). \tag{2.74}$$

While E of (2.70) is obtained as the lowest eigenvalue of the Schrödinger equation $\hat{H}_{\text{Coul}}\Psi = \Psi E$, now the equation

$$\hat{H}_{\text{Coul},\,\lambda}\Psi_\lambda = \Psi_\lambda E_\lambda \tag{2.75}$$

is to be considered. It differs from the original Schrödinger equation for E only by the notation of the dynamical variable being now $\lambda \boldsymbol{r}_i$, and by the replacement $\boldsymbol{R}_\mu \to \lambda \boldsymbol{R}_\mu$ of parameters. While the former change of notation has no effect on E, the latter replacement has. Hence,

$$E_\lambda = E(Z_\mu, \lambda \boldsymbol{R}_\mu), \quad \left.\frac{dE_\lambda}{d\lambda}\right|_{\lambda=1} = \sum_\mu \boldsymbol{R}_\mu \cdot \frac{\partial E}{\partial \boldsymbol{R}_\mu}. \tag{2.76}$$

Alternatively, this derivative may be obtained by applying the first equation (2.63) to (2.73). Comparing both results yields

$$\sum_\mu \boldsymbol{R}_\mu \cdot \frac{\partial E}{\partial \boldsymbol{R}_\mu} = -2E_{\text{kin}} - E_{\text{Coul}}. \tag{2.77}$$

If the positions $\boldsymbol{R}_\mu$ of the nuclei are chosen so that the adiabatic forces on them $\boldsymbol{F}_\mu = \partial E/\partial \boldsymbol{R}_\mu$ are zero (*adiabatic* equilibrium positions), then

$$E_{\text{Coul}} = -2E_{\text{kin}} = 2E \tag{2.78}$$

follows.

Instead of (2.69), the total Hamiltonian for the motion of both nuclei and electrons

$$\hat{H}_{\text{tot}} = \hat{T}_{\text{nucl}} + \hat{H}_{\text{Coul}}, \quad \hat{T}_{\text{nucl}} = -\sum_{\mu=1}^{M} \frac{\nabla_\mu^2}{2M_\mu} \tag{2.79}$$

may be considered, where now the nuclear positions $\boldsymbol{R}_\mu$ are dynamical variables too, rather than parameters. If the above scaling is performed in the Schrödinger equation

$$\hat{H}_{\text{tot}}\Psi_{\text{tot}} = \Psi_{\text{tot}}E_{\text{tot}}, \tag{2.80}$$

then this scaling merely results in a change of notation of the dynamical variables, hence E_{tot} remains independent of λ. Application of (2.63) now yields

$$\begin{aligned} E_{\text{COUL}} &= -2E_{\text{KIN}} = 2E_{\text{tot}}, \\ E_{\text{COUL}} &\stackrel{\text{def}}{=} \langle\Psi_{\text{tot}}|\hat{V}_{\text{Coul}}|\Psi_{\text{tot}}\rangle, \quad E_{\text{KIN}} \stackrel{\text{def}}{=} \langle\Psi_{\text{tot}}|\hat{T}_{\text{nucl}} + \hat{T}|\Psi_{\text{tot}}\rangle. \end{aligned} \tag{2.81}$$

The difference between E_{COUL} and E_{Coul} is caused by correlation between the nuclei and by a change of the correlation between nuclei and electrons due to nuclear motion, which is not contained in the adiabatic theory. This correlation energy is again negative and roughly equal to –2 times the kinetic energy of the nuclei, as can be seen from comparing (2.78) with (2.81): The kinetic energy of the nuclei is roughly the difference $E_{\text{KIN}} - E_{\text{kin}}$. The heavier the masses M_μ of the nuclei, the smaller their kinetic energy compared to that of the electrons, and hence the better the adiabatic approximation. (Non-adiabatic energy corrections are systematically of the order of $M_\mu^{-1/2}$; cf. e.g. [Born and Huang, 1968], Section 14.)

A comprehensive review of the various virial theorems may be found in [Marc and McMillan, 1985].

Another general problem of Coulomb systems is connected with the Hamiltonian (2.69), which is not suited for the thermodynamic limit N, $M \to \infty$, N/V, $M/V = \text{const}$. The problem arises because in this limit each of the three double-sums of (2.69) diverges already, if one fixes one particle index and sums over the second. As a consequence we obtain infinite (with varying sign) contributions to the total energy per particle, and the energy per volume $E/V = \langle\hat{H}_{\text{Coul}}\rangle/V$ is not defined.

This problem can be circumvented in the following manner: Consider the torus $\boldsymbol{T}^3$ defined by

$$\boldsymbol{r} \equiv \boldsymbol{r} + L\boldsymbol{R}_1, \quad \boldsymbol{r} \equiv \boldsymbol{r} + L\boldsymbol{R}_2, \quad \boldsymbol{r} \equiv \boldsymbol{r} + L\boldsymbol{R}_3 \tag{2.82}$$

with $(\boldsymbol{R}_1, \boldsymbol{R}_2, \boldsymbol{R}_3)$ linear independent. Replace the Coulomb potential by a Yukawa potential in the torus

$$v_{\text{Coul}}(r) = \frac{1}{r} \quad \longrightarrow \quad v_{\alpha L}(\boldsymbol{r}) = \sum_{\boldsymbol{l}} v_\alpha(|\boldsymbol{r} + \boldsymbol{R}_{\boldsymbol{l}}|), \tag{2.83}$$

$$v_\alpha(r) = \frac{e^{-\alpha r}}{r}, \quad \boldsymbol{R_l} = L(l_1\boldsymbol{R}_1 + l_2\boldsymbol{R}_2 + l_3\boldsymbol{R}_3),\ l_i \text{ integer}, \tag{2.84}$$

which also means that the Poisson equation is to be replaced according to

$$\Delta v_{\text{Coul}} = -4\pi\delta(\boldsymbol{r}) \quad \longrightarrow \quad (\Delta - \alpha^2)v_\alpha = -4\pi\delta(\boldsymbol{r}). \tag{2.85}$$

Instead of the Hamiltonian (2.69) consider now the Hamiltonian on the torus

$$\begin{aligned}\hat{H}_{\alpha L} = \hat{H}_{\alpha L}(r_i; Z_\mu, \boldsymbol{R}_\mu) \ &= \\ = \ -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - \sum_{i=1}^{N}\sum_{\mu=1}^{M} Z_\mu v_{\alpha L}(|\boldsymbol{r}_i - \boldsymbol{R}_\mu|) + \frac{1}{2}\sum_{i\neq j}^{N} v_{\alpha L}(|\boldsymbol{r}_i - \boldsymbol{r}_j|) + \\ +\frac{1}{2}\sum_{\mu\neq\nu}^{M} Z_\mu Z_\nu v_{\alpha L}(|\boldsymbol{R}_\mu - \boldsymbol{R}_\nu|),\end{aligned} \tag{2.86}$$

where N and M now mean the number of electrons and the number of nuclei, respectively, in the torus $\boldsymbol{T}^3$. In this Hamiltonian, with α non-zero, for $\boldsymbol{r}_i \neq \boldsymbol{R}_\mu$ and $\boldsymbol{r}_i \neq \boldsymbol{r}_j$ all sums remain finite as they were in (2.69) for a *finite* system, i.e. with N and M finite in the whole space. Denote the ground state energy of this torus by $E_{\alpha L}(Z_\mu, \boldsymbol{R}_\mu)$, which for α non-zero is finite, too.

For a finite system, a molecule or a radical say, these considerations would be irrelevant. Nevertheless, it is easily seen that the energy (2.70) is

$$E(Z_\mu, \boldsymbol{R}_\mu) = \lim_{\alpha\downarrow 0}\lim_{L\to\infty} E_{\alpha L}(Z_\mu, \boldsymbol{R}_\mu). \tag{2.87}$$

In the L-limes, due to the exponential decay of the interaction with distance, only one item of the $\boldsymbol{l}$-sum in (2.83) survives, and the α-limes replaces afterwards v_α back by v_{Coul}. The essential point in this case is that $\lim_{L\to\infty} E_{\alpha L}(Z_\mu, \boldsymbol{R}_\mu)$ is a finite continuous function of $\alpha \geq 0$. Picking a sufficiently large L and a sufficiently small α, but $\alpha \gg 1/L$, one can approximate $E(Z_\mu, \boldsymbol{R}_\mu)$ by $E_{\alpha L}(Z_\mu, \boldsymbol{R}_\mu)$ arbitrarily close. For a finite L on the other hand, the limes

$$\lim_{\alpha\downarrow 0} E_{\alpha L}(Z_\mu, \boldsymbol{R}_\mu)$$

does not exist *unless the torus is perfectly neutral with zero dipolar moment.* The individual $\boldsymbol{l}$-sums of (2.83) diverge in that limes. Only for higher multipole distributions the Coulomb potential decays fast enough with distance to ensure convergence of all lattice sums, provided the summation over the neutral assembly is carried out before the limes is taken. In a ferroelectric

crystal, this further needs a domain configuration with zero total moment (or alternatively the separation of an internal macroscopic mean electric field).

Hence, for an extended system, a different limes has to be considered. Here, for $L = 1$ one picks a large enough torus $\boldsymbol{T}_1^3$ and ensures that the torus is perfectly neutral, i.e., $N = \sum_\mu Z_\mu$. Then, for integer $L > 1$, the positions $\boldsymbol{R}_\mu$ are assumed to form L^3 identical copies of the original torus, attached to each other. Finally,

$$(E/V)(Z_\mu, \boldsymbol{R}_\mu) = \lim_{\substack{L\to\infty \\ N/L^3=\text{const.}}} \frac{1}{L^3|\boldsymbol{T}_1^3|} \lim_{\alpha\downarrow 0} E_{\alpha L}(Z_\mu, \boldsymbol{R}_\mu) \tag{2.88}$$

is the ground state energy per volume, if the right side exists and is finite. In (2.88), $|\boldsymbol{T}_1^3|$ is the volume of the torus $\boldsymbol{T}_1^3$ the limiting process was started with (e.g. unit cell in a crystal or a cutout of a disordered solid large enough to be representative). By this prescription, the last term of (2.86) as well as the external potential $\sum_\mu Z_\mu v_{\alpha L}(|\boldsymbol{r} - \boldsymbol{R}_\mu|)$ on a given electron are independent of L. In the L-limes, merely the accessible electron momenta according to (a slightly generalized) relation (1.29) become infinitely dense: The electrons are allowed to delocalize with the electrostatics kept perfectly balanced. Picking a sufficiently large L, one can approximate $(E/V)(Z_\mu, \boldsymbol{R}_\mu)$ by $\lim_{\alpha\downarrow 0} E_{\alpha L}(Z_\mu, \boldsymbol{R}_\mu)/L^3|\boldsymbol{T}_1^3|$ arbitrarily close.

Again, one can consider (E/V) as the ground state energy per volume for fixed $\boldsymbol{R}_\mu$ as just described, or alternatively, by adding $\hat{T}_{\text{nucl}}$, as the ground state energy of the dynamical system of electrons and nuclei. That for an infinite system the latter is finite was only in the mid sixties recognized as an important solvable problem, and (E/V) was first proven to be bounded from below in the end of sixties [Dyson and Lenard, 1967, Lenard and Dyson, 1968], whereby the proof turned out to be astonishingly complicated. A shorter proof with sharpening the estimate by more than ten orders of magnitude was given in [Lieb, 1976] with the help of some results of Thomas-Fermi theory. Those considerations had essentially to fight with the short-range part of the Coulomb potential (proof of absence of a Coulomb collapse), which problem was not considered above because we fixed the nuclear positions. In the latter work by Lieb, however, also the existence of the thermodynamic limit of a neutral system was proven, which is complicated due to the long-range part of the Coulomb potential.

All those proofs did not explicitly treat the Coulomb potential as the limit of an exponentially decaying potential, but sort of a two step limit like (2.88) is effectively used in any numerical calculation of Coulomb energies of

extended systems: The long-range part of the Coulomb potential is Fourier transformed, and the $\boldsymbol{q} = 0$ Fourier component is treated as if it were finite. (Note from (2.85), that $v_{\mathrm{Coul}}(q) \sim 1/q^2$ while $v_\alpha(q) \sim 1/(q^2 + \alpha^2)$.) A standard numerical approach to Coulomb energies of extended systems is described in [Fuchs, 1935, Ewald, 1921]. The total energy is then self-consistently calculated with using a discrete $\boldsymbol{k}$-mesh, and the convergence with increasing $\boldsymbol{k}$-point density is considered. In this approach, the above considerations are reflected in a delicacy of the limes $\boldsymbol{q} \to 0, \alpha \to 0$, leaving e.g. a potential constant fundamentally undetermined. Only the local electro-neutrality ensures that the total energy per volume does not depend on such a constant.

3 Thomas-Fermi Theory

After having started in Chapter 1 with the N-particle wavefunction, in the preceding chapter it was shown that the single-particle and two-particle density matrices suffice to calculate most ground state properties of a many-body system described by the Hamiltonian (1.4). If one were able to classify the set of all admissible two-particle density matrices—the corresponding single-particle density matrices follow from them via (2.13)—then the ground state energy, E, of the many-body system could be obtained as the absolute minimum of (2.46) over this set in a variational way, without reference to many-body wavefunctions as needed in (1.3).

Single-particle density matrices may be classified as all possible self-adjoint trace-class operators (in the Hilbert space of the respective representation) with real non-negative eigenvalues between zero and unity in the fermion case, and arbitrarily large in the boson case. Unfortunately, for two-particle density matrices the situation is much more complicated [Coleman, 1963, Ando, 1963].

The modern progress of density functional theory uses a philosophy which in a manner of speaking starts from the other end. Exploiting the relation between particle densities $n(\boldsymbol{r})$ and many-body wavefunctions $\Psi(x_1 \dots x_N)$ *of ground states*, one tries to find a functional expression of the ground state energy E through the ground state density $n(\boldsymbol{r})$ instead of the two-particle density matrix, and then to base a variational principle for the density on that functional relation.

Thomas-Fermi theory [Thomas, 1927, Fermi, 1927] is the earliest and most naïve version of such theories, which is, however, still of considerable conceptual importance, and it is up to now the only explicit density functional theory being asymptotically exact in a certain sense [Lieb, 1981]. A recent survey from the user's standpoint may be found in [Parr and Yang, 1989].

3.1 The Thomas-Fermi Functional and Thomas-Fermi Equation

Thomas and Fermi independently considered the first three terms of the energy expression (2.46). At that time they were not aware of the exchange energy and neglected the correlation term. Hence the only contribution not readily expressed through the particle density $n(\boldsymbol{r})$ was the kinetic energy. There is, however, one model situation where even the kinetic energy is readily expressed in terms of the particle density. This is the homogeneous interaction-free fermion gas (with spin 1/2), for which (1.45) and (1.43) yield

$$\varepsilon_{\rm kin}(n) = C_{\rm F}\, n^{2/3}, \quad C_{\rm F} = \frac{3}{10}(3\pi^2)^{2/3} = 2.8712 \tag{3.1}$$

for the average kinetic energy per particle as a function of the (constant in space) particle density n. The kinetic energy per unit volume in this situation is $n\varepsilon_{\rm kin}$, and if in a real case the particle density varies sufficiently slowly in space, then

$$E_{\rm kin} \approx \int d^3r\, n(\boldsymbol{r})\varepsilon_{\rm kin}(\boldsymbol{r}) = C_{\rm F}\int d^3r\, n^{5/3}(\boldsymbol{r}) \tag{3.2}$$

may serve as a workable approximation for the kinetic energy functional of the particle density.[4] Hence, the Thomas-Fermi functional for the total energy is

$$\begin{aligned} E_{\rm TF}[n(\boldsymbol{r}); v(\boldsymbol{r})] \;&=\; C_{\rm F}\int d^3r\, n^{5/3}(\boldsymbol{r}) + \int d^3r\, v(\boldsymbol{r})n(\boldsymbol{r}) + \\ &\quad + \frac{1}{2}\int d^3r d^3r'\, n(\boldsymbol{r}')w(|\boldsymbol{r}'-\boldsymbol{r}|)n(\boldsymbol{r}). \end{aligned} \tag{3.3}$$

It is an explicitly given functional of both the density $n(\boldsymbol{r})$ and the external potential $v(\boldsymbol{r})$ (and of course also depends on the form of the pair interaction $w(r)$). For the Coulomb system of Section 2.7 it is

$$\begin{aligned} &E_{\rm TF}[n(\boldsymbol{r}); Z_\mu, \boldsymbol{R}_\mu] \;= \\ &\;=\; C_{\rm F}\int d^3r\, n^{5/3}(\boldsymbol{r}) - \sum_{\mu=1}^{M}\int d^3r\, \frac{Z_\mu n(\boldsymbol{r})}{|\boldsymbol{r}-\boldsymbol{R}_\mu|} + \\ &\quad + \frac{1}{2}\int d^3r d^3r'\, \frac{n(\boldsymbol{r}')n(\boldsymbol{r})}{|\boldsymbol{r}'-\boldsymbol{r}|} + \frac{1}{2}\sum_{\mu\neq\nu}^{M}\frac{Z_\mu Z_\nu}{|\boldsymbol{R}_\mu-\boldsymbol{R}_\nu|}, \end{aligned} \tag{3.4}$$

[4] $E_{\rm kin} \geq C_{\rm LT}\int d^3r\, n^{5/3}(\boldsymbol{r})$ with $C_{\rm LT} = (3/10)(3\pi/4)^{2/3}$ is a rigorous result by [Lieb and Thirring, 1975].

where, as in (2.69), the last constant (i.e. density independent) term of the Coulomb interaction energy of the nuclei is added in order to be prepared for the possibility of a thermodynamic limit (cf. Section 2.7).

We are going to use (3.3) and (3.4), respectively, as variational expressions, in which the density $n(\boldsymbol{r})$ will be varied in order to find the ground state. To determine the range of admissible variations, the next step is to characterize the domain of definition of the above two functionals of $n(\boldsymbol{r})$ [Lieb, 1981, mathematical details of the present chapter will follow this basic paper]; cf. also [Lieb and Simon, 1977]. As most considerations in our context will concern some *complete* (in the sense of the discussion in Section 1.5) functional space, integration should usually be understood in the Lebesgue sense [Reed and Simon, 1973], see also Chapter 5 of the present text. In this context, given some real number $p, 1 \leq p \leq \infty$, $f \in \boldsymbol{L}^p$ means that

$$\|f\|_p \stackrel{\text{def}}{=} \left[\int dx |f(x)|^p\right]^{1/p}, \quad (1 \leq p < \infty) \tag{3.5}$$

is finite, where integration is over the respective Lebesgue measurable space, e.g. the $\boldsymbol{R}^3$ in the present context. Functions coinciding almost everywhere (a.e.), i.e. everywhere except possibly on a set of points of Lebesgue measure zero, are considered equal (their difference is not causing any difference of integrals in which they appear), and $f \in \boldsymbol{L}^\infty$ means

$$\|f\|_\infty \stackrel{\text{def}}{=} \text{ess sup}|f(x)| \tag{3.6}$$

being finite, where the essential supremum ess sup is the smallest real number r so that $|f(x)| \leq r$ a.e. (A reader not really interested in the more mathematical depths may on an intuitive heuristic level take the integrals of the present section in his usual understanding and ignore the appendage a.e.)

A more systematic treatment of the $\boldsymbol{L}^p$-spaces will be given in Chapter 5. Here we list only a few important issues without proof: If $f \in \boldsymbol{L}^p$ and $f \in \boldsymbol{L}^q$ with $p < q$, then $f \in \boldsymbol{L}^t$ for all $p \leq t \leq q$. If $f \in \boldsymbol{L}^p$ and $g \in \boldsymbol{L}^{p'}$ with $1/p + 1/p' = 1$, then $fg \in \boldsymbol{L}^1$, i.e., $\int dx\, |fg| < \infty$, and hence *a fortiori* $|\int dx\, fg| < \infty$. More generally, if $f \in \boldsymbol{L}^p \cap \boldsymbol{L}^q$ and $g = g_1 + g_2,\ g_1 \in \boldsymbol{L}^{p'},\ g_2 \in \boldsymbol{L}^{q'}$, where again $1/p + 1/p' = 1 = 1/q + 1/q'$, then $fg \in \boldsymbol{L}^1$.

The kinetic energy integral of (3.3) or (3.4) is finite, if $n(\boldsymbol{r}) \in \boldsymbol{L}^{5/3}$. On the other hand, $\int d^3r\, n(\boldsymbol{r}) = N$ demands $n(\boldsymbol{r}) \in \boldsymbol{L}^1$. Let now $n(\boldsymbol{r}) \in \boldsymbol{L}^{5/3} \cap \boldsymbol{L}^1,\ v = v_1 + v_2,\ v_1(\boldsymbol{r}) \in \boldsymbol{L}^{5/2},\ v_2(\boldsymbol{r}) \in \boldsymbol{L}^\infty,\ w = w_1 + w_2,\ w_1(|\boldsymbol{r}|) \in$

$\boldsymbol{L}^{5/2}$, $w_2(|\boldsymbol{r}|) \in \boldsymbol{L}^\infty$. Then $\int d^3r\, v_1(\boldsymbol{r})n(\boldsymbol{r})$ is finite because $v_1 \in \boldsymbol{L}^{5/2}$ and $n \in \boldsymbol{L}^{5/3}$, 2/5+3/5=1, and $\int d^3r\, v_2(\boldsymbol{r})n(\boldsymbol{r})$ is finite because $v_2 \in \boldsymbol{L}^\infty$ and $n \in \boldsymbol{L}^1$. Hence $\int d^3r\, v(\boldsymbol{r})n(\boldsymbol{r})$ is finite. Furthermore, $\int d^3r'\, n(\boldsymbol{r}')w(|\boldsymbol{r}'-\boldsymbol{r}|)$ is finite for all $\boldsymbol{r}$ for the same reason, and approaches zero for $\boldsymbol{r} \to \infty$, if $w(r)$ approaches zero for $r \to \infty$. Hence the integral as a function of $\boldsymbol{r}$ is again in $\boldsymbol{L}^\infty$, and $E_{\mathrm{TF}}[n(\boldsymbol{r}); v(\boldsymbol{r})]$ of (3.3) exists and is finite. Furthermore, in the decomposition

$$\frac{1}{|\boldsymbol{r}|} = \frac{e^{-|\boldsymbol{r}|}}{|\boldsymbol{r}|} + \frac{1-e^{-|\boldsymbol{r}|}}{|\boldsymbol{r}|} \tag{3.7}$$

the first part is in $\boldsymbol{L}^{5/2}$ for the 3-dimensional $\boldsymbol{r}$-space ($\int d^3r\, r^{-5/2}e^{-5r/2} \sim \int dr\, r^2 r^{-5/2}e^{-5r/2} = \int dr\, r^{-1/2}e^{-5r/2} < \infty$), and the second part is bounded and hence in $\boldsymbol{L}^\infty$. As we see, the above assumption on $n(\boldsymbol{r})$ guarantees the existence and finiteness of (3.4), too.

The following type of argument is repeatedly used in the context of density functional theory: Given the particle number N, let *admissible* densities $n(\boldsymbol{r})$ be all non-negative $\boldsymbol{L}^{5/3} \cap \boldsymbol{L}^1$-functions with $\int d^3r\, n(\boldsymbol{r}) = N$. Assume that every admissible density is a *ground state* density for some external potential $v(\boldsymbol{r})$ with the corresponding ground state denoted by Ψ_v. If v' is some other potential, then obviously $E[v] = \langle \Psi_v | \hat{H}_v | \Psi_v \rangle \le \langle \Psi_{v'} | \hat{H}_v | \Psi_{v'} \rangle$, from which one can expect $E[v] \approx E_{\mathrm{TF}}[n_v; v] \stackrel{\mathrm{def}}{=} \inf_n E_{\mathrm{TF}}[n; v]$ (because every n is assumed to come from some $\Psi_{v'}$). Purists may instead take the following relation for a definition:

$$E_N^{\mathrm{TF}}[v] = \inf_n \left\{ E_{\mathrm{TF}}[n; v] \,\middle|\, n \in \boldsymbol{L}^{5/3} \cap \boldsymbol{L}^1,\ n(\boldsymbol{r}) \ge 0,\ \int d^3r\, n(\boldsymbol{r}) = N \right\}. \tag{3.8}$$

Here, $E_N^{\mathrm{TF}}[v]$ is the N-particle Thomas-Fermi ground state energy related to the external potential v. If v comes from nuclear charges as in (3.4), then the left-hand side of the last relation depends on the Z_μ and on the nuclear configuration $\{\boldsymbol{R}_\mu\}$. On the right-hand side, $\{a|b\}$ is the usual notation of the set of elements a for which the relations b hold.

In practically all cases of interest, $v(\boldsymbol{r})$ is negative and approaches zero at infinity:

$$v(\boldsymbol{r}) \le 0, \quad \lim_{|\boldsymbol{r}|\to\infty} v(\boldsymbol{r}) = 0. \tag{3.9}$$

If one adds one further particle in this case, then either the potential can bind it, and hence $E_{N+1}^{\mathrm{TF}}[v] < E_N^{\mathrm{TF}}[v]$, or it cannot bind it, the additional particle

disappears at infinity in an $E \to 0$ state, and hence $E_{N+1}^{\mathrm{TF}}[v] = E_N^{\mathrm{TF}}[v]$. In this latter case there is no minimal $n(\boldsymbol{r})$ for $N+1$ particles (wherefore the infimum of (3.8) is in general not a minimum), but in this case there is always a maximum number $N_{\max}$ of particles which the potential v can bind, and the corresponding particle density $n(\boldsymbol{r})$ minimizes $E_{\mathrm{TF}}[n;v]$ for all N equal to or larger than that maximum number of bound particles (the excess particles disappearing at infinity in $E \to 0$ states). Hence, the following minimum relation

$$E_N^{\mathrm{TF}}[v] = \min_n \left\{ E_{\mathrm{TF}}[n;v] \,\middle|\, n \in \boldsymbol{L}^{5/3} \cap \boldsymbol{L}^1,\ n(\boldsymbol{r}) \geq 0,\ \int d^3r\, n(\boldsymbol{r}) \leq N \right\} \tag{3.10}$$

is always true under the condition (3.9), where the sign of equality in the last constraint of (3.8) had to be changed into $\leq$. The minimum is taken on for some n with $\int d^3r\, n = N$, if N is smaller than $N_{\max}$, and it is taken on by some n with $\int d^3r\, n = N_{\max}$ in all other cases.

It is also easy to see that the functional $E_{\mathrm{TF}}[n;v]$ is strictly convex in n for every fixed v and for every non-negative pair potential $w(r)$, i.e.,

$$E_{\mathrm{TF}}[\, cn_1 + (1-c)n_2; v\,] < cE_{\mathrm{TF}}[n_1;v] + (1-c)E_{\mathrm{TF}}[n_2;v], \quad 0 < c < 1. \tag{3.11}$$

This implies that the particle density minimizing the Thomas-Fermi energy functional is unique. This is, of course, a result of approximation because in real quantum physics the ground state may be degenerate, in general implying a degeneracy of several varying ground state densities, too.

Adding the integral condition on the density, multiplied by a Lagrange multiplier μ, to the energy functional (3.3) and varying it with respect to the density yields the Thomas-Fermi equation

$$\frac{5}{3} C_{\mathrm{F}}\, n^{2/3}(\boldsymbol{r}) = \max\{\, \mu - v(\boldsymbol{r}) - v_{\mathrm{H}}(\boldsymbol{r}),\, 0\,\} \stackrel{\mathrm{def}}{=} [\, \mu - v(\boldsymbol{r}) - v_{\mathrm{H}}(\boldsymbol{r})\,]_+\,, \tag{3.12}$$

where, as previously, the Hartree potential

$$v_{\mathrm{H}}(\boldsymbol{r}) = \int d^3r'\, n(\boldsymbol{r}')w(|\boldsymbol{r}' - \boldsymbol{r}|) \tag{3.13}$$

was introduced. If, at some point $\boldsymbol{r}$, $n(\boldsymbol{r}) > 0$, then the minimum condition demands $\delta E_{\mathrm{TF}}/\delta n - \mu = 0$ there. If, however, $n(\boldsymbol{r}) = 0$ for the minimal n,

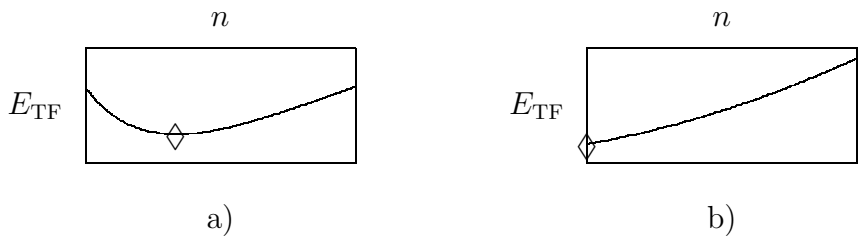

Figure 3: Minimum of $E_{\mathrm{TF}}(n)$ in the domain $0 \leq n \leq \infty$, a) inner point, b) left domain boundary.

then the slope may be positive, $\delta E_{\mathrm{TF}}/\delta n - \mu \geq 0$ at that point $\boldsymbol{r}$, because the density $n(\boldsymbol{r})$ is not allowed to become negative (case of minimum at the boundary of domain of definition; cf. Fig.3, where for the sake of simplicity $\mu = 0$ is assumed).

In the form (3.12), the Thomas-Fermi equation is an integral equation because its solution $n(\boldsymbol{r})$ appears under the integral of v_{H}. The solving density $n(\boldsymbol{r})$ is obtained non-zero everywhere, where $v(\boldsymbol{r}) + v_{\mathrm{H}}(\boldsymbol{r}) < \mu$. Otherwise n is zero. The chemical potential (Fermi level) μ is to be determined such that $\int d^3r\, n = N$. The Thomas-Fermi equation may be transformed into a differential equation for the Hartree potential with the help of Poisson's equation

$$-\nabla^2 v_{\mathrm{H}}(\boldsymbol{r}) = 4\pi n(\boldsymbol{r}). \tag{3.14}$$

With the numerical value of C_{F} from (3.1) and by inserting $n(\boldsymbol{r})$ from (3.12) into (3.14) one finds

$$-\nabla^2 v_{\mathrm{H}}(\boldsymbol{r}) = \frac{8\sqrt{2}}{3\pi}\, [\,\mu - v(\boldsymbol{r}) - v_{\mathrm{H}}(\boldsymbol{r})\,]_+^{3/2}\,. \tag{3.15}$$

Appropriate boundary conditions have to be added of course, which take on a simple form in problems of high symmetry.

Consider the enormous gain of the Thomas-Fermi approximation: Instead of the solution of the N-particle Schrödinger equation, which is an eigenvalue problem in $3N$ coordinates (where N can be arbitrarily large, depending on the case of application), *approximants* of the ground state energy and the ground state density are now obtained from the solution of a single 3-dimensional integral equation, or alternatively from the direct solution of a 3-dimensional variational problem.

3.2 The Thomas-Fermi Atom

Since an excellent textbook treatment of the Thomas-Fermi atom is given in [Landau and Lifshitz, 1977, §70], we keep the discussion brief here.

For a neutral atom,

$$N = Z, \quad v = -\frac{Z}{r}, \quad \nabla^2 v = 4\pi Z \delta(\boldsymbol{r}). \tag{3.16}$$

The Thomas-Fermi equation is solved by a spherical $n(r)$ in this case, and since the minimizing density of the Thomas-Fermi functional is unique, the Thomas-Fermi atom is always spherical. Suppose now that $n(r)$ is zero outside of some radius r_0. Then, for a neutral atom, by ordinary electrostatics,

$$v_{\text{eff}} \stackrel{\text{def}}{=} v + v_{\text{H}} = 0 \text{ for } r > r_0, \tag{3.17}$$

too, because of the neutrality. Hence, from (3.12), $\mu = 0$ in this case. (The argument remains valid for $r_0 \to \infty$.) For a positively charged ion, $v_{\text{eff}} < 0$ for $r > r_0$, and hence $\mu = v_{\text{eff}}(r_0) < 0$. (The right hand side of (3.12) is equal to $\mu - v_{\text{eff}}$ for $r \leq r_0$, where $n(r) \geq 0$, and from $n(r_0) = 0$ follows $\mu = v_{\text{eff}}$.) For a *negatively* charged ion, it is easily seen that a finite r_0 is impossible, because now $v_{\text{eff}} > 0$ would follow from electrostatics for $r > r_0$, where $n(r) = 0$, and hence (3.12) could not be fulfilled. Furthermore, $\mu \neq 0$ is also impossible, since both the effective potential and the density approach zero for $r \to \infty$.

Since we found $\mu = 0$ for $N \geq Z$ (neutral atom or negatively charged ion), and considering (3.16), the Thomas-Fermi equation for this case takes on the form

$$-\nabla^2 v_{\text{eff}} = \frac{8\sqrt{2}}{3\pi}(-v_{\text{eff}})^{3/2} - 4\pi Z \delta(\boldsymbol{r}). \tag{3.18}$$

The δ-function leads to the boundary condition $r v_{\text{eff}} \to -Z$ for $r \to 0$. Furthermore, we already discussed $v_{\text{eff}} \to 0$ for $r \to \infty$. These two conditions determine the solution of (3.18) for v_{eff} uniquely. Since, moreover, $n(r)$ is uniquely obtained from v_{eff} via Poisson's equation, a unique relation $N = N(Z)$ is obtained under the presupposition $N \geq Z$. For $r \neq 0$, (3.18) reads

$$-\frac{1}{r}\frac{d^2}{dr^2} r v_{\text{eff}} = \frac{8\sqrt{2}}{3\pi}(-v_{\text{eff}})^{3/2}, \tag{3.19}$$

and for large r, its solution is $v_{\text{eff}} \sim r^{-4}$. On the other hand, if $N(r)$ denotes the electron number inside the radius r, then Gauss' theorem of

electrostatics yields $Z - N(r) = r^2\, dv_{\rm eff}/dr$. Hence, for the solution of (3.19), $Z - N = Z - N(\infty) = 0$. Summarizing, under the assumption $N \geq Z$ we found $N = Z$. *There is a unique solution of the Thomas-Fermi theory for every neutral atom, but non for negatively charged ions.* For $N < Z$, a finite r_0 is always obtained. *The positively charged Thomas-Fermi ion has a finite radius.*

With the ansatz

$$v_{\rm eff}(r) = -\frac{Z}{r}\,\chi(\alpha r), \quad \alpha = (4\pi)^{2/3} C_{\rm F}^{-1} Z^{1/3} = 1.1295\, Z^{1/3}, \tag{3.20}$$

the Thomas-Fermi equation for the neutral atom is cast into the universal equation

$$\frac{d^2\chi(x)}{dx^2} = \frac{1}{x^{1/2}}[\chi(x)]^{3/2}, \quad \chi(0) = 1, \quad \chi(\infty) = 0, \tag{3.21}$$

which may be solved numerically (Fig.4).

The electron density $n(r)$ is obtained as

$$n(r) = \frac{32}{9\pi^3}\left[\frac{\chi(\alpha r)}{\alpha r}\right]^{3/2} Z^2 \tag{3.22}$$

with the asymptotics

$$\sim r^{-3/2} \overset{0\leftarrow r}{\longleftarrow} n(r) \overset{r\to\infty}{\longrightarrow} \sim r^{-6}. \tag{3.23}$$

(For a positively charged ion, $N < Z$, the density $n(r)$ has a bounded support, i.e., $n(r) = 0$ outside of some radius r_0 as discussed above.)

The Thomas-Fermi energy of the atom (in notation of (3.8) and the text thereafter) is

$$E_Z^{\rm TF}[Z] = -0.7687\, Z^{7/3}. \tag{3.24}$$

Compared to accurate quantum mechanical results, this is by 54% to low for hydrogen (for which the exact value is –0.5=–1Rydberg), by 35% to low for helium, and by about 15% for heavy elements ($Z \approx 100$).

The monotonous function $\chi(x)$ of Fig.4 yields a monotonous charge density with no indication of any shell structure of the Thomas-Fermi atom. The charge density diverges erroneously for $r \to 0$ and has a wrong asymptotic behavior $\sim r^{-6}$, where an atom in quantum mechanics must have an

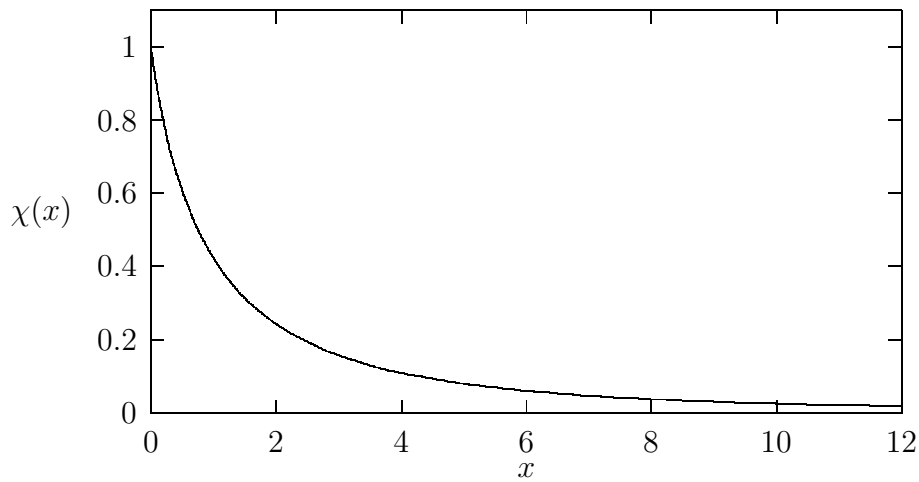

Figure 4: Universal Thomas-Fermi function $\chi(x)$ of neutral atoms.

exponential falloff [Hoffmann-Ostenhof *et al.*, 1980]. (Check that the prefactor of the asymptotics $n(r) \sim r^{-6}$ does not depend on Z; the 'size of the Thomas-Fermi atom' is independent of Z.)

Everything discussed in this section transfers accordingly to molecules, except the universality of a function χ, but including the asymptotics close to nuclei and at infinity, and including the non-existence of negatively charged entities [Lieb, 1981].

3.3 The Thomas-Fermi Screening Length

Consider a homogeneous electron liquid of density $n(\boldsymbol{r}) = n^{(0)} = \text{const.}$ in a positive neutralizing background density $n_+ = n^{(0)}$ so that the effective potential $v_{\text{eff}}^{(0)}(\boldsymbol{r}) = v^{(0)}(\boldsymbol{r}) + v_{\text{H}}^{(0)}(\boldsymbol{r}) = \text{const.}$, too. The external potential v produced by the background charge density, and the Hartree potential v_{H} separately are, of course, not constant in space, since they must fulfill the respective Poisson equations so that their Laplacian derivative is constant, and from (3.14, 3.15)

$$4\pi n^{(0)} = -\nabla^2 v_{\text{H}}^{(0)}(\boldsymbol{r}) = \frac{8\sqrt{2}}{3\pi}[\,\mu - v_{\text{eff}}^{(0)}\,]^{3/2} = \frac{8\sqrt{2}}{3\pi}\varepsilon_f^{3/2}, \tag{3.25}$$

where the Fermi energy ε_f is the distance from the constant effective potential to the chemical potential μ. It is, according to these relations, connected with the density by

$$\varepsilon_f = \frac{1}{2}(3\pi^2 n^{(0)})^{2/3}, \tag{3.26}$$

exactly as in (1.43), from which the Thomas-Fermi kinetic energy functional was derived.

Put now a small additional test charge δZ in the coordinate origin, so that an additional external potential $\delta v = -\delta Z/r$ appears, for which $\nabla^2 \delta v = 0$ and hence $\nabla^2 \delta v_{\mathrm{H}} = \nabla^2 \delta v_{\mathrm{eff}}$ at $\boldsymbol{r} \neq 0$. Thus, now (3.15) yields

$$-\nabla^2 (v_{\mathrm{H}}^{(0)} + \delta v_{\mathrm{eff}}) = \frac{8\sqrt{2}}{3\pi}(\varepsilon_f - \delta v_{\mathrm{eff}})^{3/2} \approx \frac{8\sqrt{2}}{3\pi}\varepsilon_f^{3/2}\left(1 - \frac{3}{2}\frac{\delta v_{\mathrm{eff}}}{\varepsilon_f}\right). \tag{3.27}$$

The last expression has been linearized with respect to the perturbation. Subtraction of (3.25) leads to

$$\nabla^2 \delta v_{\mathrm{eff}} = \frac{4\sqrt{2\varepsilon_f}}{\pi}\delta v_{\mathrm{eff}} \stackrel{\mathrm{def}}{=} \frac{1}{\lambda_{\mathrm{TF}}^2}\delta v_{\mathrm{eff}}. \tag{3.28}$$

As is readily obtained from this relation, the screened potential perturbation has a Yukawa form

$$\delta v_{\mathrm{eff}}(r) \sim \frac{1}{r} e^{-r/\lambda_{\mathrm{TF}}} \tag{3.29}$$

with the Thomas-Fermi screening length λ_{TF}. In truth, (3.29) gives the asymptotics for large r, because linearization in (3.27) was justified for sufficiently large r only. With the electron density parameter r_s defined by

$$\frac{4\pi}{3} r_s^3 \stackrel{\mathrm{def}}{=} n^{-1} \tag{3.30}$$

one finally obtains

$$\lambda_{\mathrm{TF}}^2 = \frac{\pi}{4k_f} = \left(\frac{\pi}{12}\right)^{2/3} r_s \tag{3.31}$$

for the Thomas-Fermi screening length. For ordinary metals, $r_s = 2 \ldots 6$ and hence $\lambda_{\mathrm{TF}} = 0.9 \ldots 1.5$ in natural units.

3.4 Scaling Rules

In this section we analyze the dependence of the functional (3.4) on the parameters Z_μ, $\boldsymbol{R}_\mu$. Using ideas of physical similarity, we scale lengths and densities according to $\boldsymbol{R}_\mu \to \tilde{\boldsymbol{R}}_\mu = \gamma \boldsymbol{R}_\mu$, $n(\boldsymbol{r}) \to \tilde{n}(\boldsymbol{r}) = n(\gamma^{-1}\boldsymbol{r})$, so that density profiles are stretched on the same scale as atom position vectors $\boldsymbol{R}_\mu$. We assume that this length scaling can be compensated by according scalings of density amplitudes, charges and energies with appropriate factors. Thus, assuming

$$E_{\mathrm{TF}}[n(\boldsymbol{r}); Z_\mu, \boldsymbol{R}_\mu] = \eta E_{\mathrm{TF}}[\alpha n(\gamma^{-1}\boldsymbol{r}); \beta Z_\mu, \gamma \boldsymbol{R}_\mu], \tag{3.32}$$

from the four terms of (3.4) the conditions

$$\eta = \gamma^{-3}\alpha^{-5/3} = \gamma^{-2}\beta^{-1}\alpha^{-1} = \gamma^{-5}\alpha^{-2} = \gamma\beta^{-2} \tag{3.33}$$

follow, hence

$$E_{\mathrm{TF}}[n(\boldsymbol{r}); Z_\mu, \boldsymbol{R}_\mu] = \gamma^7 E_{\mathrm{TF}}[\gamma^{-6} n(\gamma^{-1}\boldsymbol{r}); \gamma^{-3} Z_\mu, \gamma \boldsymbol{R}_\mu]. \tag{3.34}$$

With this result, from (3.8, 3.12, 3.13) and (3.17) the following scaling relations are obtained:

$$n(\boldsymbol{r}; Z_\mu, \boldsymbol{R}_\mu) = \gamma^6 n(\gamma\boldsymbol{r}; \gamma^{-3} Z_\mu, \gamma \boldsymbol{R}_\mu), \tag{3.35}$$

$$N(Z_\mu, \boldsymbol{R}_\mu) = \gamma^3 N(\gamma^{-3} Z_\mu, \gamma \boldsymbol{R}_\mu), \tag{3.36}$$

$$v_{\mathrm{eff}}(\boldsymbol{r}; Z_\mu, \boldsymbol{R}_\mu) = \gamma^4 v_{\mathrm{eff}}(\gamma\boldsymbol{r}; \gamma^{-3} Z_\mu, \gamma \boldsymbol{R}_\mu), \tag{3.37}$$

$$\mu(Z_\mu, \boldsymbol{R}_\mu) = \gamma^4 \mu(\gamma^{-3} Z_\mu, \gamma \boldsymbol{R}_\mu), \tag{3.38}$$

$$E_N^{\mathrm{TF}}(Z_\mu, \boldsymbol{R}_\mu) = \gamma^7 E^{\mathrm{TF}}_{\gamma^{-3}N}(\gamma^{-3} Z_\mu, \gamma \boldsymbol{R}_\mu). \tag{3.39}$$

The density minimizing the left side of (3.34) is $n(\boldsymbol{r}; Z_\mu, \boldsymbol{R}_\mu)$, that one minimizing the right side is $n(\boldsymbol{r}; \gamma^{-3} Z_\mu, \gamma \boldsymbol{R}_\mu)$. Eq. (3.34) also says that this latter density is expressed by the former one as $n(\boldsymbol{r}; \gamma^{-3} Z_\mu, \gamma \boldsymbol{R}_\mu) = \gamma^{-6} n(\gamma^{-1}\boldsymbol{r}; Z_\mu, \boldsymbol{R}_\mu)$. This way, (3.35) is obtained. With its use, the remaining relations follow. The result (3.24) is just a special case of the last scaling relation.

There is one key issue of the modern treatment of Thomas-Fermi theory, connected with this type of scaling. If $E_N(Z_\mu, \boldsymbol{R}_\mu)$ is the exact quantum mechanical ground state energy corresponding to an assembly (3.4), i.e., the

infimum of the spectrum of the corresponding N-electron Hamiltonian with Coulomb interactions and the nuclei at fixed positions, then

$$\lim_{\gamma\to 0} \gamma^7 E_{\gamma^{-3}N}(\gamma^{-3}Z_\mu, \gamma \boldsymbol{R}_\mu) = E_N^{\mathrm{TF}}(Z_\mu, \boldsymbol{R}_\mu). \tag{3.40}$$

This is an $N \to \infty$ limit with all $z_\mu = Z_\mu/N$ fixed and interatomic distances uniformly decreasing [Lieb and Simon, 1977, Lieb, 1981]. There is a corresponding relation (in a certain weak topological sense) for the densities [ibid.]. *Thomas-Fermi theory is asymptotically exact for large nuclear charges and electron numbers.* Unfortunately, the convergence of (3.40) is very slow on a scale of the real world (cf. the numbers given after (3.24)).

Another exact relation is obtained by multiplying the Thomas-Fermi equation (3.12) by $n(\boldsymbol{r})$ and integrating over $\boldsymbol{r}$-space. If we denote the three terms of (3.3) *for the minimizing density* in turn by K, U and W, then the result may be expressed as

$$\frac{5}{3}K = \mu N - U - W. \tag{3.41}$$

Equally simply the virial theorem is proved in Thomas-Fermi theory *for an atom*: Let $n(r)$ be the solution of the Thomas-Fermi equation for an atom, and let $n_\lambda(r) = \lambda^3 n(\lambda r)$ (so that N is unchanged when varying λ). Then $E_{\mathrm{TF}}[n_\lambda; v]$ has its minimum at $\lambda = 1$,

$$0 = \left.\frac{\partial E_{\mathrm{TF}}[n_\lambda(r); v(r)]}{\partial \lambda}\right|_{\lambda=1} = 2K - U - W. \tag{3.42}$$

The $n^{5/3}$-dependence ensures the correct scaling behavior of the kinetic energy term K thus establishing the virial theorem.

One could try to derive the virial theorem for molecules, too. However, as another basic defect, *there is no stable molecule in Thomas-Fermi theory.* The Thomas-Fermi energy (3.10) of a molecule is monotonically decreasing for increasing interatomic distances. This is Teller's famous no-binding theorem [Teller, 1962], the mathematically rigorous proof of which again is given in [Lieb, 1981].

3.5 Correction Terms

The comprehensive description (with mathematical rigor) of the state of art with respect to the content of this section may again be found in [Lieb, 1981].

For quantum chemistry applications see also [Parr and Yang, 1989]. We content ourselves with brief comments only.

As was mentioned at the beginning of Section 3.1, Thomas and Fermi did not consider the exchange and correlation energy. According to (2.46) and for a homogeneous electron liquid of density n, this energy may be written as

$$W_{\mathrm{XC}} = \frac{1}{2}\int d^3r d^3r'\, w(|\boldsymbol{r}' - \boldsymbol{r}|) h(\boldsymbol{r}', \boldsymbol{r}) = \int d^3r'\, 2\pi \int_0^\infty dr r^2\, w(r) h(r). \tag{3.43}$$

For the Coulomb interaction $w(r) = 1/r$ and replacing the pair correlation function $h(r)$ with the density times the exchange hole (2.53), the last integral gives

$$\begin{aligned} 2\pi n \int_0^\infty dr r\, h_{\mathrm{X}}(r) \;&= \\ &= -9\pi n^2 \int_0^\infty dr r \left[\frac{\sin k_f r - k_f r \cos k_f r}{(k_f r)^3}\right]^2 = \\ &= -9\pi \frac{n^2}{k_f^2} \int_0^\infty dx x \left[\frac{\sin x - x\cos x}{x^3}\right]^2 = \\ &= -\frac{9\pi}{4}\frac{n^2}{k_f^2} = -C_{\mathrm{X}} n^{4/3}, \qquad C_{\mathrm{X}} = \frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3} = 0.7386. \end{aligned} \tag{3.44}$$

For the final expression, the Fermi momentum $k_f = (3\pi^2 n)^{1/3}$ was replaced by the density expression according to (1.43). In the spirit of the Thomas-Fermi approximation, Dirac [Dirac, 1930] suggested a correction term

$$-C_{\mathrm{X}} \int d^3r\, n^{4/3}(\boldsymbol{r}) \approx W_{\mathrm{X}} \tag{3.45}$$

to be added to the Thomas-Fermi functional.

This Thomas-Fermi-Dirac theory is based on the functional

$$E_{\mathrm{TFD}}[n(\boldsymbol{r}); v(\boldsymbol{r})] = E_{\mathrm{TF}}[n(\boldsymbol{r}); v(\boldsymbol{r})] - C_{\mathrm{X}} \int d^3r\, n^{4/3}(\boldsymbol{r}) \tag{3.46}$$

and leads to the Thomas-Fermi-Dirac equation

$$\frac{5}{3} C_{\mathrm{F}}\, n^{2/3}(\boldsymbol{r}) - \frac{4}{3} C_{\mathrm{X}}\, n^{1/3}(\boldsymbol{r}) = [\,\mu - v(\boldsymbol{r}) - v_{\mathrm{H}}(\boldsymbol{r})\,]_+. \tag{3.47}$$

The functional space for n is the same as previously, the theory, however, is considerably complicated by the negatively curved additional functional term, which spoils the convexity of $E_{\mathrm{TFD}}[n]$. This difficulty is overcome by introducing a complexified functional which leads for a definite class of external potentials, including those of assemblies of point charges, to the same energy minimum E_N^{TFD} and minimizing density n as the original functional $E_{\mathrm{TFD}}[n;v]$ (in cases where the minimizing density exists). Moreover, even for zero external potential v the Dirac exchange term leads to unphysical negative total energy values for densities leaking away in space with small corrugations. There is no minimizing density in this case; the energy approaches $-\varepsilon N$ asymptotically with a certain positive constant ε. Hence, if $N_{\max}$ is the maximum number of particles v can bind (in the considered theory), then for $N > N_{\max}$, the Thomas-Fermi-Dirac theory makes no physical sense.

Unfortunately, the Thomas-Fermi-Dirac theory does not remedy any of the defects of ordinary Thomas-Fermi theory. The total energy of atoms is obtained even lower (since the Dirac term is always negative), there are no negatively charged ions or radicals either, and Teller's no-binding theorem remains in force although neutral atoms or molecules now have a finite radius. All these defects derive from the poor kinetic energy functional (3.2) rather than from the Dirac term. The Dirac term was, however, very effective in a different context (see Section 4.3 below), and in this respect it should be mentioned, that the value of C_{X} given in (3.44) is not sacrosanct since a Hartree-Fock state, from which the exchange hole (2.53) and hence the Dirac term was derived, always overestimates the exchange energy due to neglect of correlations.

The basic deficiency of Thomas-Fermi theory is the poor approximation of the kinetic energy. In order to improve upon it (originally in the problem of nucleon motion in a nucleus), von Weizsäcker [von Weizsäcker, 1935] considered modified plane waves $(1+\boldsymbol{a}\cdot\boldsymbol{r})\exp(i\boldsymbol{k}\cdot\boldsymbol{r})$ in order to have an inhomogeneous situation and found a gradient correction term

$$K_{\mathrm{W}}[n(\boldsymbol{r})] = \frac{1}{2}\int d^3r\,[\nabla n^{1/2}(\boldsymbol{r})]^2 \tag{3.48}$$

to the kinetic energy functional. Later on Kirshnits [Kirshnits, 1957] showed that the correct term in a systematic expansion of the kinetic energy functional is $(1/9)K_{\mathrm{W}}$. One may, however, from different view-points come up with different coefficients λ, $(1/9) \le \lambda \le 1$, in front of K_{W}. Without diving into the details of this theory, we mention that considerable qualitative

improvement is achieved in various respects: The electron density at atomic nuclei is now finite, and at infinity it decays exponentially. Negative ions are formed and molecules may bind. It is, however, up to now not known, to which extent the Thomas-Fermi-λWeizsäcker theory provides accurate correction terms to the asymptotics (3.40).

The gradient expansion of the kinetic energy functional can be followed up to higher orders. We do not consider this here, since it seems, unfortunately, to be a dead road, and the sixth order already diverges. (See [Dreizler and Gross, 1990, Chapter 5.] for a comprehensive and up-to-date survey.)

4 Hohenberg-Kohn Theory

The history of the Thomas-Fermi and Hohenberg-Kohn theories presents an instructive example of the way knowledge is gathered in many-body physics (or in physics in general). Thomas and Fermi introduced their ideas on a naïve and pragmatic level with the aim to test how the structure of heavier atoms and hopefully chemistry may be obtained from the newly developed apparatus of quantum physics, using the numerical means available at that time. After the promising first results, only about thirty years later the absence of molecular binding in numerical results (connected with the rapid post-war growth in numerical capabilities) became a pressing problem, and only 35 years after the foundation of Thomas-Fermi theory it was *analytically* shown by Teller that there is no chemical binding in that theory (without the Weizsäcker term). Partly because of its destructive character, Teller's theorem was doubted on grounds of rigor, and certainly this situation played a role in keeping alive the interest of mathematicians in the theory. Besides confirmation of Teller's theorem, a great number of constructive and very important results have appeared up to now from those activities.

Meanwhile, quantum chemistry had developed on the basis of Hartree-Fock theory. Again a naive approximation had to be introduced due to numerical limitations despite the rapid development of computer techniques. Slater's Xα method was born, this time, however, with an astonishing success sometimes even jealously noticed by many-body theorists.

The Hohenberg-Kohn theory is formal in nature since there is only little more than no hope to aquire rigorous knowledge of the basic ingredient, the Hohenberg-Kohn functional $F[n]$. Nevertheless, with a clever trick contributed by Kohn and Sham, it simultaneously put the Xα method on a much broader theoretical basis than the Hartree-Fock-Slater approximation had been (and hence at least partially explained its success) and opened many ways to generalize both the Thomas-Fermi and the Xα approaches. In that situation, the marriage of this approach with sophisticated many-body theory and the development of the mathematical basis became extremely fruitful. To stress this again, the gap between the formally rigorous part of the theory and its pragmatic approximate versions will probably never be closed. (At present there is a tendency to increase the gap by the rapidly

growing fields of application of approximate versions of the theory.) Nevertheless, the role of the formal part as a guide can hardly be overestimated.

In the whole of this chapter the theory is developed on a heuristic level with focus on the underlying physics. The mathematically strict development comes in Chapter 6.

Monographs and surveys on density functional theory in the understanding of Hohenberg and Kohn appear rather regularly. Here we cite a more or less representative selection: [Dreizler and da Providencia, 1985], [Parr and Yang, 1989], [Trickey, 1990], [Dreizler and Gross, 1990], and more recently [Gross and Dreizler, 1995].

4.1 The Basic Theorem by Hohenberg and Kohn

In this chapter we fix the particle number N and consider Hamiltonians as given in (1.24):[5]

$$\hat{H}[v] = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 + \sum_{i=1}^{N} v(x_i) + \frac{1}{2}\sum_{i\neq j}^{N} w(|\boldsymbol{r}_i - \boldsymbol{r}_j|) = \hat{T} + \hat{U} + \hat{W}, \quad (4.1)$$

where the external potential v is explicitly indicated as a functional variable. The functional dependence of $\hat{H}$ on v is of course affine-linear:

$$\hat{H}[\alpha v_1 + \beta v_2] = \alpha\hat{H}[v_1] + \beta\hat{H}[v_2] \quad \text{for} \quad \alpha + \beta = 1. \quad (4.2)$$

As a reference system,

$$\hat{H}^0[v] = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 + \sum_{i=1}^{N} v(x_i) \quad (4.3)$$

will be needed. In the following, a superscript 0 will systematically refer to the interaction-free case $w \equiv 0$. All Hamiltonians considered will be assumed bounded below, and, for a given interaction w not explicitly indicated, the ground state energy

$$E[v] \stackrel{\text{def}}{=} \inf\{\langle\Psi|\hat{H}[v]|\Psi\rangle \mid \Psi \in \mathcal{W}_N\}, \quad (4.4)$$

$$\begin{aligned}\mathcal{W}_N \stackrel{\text{def}}{=} \{\Psi \quad | \quad & \Psi(x_1 \ldots x_N) \text{ (anti)symmetric,} \\ & \langle\Psi|\Psi\rangle = 1,\ \langle\nabla_i\Psi|\nabla_i\Psi\rangle < \infty \text{ for } i = 1 \ldots N\},\end{aligned} \quad (4.5)$$

[5] Only in the last section the more general potential case $v_{ss'}(\boldsymbol{r})$ due to a general external magnetic field which couples to the spin is considered.

is defined even if there is no ground state $\Psi_0[v]$ minimizing (4.4) (for instance when v cannot bind N particles as discussed for the Thomas-Fermi case before (3.10)). If there is such a ground state (not necessarily unique; degeneracy is permitted), it is obviously gauge invariant with respect to potential constants, i.e., it is the same for all potentials v+const., and

$$E[v + \text{const.}] = E[v] + N \cdot \text{const.} \tag{4.6}$$

Therefore, potentials essentially differing only by such a gauge constant are not considered different in the following, and $v_1 \neq v_2$ means $v_1 - v_2$ is not a.e. a constant.

For the rest of this section we restrict the consideration to *spin independent* potentials. Nevertheless we keep in all what follows the notation $v(x)$ and $n(x)$ with x meaning $\boldsymbol{r}$ in the spin independent case. The implications of spin dependence are discussed in Sections 4.7 and 4.8. We now consider potentials like those discussed in the text around (3.7) and denote this situation as $v \in \oplus \boldsymbol{L}^p$ for some p's ($p = \infty$ allowed), and which do have an N-particle ground state. We define this class of potentials:

$$\mathcal{V}_N \stackrel{\text{def}}{=} \{v \,|\, v \in \oplus \boldsymbol{L}^p \text{ for some } p\text{'s}, \ \hat{H}[v] \text{ has a ground state }\}. \tag{4.7}$$

(Of course, $\mathcal{V}_N$ depends on w: $\mathcal{V}_N^0$ for the interaction-free case is generally different from $\mathcal{V}_N$. E.g. $v = -Z/|\boldsymbol{r}| \in \mathcal{V}_N^0$ for every $N < \infty$: There is an infinite number of interaction-free single-particle states bound by the Coulomb potential of a nucleus of charge Z, their energies clustering towards the continuum edge. However, with $w(r) = 1/r$, $v = -Z/|\boldsymbol{r}| \in \mathcal{V}_N$ for $N \leq N_{\max}(Z) < \infty$ only, where $N_{\max}(Z)$ is the maximum number of repelling each other particles, which the potential $v = -Z/|\boldsymbol{r}|$ can bind.)

Now, for $v \in \mathcal{V}_N$, we have

$$\hat{H}[v]\Psi_0[v] = \Psi_0[v]E[v] \tag{4.8}$$

with at least one ground state $\Psi_0[v]$. In case of degeneracy, $\Psi_0[v]$ denotes in the following any one of the degenerate ground states. Consider $v_1 \neq v_2 \in \mathcal{V}_N$ and suppose $\Psi_0[v_1] = \Psi_0[v_2] \equiv \Psi_0$. Subtracting the two Schrödinger equations from each other yields

$$\sum_i (v_1(x_i) - v_2(x_i))\Psi_0 = \Psi_0(E[v_1] - E[v_2]) = \Psi_0 \cdot \text{const.} \tag{4.9}$$

Since the left side is different from a constant on a domain of non-zero measure, Ψ_0 must be zero there. It is a conjecture without doubt although

not easily proved mathematically, that this cannot be for $v \in \oplus \boldsymbol{L}^p$. ($v \in \oplus \boldsymbol{L}^p$ excludes hard potential barriers behind which $\Psi_0 \equiv 0$ would be possible.) Hence, $\Psi_0[v_1] \neq \Psi_0[v_2]$. Since $\Psi_0[v_i]$ is non-zero where the potentials are different, $\Psi_0[v_2]$ does not satisfy the Schrödinger equation for v_1, and hence a strict inequality

$$E[v_1] < \langle \Psi_0[v_2] | \hat{H}[v_1] | \Psi_0[v_2] \rangle = E[v_2] + \int dx\, n[v_2](v_1 - v_2) \tag{4.10}$$

holds, where $n[v]$ is the particle density corresponding to the ground state $\Psi_0[v]$. The same strict inequality holds with subscripts 1 and 2 reversed.

Suppose that still $n[v_1] = n[v_2]$ a.e. Then, adding the inequality (4.10) and that obtained by interchanging subscripts 1 and 2 yields

$$E[v_1] + E[v_2] < E[v_2] + E[v_1], \tag{4.11}$$

which is a contradiction. Hence,

$$n[v_1] \neq n[v_2] \text{ for } v_1 \neq v_2, \tag{4.12}$$

or, in other words, for every given $n(x)$ (taken as function on the whole x-space) there is at most one potential function $v(x)$mod(const.) for which $n(x)$ is the ground state density. This is the basic theorem by Hohenberg and Kohn:

> *$v(x)$mod(const.) $\in \mathcal{V}_N$ is a unique function of the ground state density $n(x)$.*

($v(x)$mod(const.) is a family of potentials, $\mathcal{V}_N$ is the set of such families, and it is one of these families which is uniquely defined by $n(x)$.) For this reason, $\mathcal{V}_N$ may be called the set of n-representable potentials, although this name is not much in use in the literature.

The use of the theorem is to transfer every functional dependence on v into a functional dependence on n by substituting $v[n]$. To this end we have to define the class of densities

$$\mathcal{A}_N \stackrel{\text{def}}{=} \{n(x) \,|\, n \text{ comes from an } N\text{-particle ground state}\}, \tag{4.13}$$

being called the class of pure-state v-representable densities in the literature. Now, define the density functional by Hohenberg and Kohn as

$$F_{\text{HK}}[n] \stackrel{\text{def}}{=} E[v[n]] - \int dx\, v[n]n, \quad n \in \mathcal{A}_N. \tag{4.14}$$

It is clear from (4.6) that the undefined potential constant drops out on the r.h.s. of (4.14); $F_{\mathrm{HK}}[n]$ is uniquely and well defined on $\mathcal{A}_N$. Now consider $F_{\mathrm{HK}}[n]+\int dx\, nv$ as a functional of the two *independent* variables n and v (v in the integral not necessarily being $v[n]$). Let $n \in \mathcal{A}_N$. This implies that there is some potential $v_n = v[n]$ for which n is a ground state density. Pick $v \in \mathcal{V}_N$ independently. Then, $F_{\mathrm{HK}}[n] + \int dx\, nv = F_{\mathrm{HK}}[n] + \int dx\, nv_n + \int dx\, n(v - v_n) = E[v_n] + \int dx\, n(v - v_n) = \langle \Psi_0[v_n]|\hat{H}[v_n]|\Psi_0[v_n]\rangle + \int dx\, n(v - v_n) = \langle \Psi_0[v_n]|\hat{H}[v]|\Psi_0[v_n]\rangle \geq E[v]$. This is just the same argument as used before (3.8), now based on rigor. Here, the issue is the Hohenberg-Kohn variational principle

$$E[v] = \min_{n\in\mathcal{A}_N} \left\{ F_{\mathrm{HK}}[n] + \int dx\, nv \right\} \tag{4.15}$$

for $v \in \mathcal{V}_N$.

Recall that in these considerations neither $\Psi_0[v]$ nor $n[v]$ were supposed unique. The ground state may be degenerate, and so may be the ground state density. The basic theorem states the mapping $v = v[n]$ to be single-valued, and that is enough for the whole theory. This point was first made by Lieb [Lieb, 1983] (cf. also [Kohn, 1985].) In the original paper [Hohenberg and Kohn, 1964] the analysis was confined to the class $\mathcal{V}'_N$ of potentials having a *non-degenerate* ground state and to densities $n \in \mathcal{A}'_N$ coming from a non-degenerate ground state. In that case, the mapping between n and v is one-to-one, and, consequently, there is even a one-to-one mapping between n and Ψ_0. This makes the alternative definition

$$F_{\mathrm{HK}}[n] = \langle \Psi_0[n]|\hat{T} + \hat{W}|\Psi_0[n]\rangle, \quad n \in \mathcal{A}'_N \tag{4.16}$$

of the Hohenberg-Kohn density functional possible, which of course coincides with (4.14) on the narrower class $\mathcal{A}'_N \subset \mathcal{A}_N$. As degeneracy of the ground state is quite common in physics and the theory makes little profit from the mappings being one-to-one, (4.14) should be taken as the basic definition of the Hohenberg-Kohn density functional, with the restriction to non-degenerate ground states released. Note, however, the discussion at the very end of this chapter.

The really serious problems remaining are connected with the fact that neither the classes $\mathcal{V}_N$ and $\mathcal{A}_N$ nor the functional $F_{\mathrm{HK}}[n]$ are known explicitly. From an information theoretical point of view, one even could say that nothing was gained by the variational principle (4.15), because a guess of $F_{\mathrm{HK}}[n]$ seems equally hopeless as a direct guess of $E[v]$, both functionals

being no doubt extremely involved. On the other hand, the very simple guess

$$F_{\mathrm{HK}}[n] \approx C_{\mathrm{F}} \int d^3r\, n^{5/3}(\boldsymbol{r}) + \frac{1}{2} \int d^3r d^3r'\, n(\boldsymbol{r}')w(|\boldsymbol{r}' - \boldsymbol{r}|)n(\boldsymbol{r}) \quad (4.17)$$

in connection with the variational principle (4.15) yields precisely the Thomas-Fermi theory, which was shown in Chapter 3 to produce a number of encouraging estimates. Also the corrected versions of Section 3.5 of Thomas-Fermi theory all fit in the frame of Hohenberg-Kohn theory.

We are faced with two principally different types of theory in this field: The Hohenberg-Kohn variational principle and its subsequently considered versions are rigorously based on quantum theory but are not given explicitly. Thomas-Fermi theory, its variants, and the related local-density approximation considered below are explicitly given variational principles, but only a few rigorous statements can be made about their connection to exact quantum theory, the strongest being (3.40). It is, however, a general experience in physics that a guess in a variational expression can be much more fruitful than a direct guess of a quantity to be estimated. The recent developments in density functional theory may serve as a brilliant example for this experience.

4.2 The Kohn-Sham Equation

The real break-through in modern density functional theory, which put it at once on one level with and tightly linked it to the Hartree-Fock-Slater approximation of many-fermion theory—but meanwhile led far beyond—, came from Kohn and Sham [Kohn and Sham, 1965] with the suggestion of the Kohn-Sham equation. Recall that the weakest part of Thomas-Fermi theory was the treatment of the kinetic energy functional, and we are now going to explain the Kohn-Sham trick in handling this part.

The considerations of last section can be carried through for any reasonable w (so that the considered Hamiltonians are bounded below). Particularly for $w \equiv 0$ the Hohenberg-Kohn functional

$$T[n] \stackrel{\mathrm{def}}{=} E^0[v^0[n]] - \int dx\, v^0[n]n, \quad n \in \mathcal{A}_N^0 \quad (4.18)$$

is just the kinetic energy of the ground state of the *interaction-free* N-particle system as a functional of the ground state density. Even this functional is not explicitly known, but its existence (i.e. the property of the mapping

$n \mapsto T$ to be single-valued) is again guaranteed by the Hohenberg-Kohn theorem. Among the ground states of an interaction-free fermion system in an external field v there are always determinantal states (1.15). In case of degeneracy, linear combinations of degenerate determinantal states may also serve as ground states. A subtlety here is, that a density derived from a linear combination of degenerate determinantal ground states may not necessarily be derivable from a single determinantal ground state [Lieb, 1983]. Therefore, slightly deviating from the general scheme, we define

$$\mathcal{A}_N^0 \stackrel{\text{def}}{=} \{n(x) \,|\, n \text{ comes from a } \textit{determinantal } N\text{-particle ground state}\}. \tag{4.19}$$

as the domain of $T[n]$.

The density of a determinantal state is

$$n(x) = \sum_{i=1}^{N} \phi_i(x)\phi_i^*(x)$$

$$\text{for} \quad \Psi_0^0(x_1 \dots x_N) = \frac{1}{\sqrt{N!}} \det \|\phi_i(x_k)\|, \quad \langle \phi_i | \phi_j \rangle = \delta_{ij}. \tag{4.20}$$

Its kinetic energy is

$$E_{\text{kin}}^0 = \langle \Psi_0^0 | \hat{T} | \Psi_0^0 \rangle = -\frac{1}{2} \sum_{i=1}^{N} \langle \phi_i | \nabla^2 | \phi_i \rangle. \tag{4.21}$$

Hence,

$$T[n = \sum \phi_i \phi_i^*] = -\frac{1}{2} \sum_{i=1}^{N} \langle \phi_i | \nabla^2 | \phi_i \rangle \tag{4.22}$$

under the orthonormality conditions of (4.20) and under the condition that the orbitals $\phi_i, i = 1, \dots, N$ form a determinantal ground state. In other words,

$$T[n] = \min_{\phi_i^*, \phi_i} \left\{ -\frac{1}{2} \sum_{i=1}^{N} \langle \phi_i | \nabla^2 | \phi_i \rangle \;\middle|\; \langle \phi_i | \phi_j \rangle = \delta_{ij}, \; \sum_{i=1}^{N} \phi_i \phi_i^* = n \right\}. \tag{4.23}$$

For n fixed, the minimum of E^0 is taken on at minimal E_{kin}^0. Actually this expression for T is already a continuation of the definition (4.18) beyond $\mathcal{A}_N^0$,

since it also gives a value for T when no potential $v^0[n]$ exists for that density n. We come back to this type of generalization in a more general context in the next but one section. The relation (4.23) may now be inserted into the Hohenberg-Kohn variational principle (4.15), which in the considered case reads

$$E^0[v] = \min_n \left\{ T[n] + \int dx\, nv \right\}. \tag{4.24}$$

The density n in the last integral may also be replaced with the conditions of (4.23), $\int dxnv = \sum \langle \phi_i | v | \phi_i \rangle$, $\langle \phi_i | \phi_j \rangle = \delta_{ij}$, and then, instead of a two-step minimization, first of (4.23) and then of (4.24), the minimum may likewise be sought in one step:

$$E^0[v] = \min_{\phi_i^*,\, \phi_i} \left\{ \sum_{i=1}^{N} \left(-\frac{1}{2} \langle \phi_i | \nabla^2 | \phi_i \rangle + \langle \phi_i | v | \phi_i \rangle \right) \;\middle|\; \langle \phi_i | \phi_j \rangle = \delta_{ij} \right\}. \tag{4.25}$$

Now, after introducing Lagrange multipliers ε_i for the side conditions of orbital normalization (orthogonality will automatically be provided), each term of the i-sum may be varied separately which yields just the one-particle Schrödinger equation

$$\left(-\frac{\nabla^2}{2} + v(x) \right) \phi_i(x) = \phi_i(x)\, \varepsilon_i \tag{4.26}$$

for the N orbitals lowest in energy of non-interacting electrons.

In the interacting case $w \not\equiv 0$, we decompose the Hohenberg-Kohn functional according to

$$F_{\mathrm{HK}}[n] = T[n] + E_{\mathrm{H}}[n] + E_{\mathrm{XC}}[n], \quad n \in \mathcal{A}_N, \tag{4.27}$$

with the Hartree energy E_{H} as introduced in (2.46) (its variation with respect to the density n giving the Hartree potential (3.13)). This decomposition *defines* E_{XC} as a functional of n on the considered domain, since all the remaining terms of (4.27) were already defined as functionals of n on that domain. (The Hartree energy (2.46) was defined for any density n, and the domain of definition of $T[n]$ by (4.23) is a whole functional space containing all of $\mathcal{A}_N$ as will be discussed in Section 4.4. Since, however, $F_{\mathrm{HK}}[n]$ was only defined on $\mathcal{A}_N$, E_{XC} is so far likewise only defined on that set of densities.) Compare the corresponding decomposition (2.65).

Now, to proceed one replaces $T[n]$ with the r.h.s. of (4.23) and n in $E_{\rm H}$ and in $E_{\rm XC}$ with (4.20). Then one can again combine the variations of (4.23) and of (4.15) into a one-step variation

$$E[v] = \min_{\phi_i^*,\,\phi_i} \left\{ \sum_{i=1}^{N} \left(-\frac{1}{2}\langle\phi_i|\nabla^2|\phi_i\rangle + \langle\phi_i|v|\phi_i\rangle \right) + E_{\rm H}[n] + E_{\rm XC}[n] \;\middle|\; \langle\phi_i|\phi_j\rangle = \delta_{ij} \right\}. \tag{4.28}$$

This time, however, some functional analysis is necessary. The variation of $E_{\rm XC}[n]$ with respect to a variation of ϕ_i^* may be expressed via (4.20), the latter implying $\delta n(x)/\delta\phi_i^*(x') = \delta(x-x')\phi_i(x')$:

$$\frac{\delta E_{\rm XC}}{\delta\phi_i^*(x)} = \int dx' \, \frac{\delta E_{\rm XC}}{\delta n(x')} \frac{\delta n(x')}{\delta\phi_i^*(x)} = \frac{\delta E_{\rm XC}}{\delta n(x)}\phi_i(x). \tag{4.29}$$

(In the existing literature the whole approach is often very sketchy represented by *formally* treating the kinetic energy this way:

$$-\frac{\nabla^2}{2}\phi_i(x) = \frac{\delta E^0_{\rm kin}}{\delta\phi_i^*(x)} = \int dx' \, \frac{\delta T}{\delta n(x')} \frac{\delta n(x')}{\delta\phi_i^*(x)} = \frac{\delta T}{\delta n(x)}\phi_i(x). \tag{4.30}$$

This chain of equations does not mean that $\delta T/\delta n(x) = -\nabla^2/2$, since it does not hold for every function $\phi_i(x)$ and for every variation $\delta\phi_i^*(x)$ of a complete set. It just holds for those orbitals $\phi_i(x)$ which make up the determinantal *ground state* with density $n(x)$ and under an integral with $\delta\phi_i^*(x)$ orthogonal to all of them. Such a sketchy presentation of course requires a certain knowledge of convex analysis on the side of the reader. The simple but essential point is that given n, $T[n]$ is the *minimum* (4.23) of (4.22) rather then (4.22) itself for a general set of ϕ_i.)

The derivations given here still are a bit formal and on an heuristic level. On this level on may orient by partial derivatives and use the following formal correspondences:

$$\begin{aligned}
n_i,\ i = 1\ldots M \quad &: \quad n(x),\ x \in \boldsymbol{R}^3, \\
\frac{\partial n_i}{\partial n_j} = \delta_{ij} \quad &: \quad \frac{\delta n(x)}{\delta n(y)} = \delta(x-y), \\
\frac{\partial}{\partial n_i}\sum_j a_j n_j = a_i \quad &: \quad \frac{\delta}{\delta n(x)}\int dx' \, f(x')n(x') = f(x), \\
\frac{\partial}{\partial\phi_k} f(n_i(\phi_k)) = \sum_i \frac{\partial f}{\partial n_i}\frac{\partial n_i}{\partial\phi_k} \quad &: \quad \frac{\delta}{\delta\phi(x)}F[n[\phi]] = \int dx' \, \frac{\delta F}{\delta n(x')}\frac{\delta n(x')}{\delta\phi(x)}.
\end{aligned} \tag{4.31}$$

(The third line is just a special case of Euler's lemma for the first variation: $(\delta/\delta n(x)) \int dx' f(x') n(x') = 0 \Rightarrow f(x) \equiv 0$. Generally, the linear increment (total differential) $\delta F[\delta n] = \int dx\, (\delta F/\delta n(x))\, \delta n(x)$ is to be considered as a linear functional of δn, which may be viewed as represented by some (generalized) integral kernel $\delta F/\delta n(x)$ being often a generalized function as e.g. a δ-function. A correct definition of functional derivatives will be given in Section 5.7.)

The variation of (4.28) is now straightforward. Again the side conditions are treated with Lagrange multipliers. As the resulting equation will turn out to be Hermitian, again one needs only to consider the normalization of orbitals ϕ_i with multipliers ε_i while the orthogonality will automatically be provided by the Hermiticity of the resulting Kohn-Sham equation:

$$\left(-\frac{\nabla^2}{2} + v_{\text{eff}}(x) \right) \phi_i(x) = \phi_i(x)\, \varepsilon_i, \quad v_{\text{eff}} \stackrel{\text{def}}{=} v + v_{\text{H}} + v_{\text{XC}}, \tag{4.32}$$

where the Kohn-Sham exchange and correlation potential is defined as

$$v_{\text{XC}}(x) \stackrel{\text{def}}{=} \frac{\delta E_{\text{XC}}}{\delta n(x)}. \tag{4.33}$$

Multiplying (4.26) by $\phi_i^*(x)$, integrating, summing over i and considering (4.20, 4.21) yields immediately

$$E^0[v] = \sum_{i=1}^{N} \varepsilon_i, \quad \varepsilon_1 \le \varepsilon_2 \le \ldots, \tag{4.34}$$

where for the minimum of (4.15) in the considered case clearly the orbitals with the lowest N eigenvalues ε_i have to be taken. (In case of degeneracy $\varepsilon_N = \varepsilon_{N+1} = \ldots$ this choice is not unique, see also Section 4.5.)

Note however that, by definition, $T[n]$ in (4.27) is further on the kinetic energy of an interaction-free ground state with density n (and hence at best corresponding to an external potential $v^0[n]$, different from v). The Kohn-Sham E_{XC} contains the change in kinetic energy due to interaction and can be expressed through a coupling constant integral over the w-term as given in (2.67, 2.68).

A consideration analogous to that which led to (4.34) now yields

$$E[v] \le \sum_{i=1}^{N} \varepsilon_i - E_{\text{H}}[n] - \int dx\, n v_{\text{XC}} + E_{\text{XC}}[n], \qquad n(x) = \sum_{i=1}^{N} |\phi_i(x)|^2. \tag{4.35}$$

The $\leq$ sign was needed here, because the right side can only be the minimum of (4.15) over $\mathcal{A}_N \cap \mathcal{A}_N^0$: $E_{\mathrm{XC}}[n]$ was only defined on $\mathcal{A}_N$, and the solution of (4.32) yields via (4.20) a density out of $\mathcal{A}_N^0$. There is another principal difference between the equations (4.26) and (4.32): (4.26) is a linear Schrödinger equation, and the eigenenergies ε_i and eigenfunctions ϕ_i depend only on the external potential v. (4.32) on the other hand is a nonlinear problem, because the effective potential v_{eff} depends on the Kohn-Sham orbitals ϕ_i. Hence it must be solved iteratively (self-consistently) like, e.g., the Hartree-Fock problem. As one consequence, the same level-crossing problems as in Hartree-Fock theory may appear (cf. the text after (1.62)). Note that such a level crossing situation implies that the non-interacting reference ground state is degenerate; one more reason not to rely on non-degeneracy of the ground states. To cope with such situations we will need another generalization of the theory which will be considered in Section 4.5. Yet, in many applications (with approximate functional E_{XC}, of course), a self-consistent solution of (4.32) is found with the Kohn-Sham orbitals for the lowest N levels ε_i occupied. In that case, automatically $n \in \mathcal{A}_N^0$, since an external potential $v^0[n]$ equal to the self-consistent v_{eff} would, for an interaction-free fermion system, give just that ground state density. *Hence, every n obtained via* (4.20) *from a solution of the Kohn-Sham equations is in $\mathcal{A}_N^0$.* Since $\mathcal{A}_N \subseteq \mathcal{A}_N^0$ has not been proved, possibly not every density $n \in \mathcal{A}_N$ can be obtained via (4.20) from a solution of the Kohn-Sham equations. Solving a variational problem by means of Euler's equation needs a careful investigation of the existence of the functional derivatives which cannot be provided at this stage since we do not even know the topology of the domains $\mathcal{A}_N$ and $\mathcal{A}_N^0$.

4.3 The Link to the Hartree-Fock-Slater Approximation

The direct solution of the Hartree-Fock equations (1.55) for large systems ($N > 10^3$) is limited by the involved structure of the exchange potential operator (1.57). To overcome this limitation, albeit with approximate results (being aware that Hartree-Fock theory itself is an approximation, although giving definite (upper) bounds for the total energy), [Slater, 1951] proposed to estimate the exchange potential term in the spirit of Thomas and Fermi. While Slater used a Fermi surface average of the exchange potential for the homogeneous system, [Gáspár, 1954] proposed to approximate the exchange energy term of (1.54) by the Dirac expression (3.45), with the rules (4.31)

yielding

$$v_{\mathrm{X}}(\boldsymbol{r}) = -\frac{4}{3} C_{\mathrm{X}}\, n^{1/3}(\boldsymbol{r}). \tag{4.36}$$

The potential expression originally used by Slater had a factor 2 instead of 4/3 in front.

Since both approaches have their justifications in particular contexts, and, as was soon emphasized by Slater, correlation modifies the effect of exchange, generally reducing it, a parameterization

$$v_{\mathrm{X}\alpha}(\boldsymbol{r}) = -2\alpha C_{\mathrm{X}}\, n^{1/3}(\boldsymbol{r}) \tag{4.37}$$

became very popular in the sixties and seventies and was used with great success in atomic, molecular and solid state calculations of the electronic state. A most popular variant of this X_α-approach was to determine α so that the virial equation $\langle \hat{H} \rangle = -\langle \hat{T} \rangle$ was fulfilled (cf. (2.78)).

The Hartree-Fock-Slater equation is obtained by replacing the exchange term of (1.55) with the local potential term using (4.37). Comparing it with the Kohn-Sham equation (4.32), one observes immediately that the former fits into the latter frame, if one understands (4.37) as an approximation to the Kohn-Sham exchange and correlation potential or, equivalently, if one understands $3\alpha/2$ times the Dirac expression (3.45) as an approximation to $E_{\mathrm{XC}}[n]$. With simple approximate expressions of that type for $E_{\mathrm{XC}}[n]$ (termed local density approximations), Hohenberg-Kohn-Sham theory is computationally as equally simple as the Hartree-Fock-Slater approximation.

Note, however, the big conceptual difference between the two approaches: Hartree-Fock theory searches for the best determinantal approximation to the ground state, whereas Hohenberg-Kohn theory searches for the best approximation to the density of the interacting ground state the latter of which cannot be a determinant. In particular, there is no virial theorem connecting $E[v[n]]$ with $T[n]$, since T contains only part of the kinetic energy, part being contained in E_{XC}. (It is not difficult to show that $\langle \hat{T} \rangle - T[n] \geq 0$ [Levy, 1982, cf. also (4.45)].) This explains why the α-value in (4.37) which gives the lowest total energy for an actual case, was always found to be different from the value that satisfied the virial equation.

4.4 Constrained Search Density Functionals

Hohenberg-Kohn theory as described up to here, although alluring, leaves one uncomfortable with v-representability problems: How relevant is $\mathcal{A}_N \cap$

$\mathcal{A}_N^0$? Both sets of search are unknown, but it *is* known that they are not convex. This last statement might pose the most severe problem (cf. next chapter).

A generalization which circumvents the mapping of ground state densities n on potentials v was independently considered by Levy and Lieb [Levy, 1982, Lieb, 1983]. Define instead of the Hohenberg-Kohn functional the Levy-Lieb functional

$$F_{\mathrm{LL}}[n] \stackrel{\mathrm{def}}{=} \inf\{\, \langle\Psi|\hat{T}+\hat{W}|\Psi\rangle \mid \Psi \mapsto n,\ \Psi \in \mathcal{W}_N \,\}. \tag{4.38}$$

Here, the infimum search is over *all* N-particle wavefunctions (not only ground states) *yielding a given density* $n(x)$. Since $E[v] = \inf\{\langle\Psi|\hat{H}|\Psi\rangle \mid \Psi \in \mathcal{W}_N\} = \inf\{\langle\Psi|\hat{T}+\hat{W}|\Psi\rangle + \int dx\, vn[\Psi] \mid \Psi \in \mathcal{W}_N\}$, F_{LL} can trivially replace F_{HK} in (4.15).

Now the question arises, what characterizes a density n coming from an N-particle wavefunction. Fortunately, this question has got a simple final answer [Gilbert, 1975, Harriman, 1980, Lieb, 1983]: Any non-negative density integrating to N and such that $\int dx\, |\nabla n^{1/2}(\boldsymbol{r})|^2 < \infty$ (i.e. $\nabla n^{1/2} \in \boldsymbol{L}^2$, eventually for each spin component) comes from a $\Psi \in \mathcal{W}_N$, in the fermion case even from a determinantal one. Lieb proved the statement in both directions, i.e., every N-particle wavefunction has a density with the asserted properties. Hence, defining the set of N-representable densities

$$\mathcal{J}_N \stackrel{\mathrm{def}}{=} \{\, n \mid n(x) \geq 0,\ \nabla n^{1/2} \in \boldsymbol{L}^2,\ \int dx\, n = N \,\}, \tag{4.39}$$

we have

$$E[v] = \inf_{n \in \mathcal{J}_N} \left\{ F_{\mathrm{LL}}[n] + \int dx\, nv \right\}. \tag{4.40}$$

instead of (4.15).

The immediate gain is that $\mathcal{J}_N$ is explicitly known. It is even convex, since it is directly seen that $cn_1 + (1-c)n_2 \in \mathcal{J}_N,\ 0 \leq c \leq 1$, if $n_1,\ n_2 \in \mathcal{J}_N$. Moreover, the statement of Gilbert, Harriman and Lieb implies that

$$\mathcal{A}_N \subset \mathcal{J}_N, \quad \mathcal{A}_N^0 \subset \mathcal{J}_N. \tag{4.41}$$

Equality of the sets is excluded here because it is known due to Lieb that $\mathcal{A}_N$ and $\mathcal{A}_N^0$ are not convex. A further formally important statement by Lieb is that in (4.38) there is always a minimizing $\Psi \in \mathcal{W}_N$, i.e.,

$$F_{\mathrm{LL}}[n] = \min\{\, \langle\Psi|\hat{T}+\hat{W}|\Psi\rangle \mid \Psi \mapsto n,\ \Psi \in \mathcal{W}_N \,\}, \quad n \in \mathcal{J}_N. \tag{4.42}$$

The minimizing Ψ need not be unique, and among the minimizing Ψ's there need not be a ground state (for instance, if n is not a ground state density). From (4.40) (end hence from (4.38)) it follows immediately, however, that, if n *is* a ground state density, then a ground state minimizes (4.42):

$$F_{\mathrm{LL}}[n] = F_{\mathrm{HK}}[n] \;\text{ for }\; n \in \mathcal{A}_N. \tag{4.43}$$

Thus F_{LL} is just a continuation of F_{HK} from $\mathcal{A}_N$ onto a convex and explicitly given domain. Unfortunately, $F_{\mathrm{LL}}[n]$ itself is not convex on $\mathcal{J}_N$ [Lieb, 1983].

Again, as in the original Hohenberg-Kohn version, everything said so far holds true for any reasonable interaction w, particularly also for $w \equiv 0$. So,

$$T_{\mathrm{LL}}[n] = \min\{\, \langle\Psi|\hat{T}|\Psi\rangle \mid \Psi \mapsto n,\; \Psi \in \mathcal{W}_N \,\}, \quad n \in \mathcal{J}_N. \tag{4.44}$$

is another continuation of $T[n]$ from (4.18) onto $\mathcal{J}_N$. (As we now know that every $n \in \mathcal{J}_N$ comes from a determinantal state, the domain of definition of (4.23) is $\mathcal{J}_N$, too.) On this domain, although for every n there is a determinantal Ψ it comes from, nevertheless, since n does not uniquely determine Ψ, the minimizing Ψ for (4.44) need not be a determinant. However, if $n \in \mathcal{A}_N^0$ then there is a minimizing determinant Ψ; and in that case (4.44) coincides with both (4.18) and (4.23). Of course, for a general density, $T_{\mathrm{LL}}[n] \le T[n]$, since the search of (4.23) is more restricted than that of (4.44). Now, the same decomposition (4.27) (further with T, not with T_{LL}) can be made for F_{LL} yielding the Kohn-Sham equations. This time, E_{XC} is already defined on $\mathcal{J}_N$ because now all the other entries in (4.27) are defined on that domain. Hence, with F_{LL} we are now more comfortable with the Hohenberg-Kohn variational principle, but not yet with the Kohn-Sham equations which further on yield densities $n \in \mathcal{A}_N^0$ only.

There is one simple but maybe important side product of (4.44). Given a ground state density n for $\hat{H}$ (possible with interaction), the kinetic energy in that ground state is $\langle\hat{T}\rangle = \langle\Psi_0|\hat{T}|\Psi_0\rangle$, where the ground state wavefunction Ψ_0 minimizes $\langle\hat{H}\rangle$ (and gives n). For that n, $T_{\mathrm{LL}}[n] = \langle\Psi|\hat{T}|\Psi\rangle$, where Ψ minimizes $\langle\hat{T}\rangle$ among all wavefunctions giving n. Consequently,

$$\langle\Psi_0|\hat{T}|\Psi_0\rangle \ge T_{\mathrm{LL}}[n[\Psi_0]], \tag{4.45}$$

where $n[\Psi_0]$ denotes the density of the ground state Ψ_0.

4.5 Ensemble State Density Functionals

Instead of considering quantities coming from N-particle pure states Ψ the theory may be generalized to mixed states given by N-particle density matrices. Such a generalization was first introduced in order to treat

non-zero temperatures with density functional theory [Mermin, 1965]. For zero temperature, a step in this direction was made by [Janak, 1978] with the introduction of fractional orbital occupation numbers, and later on linked to the general mixed state theory [Perdew and Zunger, 1981]. With the possibility of fractional orbital occupation numbers this generalization also increases the scope of applicability of the Kohn-Sham equations [Englisch and Englisch, 1984a, b].

An admissible N-particle density matrix, i.e. an ensemble state, has the form

$$\gamma_N(x_1 \dots x_N; x_1' \dots x_N') = \sum_{K=1}^{\infty} \Psi_K(x_1 \dots x_N) g_{N;K} \Psi_K^*(x_1' \dots x_N'),$$
$$0 \le g_{N;K}, \qquad \sum_{K=1}^{\infty} g_{N;K} = 1 \tag{4.46}$$

(cf. e.g. [Parr and Yang, 1989, Section 2.2]). The map of density matrices onto densities n is always linear (an advantage against wavefunctions), hence the density coming from the above γ_N is (in accordance with (2.1–2.3))

$$n(x) = \sum_{K=1}^{\infty} g_{N;K} n_K(x), \tag{4.47}$$

where $n_K(x)$ comes from the pure state Ψ_K. Notice that $n_K \in \mathcal{J}_N$, $\mathcal{J}_N$ is convex, and (4.47) is obviously an affine combination, hence $n \in \mathcal{J}_N$ too.

Admitting ensemble states increases the variational freedom and eventually increases the set of ground states, incorporating mixtures of degenerate pure ground states; for the *ground state energy* it does not imply any change. The ensemble-state expectation value of the N-particle Hamiltonian is

$$\langle \hat{H} \rangle = \operatorname{tr} \hat{H} \gamma_N = \sum_K g_{N;K} \langle \hat{H} \rangle_K, \tag{4.48}$$

where $\langle \hat{H} \rangle_K$ is the pure-state expectation value for the state Ψ_K. The ground state energy is further on defined as

$$E[v] = \inf_{\gamma_N} \langle \hat{H}[v] \rangle. \tag{4.49}$$

The expectation value (4.48) depends linearly on the mixing coefficients $g_{N;K}$, hence, as any linear function, if it has an extremum, it takes on its

extremum on the boundary of the domain of definition of those $g_{N;K}$, given by the last two relations (4.46). If one depicts the $g_{N,K}$ as the K-components of a real Euclidean vector, this domain forms a simplex with corners $g_{N;K} = \delta_{K\tilde{K}}$, $\tilde{K} = 1, 2, \ldots$. (A simplex is a polyhedron with minimum number (dimension+1) of corners: a point, stretch, triangle, tetrahedron,....) The boundary of a simplex consists of simplices of lower dimension; on each of those the infimum of the linear function is again taken on on its boundary. Hence it is taken on on a corner, i.e. on a pure state. If there are several degenerate pure ground states, the infimum is taken on by any state of the simplex spanned by those pure ground states, i.e. having them as its corners. (It is not difficult to realize that, if the function has no minimum but a *finite* infimum, then there must be a sequence of corners, that is of pure states, whose function values converge to that infimum.)

Note the principal difference compared to the expansion of a pure state into a basis, where the expectation values are *bilinear* in the expansion coefficients, and none of the above considerations apply. The expansion of a pure state into a basis compares to a *coherent* superposition of waves, while a mixed state compares to an *incoherent* superposition; in (4.48) no interference terms between pure states appear.

There are physical reasons to include ensemble states into consideration even for the ground state, that is, for zero temperature. Consider for instance a boron atom. Its ground state can be characterized by the configuration $(1s)^2(2s)^2(2p)$. We have in mind the many-particle ground state and use the orbital configuration for its approximate characterization only. It has the orbital angular momentum quantum number (which is a good quantum number for the many-particle state of the interacting electrons) equal to $L = 1$. Taking spin into account, the ground state has a total angular momentum (again a good quantum number) of either $J = 1/2$ or $J = 3/2$. If, as in our present theory, spin-orbit coupling is neglected, these states are degenerate: the ground state is a sextet with quantum numbers $L^2 = L(L+1) = 2$, $L_z = 0, \pm 1$, $S^2 = 3/4$, $S_z = \pm 1/2$. (In reality with spin-orbit splitting the ground state has $J = 1/2$ and is a duplet.) In the particular case under consideration with only one electron in an unfilled shell, the spin does not influence the symmetry of the orbital part of the wave function, hence we can forget about it and consider the orbitally threefold degenerate state with $L = 1$. Its spatial electron density consists of a spherical part (of the atomic core) and a part of $L = 1$ symmetry, which can for instance be chosen $\sim x^2$ or $\sim y^2$ or $\sim z^2$ comprising three orthogonal to each other ground states. Any linear combination of these three states with real coefficients α, β, γ,

$\alpha^2 + \beta^2 + \gamma^2 = 1$ is again a ground state with the symmetry axis of the aspherical charge density pointing in the direction of the vector (α, β, γ). Likewise, with complex coefficients states can be build which are eigenstates of the angular momentum projection $L_{(\alpha,\beta,\gamma)}$ in any given direction (α, β, γ). However, in real physics, even with spin-orbit coupling neglected, the atom is nearly always in a *paramagnetic* state with $L^2 = 2$ and $\langle L^2_{(\alpha,\beta,\gamma)}\rangle \neq 0$ but $\langle L_{(\alpha,\beta,\gamma)}\rangle = 0$ for all projections onto all directions (α, β, γ). This state is

$$\begin{aligned}\hat{\gamma} &= |x\rangle\frac{1}{3}\langle x| + |y\rangle\frac{1}{3}\langle y| + |z\rangle\frac{1}{3}\langle z| = \\ &= |L_z = 1\rangle\frac{1}{3}\langle L_z = 1| + |L_z = 0\rangle\frac{1}{3}\langle L_z = 0| + |L_z = -1\rangle\frac{1}{3}\langle L_z = -1|\end{aligned} \tag{4.50}$$

where the anisotropic part of the wavefunction $|x\rangle$ behaves like x/r, that of $|L_z = 1\rangle$ like $(x + iy)/r$ and so on. It is a simple exercise to prove that both lines of (4.50) are equal. This state has a spherical charge density. It is the paramagnetic ensemble state and is degenerate with the three polarized pure ground states. Note that it is an important characteristic of the paramagnetic state that it fluctuates isotropically. For instance the pure state $|z\rangle = |L_z = 0\rangle$ yields $\langle L_{(\alpha,\beta,\gamma)}\rangle = 0$ for all projections onto all directions (α, β, γ), but L_z does not fluctuate: $\langle L_z^2\rangle = 0$. Hence it cannot be influenced by a magnetic field in z-direction.

Given a complete orthonormal set of orbitals ψ_i, $\langle\psi_i|\psi_j\rangle = \delta_{ij}$, the determinantal states (1.15) form a complete orthonormal set of N-particle states

$$\Phi_L(x_1 \ldots x_N) = \frac{1}{\sqrt{N!}} \det \|\psi_{l_i}(x_k)\|, \quad \langle\Phi_{L'}|\Phi_L\rangle = \delta_{L'L}, \tag{4.51}$$

being labeled by an N-tupel of integers $L = (l_1 \leq \ldots \leq l_N)$ or likewise by a sequence of orbital occupation numbers n_i^L, $i = 1 \ldots \infty$ with $n_i^L = 1$ if $l_i \in L$ and $n_i^L = 0$ otherwise. Expanding $\Psi_K = \sum_L C_K^L \Phi_L$ transforms (4.46) into

$$\begin{aligned}\gamma_N &= \sum_{LL'} \Phi_L \gamma_{N;LL'} \Phi^*_{L'}, \\ \gamma_{N;LL'} &= \sum_K C_K^L g_{N;K} C_K^{L'*} = \gamma^*_{N;L'L}, \quad \operatorname{tr}\gamma_N = \sum_L \gamma_{N;LL} = 1.\end{aligned} \tag{4.52}$$

From this expression, using (2.1) and (2.3), and recalling that integration over one particle variable x_i of a product of two determinantal states can be

non-zero only, if both determinants contain at least one common orbital, it is not difficult to find

$$n(x) = \sum_{ij} \psi_i(x) \tilde{n}_{N;ij} \psi_j^*(x), \quad \tilde{n}_{N;ij} = \sum_{i(LL')_j} n_i^L \gamma_{N;LL'} n_j^{L'}, \tag{4.53}$$

where the last sum runs over pairs (LL') for which $n_k^L = n_k^{L'}$ except possibly for $k = i, j$. (Recall that the n_k^L take on only two values: 0 or 1.) Clearly $\tilde{n}_{N;ij} = \tilde{n}_{N;ji}^*$. Hence, this matrix may be diagonalized, $\|\tilde{n}_{N;ij}\| \to \mathrm{diag}(n_{N;1}, n_{N;2}, \ldots)$, by a unitary transformation to new orbitals $\phi_i(x)$:

$$n(x) = \sum_i n_{N;i} |\phi_i(x)|^2, \quad 0 \le n_{N;i} \le 1, \quad \sum_i n_{N;i} = N. \tag{4.54}$$

The unitarity of the transformation from the ψ_i to the ϕ_i transfers the orthonormality property from the set of the former to that of the latter. The constraints on the $n_{N;i}$ follow most easily from the constraints on the $g_{N;K}$ and from $\sum_L |C_K^L|^2 = 1$, if one thinks of using the diagonalizing orbitals ϕ_i to build the Φ_L. The same analysis yields the single-particle density matrix

$$\gamma_1(x; x') = \sum_i \phi_i(x) n_{N;i} \phi_i^*(x') \tag{4.55}$$

with the same constraints (4.54) on the $n_{N;i}$. The orbitals ϕ_i of this general single-particle density matrix expression were named natural orbitals by Löwdin. (Note that, contrary to the case (2.8) of a pure determinantal state, here the i-sum may run over an infinite number of non-zero items; this may already happen for a correlated pure state. Moreover, for (4.55) generally $\gamma_1^2 \neq \gamma_1$.)

One now defines a density functional for N-particle density matrix states

$$F_{\mathrm{DM}}[n] \stackrel{\mathrm{def}}{=} \inf\{\ \mathrm{tr}\,(\hat{T} + \hat{W})\gamma_N \,|\, \gamma_N \mapsto n\,\}, \tag{4.56}$$

where the infimum search is over all ensemble states (N-particle density matrices) giving n (again, the infimum is in fact a minimum). The corresponding interaction-free functional is by means of the general single-particle density matrix (4.55) (cf. (2.32))

$$T_{\mathrm{DM}}[n] \stackrel{\mathrm{def}}{=} \inf\{\ \mathrm{tr}\,\hat{T}\gamma_N \,|\, \gamma_N \mapsto n\,\} \ =$$

$$= \min\left\{ \sum_i n_{N;i} \int dx\, \phi_i^* \left(-\frac{\nabla^2}{2}\right) \phi_i \,\middle|\, \sum_i n_{N;i} |\phi_i|^2 = n,\ 0 \le n_{N;i} \le 1 \right\}. \tag{4.57}$$

The minimum search is now over all sets of orthonormal orbitals ϕ_i and over all sets of real numbers $n_{N;i}$, together fulfilling the constraints indicated in the right part between the curled brackets. The same reasoning as previously leads to

$$E[v] = \inf_{n \in \mathcal{J}_N} \left\{ F_{\mathrm{DM}}[n] + \int dx n v \right\}. \tag{4.58}$$

Although obviously by definition $F_{\mathrm{DM}}[n] = F_{\mathrm{LL}}[n]$ *for pure-ground state densities* n, in general only $F_{\mathrm{DM}}[n] \leq F_{\mathrm{LL}}[n]$ holds, which relation is equally obvious by definition. Since the relation between γ_N and n is linear, $F_{\mathrm{DM}}[n]$ is convex, hence $F_{\mathrm{DM}}[n] \leq$ convex hull of $F_{\mathrm{LL}}[n]$ (cf. Section 6.3 below). The two continuations F_{LL} and F_{DM} of F_{HK} from $\mathcal{A}_N$ onto $\mathcal{J}_N$ are different, nevertheless for both the Hohenberg-Kohn variational principle (the stationary solutions n of which may be chosen $\in \mathcal{A}_N$) holds. (As distinct from the situation with pure states, an affine-linear combination of two densities n_1 and n_2 coming from two degenerate ground states is always again a density coming from an ensemble ground state, obtained by mixing of the two original states. Cf. the discussion after (4.49).)

Of course, now likewise $T_{\mathrm{DM}}[n] \leq T_{\mathrm{LL}}[n] \leq T[n]$ holds. Using a decomposition of F_{DM} analogous to (4.27), this time, however, with T_{DM} instead of T, leads to

$$E[v] = \inf \left\{ \sum_i n_{N,i} \int dx\, \phi_i^* \left(-\frac{\nabla^2}{2} \right) \phi_i + E_{\mathrm{H}}[n] + E_{\mathrm{XC}}[n] + \int dx\, n v \,\middle|\, \sum_i n_{N,i} |\phi_i|^2 = n,\ 0 \leq n_{N,i} \leq 1,\ \sum_i n_{N,i} = N,\ (\phi_i|\phi_j) = \delta_{ij} \right\}. \tag{4.59}$$

Varying n through ϕ_i^*, the usual Kohn-Sham equation (4.32) is obtained from (4.59). Now putting the resulting Kohn-Sham orbitals into the variational expression $\{\cdots\}$ of (4.59) for $E[v]$ and varying the orbital occupation numbers $n_{N;i}$ one has

$$\begin{aligned} \frac{\partial\{\cdots\}}{\partial n_{N;i}} &= \int dx\, \phi_i^* \left(-\frac{\nabla^2}{2} \right) \phi_i + \\ &\quad + \int dx' \frac{\delta}{\delta n(x')} \left\{ E_{\mathrm{H}}[n] + E_{\mathrm{XC}}[n] + \int dx\, n v \right\} \frac{\partial n(x')}{\partial n_{N;i}} = \\ &= \int dx\, \phi_i^* \left(-\frac{\nabla^2}{2} \right) \phi_i + \int dx\, \phi_i^* v_{\mathrm{eff}} \phi_i = \varepsilon_i. \end{aligned} \tag{4.60}$$

This result is called Janak's theorem. Of course, only occupation number variations $\sum_i \delta n_{N;i} \partial\{\cdots\}/\partial n_{N;i}$ with $\sum_i \delta n_{N;i} = 0$ are allowed due to the constraint $\sum_i n_{N;i} = N$ in (4.59). Pick a pair ij and pick $\delta n_{N;i} = -\delta n_{N;j} = \delta n$. Then from (4.60), if $n_{N;i} = 0$ and $n_{N;j} > 0$, then $\varepsilon_i \geq \varepsilon_j$ ($\partial\{\cdots\}/\partial\delta n \geq 0$, because $\delta n < 0$ is not allowed in that case (cf. Fig.3 on page 66). If $n_{N;i} < 1$ and $n_{N;j} = 1$, then the minimum condition demands again $\varepsilon_i \geq \varepsilon_j$, because $\delta n < 0$ (i.e. $\delta n_{N;j} > 0$) is again not allowed. If $0 < n_{N;i}$, $n_{N;j} < 1$, then $\delta n = 0$ is an inner point of the domain of variation for δn, and $\varepsilon_i = \varepsilon_j$ must hold true.

This result [Englisch and Englisch, 1984a,b] generalizes the occupation rule of Section 4.2: If the Kohn-Sham energies are again ordered according to

$$\varepsilon_1 \leq \varepsilon_2 \leq \ldots, \tag{4.61}$$

then

$$n_{N;i} = 1 \text{ for } \varepsilon_i < \varepsilon_N,\ 0 \leq n_{N;i} \leq 1 \text{ for } \varepsilon_i = \varepsilon_N,\ n_{N;i} = 0 \text{ for } \varepsilon_i > \varepsilon_N. \tag{4.62}$$

This aufbau principle releases problems with level crossing if self-consistency is balanced with partial occupation of crossing levels keeping them degenerate. If in this way a self-consistent v_{eff} can be found, then there is again a $v^0[n]$ equal to that v_{eff} for which the obtained density is a non-interacting ground state density (but no longer necessarily a determinantal one), and hence the ϕ_i, $n_{N;i}$ are minimizing (4.57) and (4.58).

There are more attempts to build up a density-matrix functional theory based on γ_1 [Zumbach and Maschke, 1985].

4.6 Dependence on Particle Number N

In a mixed state, the particle number N need not be fixed. It need even not be fixed in a pure state if consideration is based on the Fock space of Section 1.5. All determinants (4.51) for all integer numbers $N = 1, 2, \ldots$ together with the vacuum state form a complete orthonormal basis of the Fock space $\mathcal{F}$,

$$|\Phi_L\rangle = |n_1^L \ldots n_i^L \ldots\rangle \tag{4.63}$$

in notation of (1.63). The density is now given by (2.23, 2.29) for a pure state. For a general mixed state

$$\hat{\gamma} = \sum_K |\Psi_K\rangle g_K \langle\Psi_K|, \quad 0 \leq g_K, \quad \sum_K g_K = 1, \tag{4.64}$$

where the $|\Psi_K\rangle$ are general Fock space vectors, the single-particle density matrix, for instance, is defined as

$$\gamma_1(x;x') = \operatorname{tr}\hat{\psi}(x)\hat{\gamma}\hat{\psi}^\dagger(x') = \sum_K g_K \langle\Psi_K|\hat{\psi}^\dagger(x')\hat{\psi}(x)|\Psi_K\rangle. \tag{4.65}$$

All relations (4.51) through (4.55), with the subscript N omitted and the L-sums running over all basis states (4.63) of the Fock space, remain valid in the present more general case on the basis of the same reasoning as there. The only change is that $\sum_i n_i = N$ may now be any non-negative *real* number, the expectation value of the total particle number operator ($N = \langle\hat{N}\rangle = \operatorname{tr}\hat{N}\hat{\gamma}$).

The ground state energy (cf. (4.4)) is now defined as a functional of v *and* a function of real non-negative N:

$$E[v,N] \stackrel{\text{def}}{=} \inf\{\ \operatorname{tr}\hat{H}[v]\hat{\gamma}\ \mid\ \operatorname{tr}\hat{N}\hat{\gamma} = N\ \}. \tag{4.66}$$

This definition modifies the definition of (4.4) in two steps. First, in (4.4) the search was over pure particle number eigenstates (eigenstates of $\hat{N}$), whereas the pure states $|\Psi_K\rangle$ of (4.64) are general Fock space states, which may be linear combinations $|\Psi_K\rangle = \sum_{N'}|\Psi_{N'}\rangle c_{N'}$, $\sum_{N'}|c_{N'}|^2 = 1$ of particle number eigenstates $\hat{N}|\Psi_{N'}\rangle = |\Psi_{N'}\rangle N'$ (cf. Section 1.5). *Since* $\hat{H}[v]$ *and* $\hat{N}$ *commute*, one has $\langle\Psi_K|\hat{H}[v]|\Psi_K\rangle = \sum_{N'}|c_{N'}|^2\langle\Psi_{N'}|\hat{H}[v]|\Psi_{N'}\rangle$. This step already finds in (4.66) the minimum over affine-linear combinations of pure-state minima of (4.4) for various N': If we explicitly denote the N-dependence of (4.4) by $E_N[v]$, then we have in (4.66) a search over all $\sum_{N'}|c_{N'}|^2 E_{N'}[v]$, $\sum_{N'}|c_{N'}|^2 N' = N$, $\sum_{N'}|c_{N'}|^2 = 1$. The second step consists in considering mixtures (4.64) of pure states $|\Psi_K\rangle$, which, however, does not further lower the infimum of (4.66) because we know already from the last section that this infimum over mixed states is the same as over pure states. Hence,

$$E[v,N] = \min_{\{c_{N'}\}}\left\{\sum_{N'}|c_{N'}|^2 E_{N'}[v]\ \middle|\ \sum_{N'}|c_{N'}|^2 N' = N,\ \sum_{N'}|c_{N'}|^2 = 1\right\}, \tag{4.67}$$

and this means $E[v,N]$ is the convex hull over all $E_{N'}[v]$ with integer N'; in particular $E[v,N] \le E_N[v]$ for integer N, and equality holds, if $E_N[v]$ (with integer N and fixed v) is convex (cf. Fig.5).

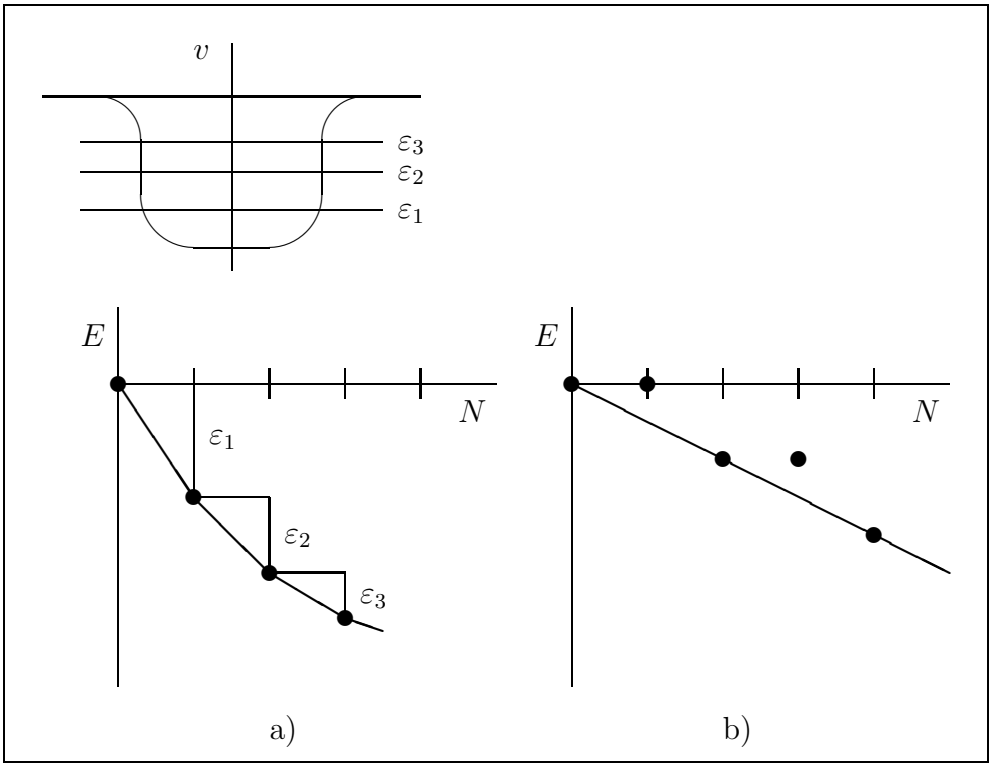

Figure 5: Ground state energies $E_N[v]$ ($\bullet$ for integer N) and $E[v, N]$ (thick line for real N) as a function of N at a fixed v: a) for non-interacting particles in the external potential v, b) for particles moving in constant zero external potential and binding in pairs: the energy lowers every other step by the negative binding energy of the next formed pair.

Given $v(x)$ for an interaction-free system with single-particle levels $\varepsilon_1 \leq \varepsilon_2 \leq \ldots$, the ground state energy for *integer* N is apparently

$$E_N^0[v] = \sum_{i=1}^{N} \varepsilon_i, \tag{4.68}$$

which can directly be deduced by using (4.55) in (4.66). Varying ϕ_i^* yields the single-particle Schrödinger equation, and varying n_i yields (4.62) and hence (4.68). Due to the monotone order of the ε_i, which has to be used in (4.68), trivially

$$2E_N^0[v] \leq E_{N-1}^0[v] + E_{N+1}^0[v]. \tag{4.69}$$

Hence, the variation of the n_i for *non-integer* N leads, according to (4.62), just to linear interpolation between the values (4.68) for *neighboring* integer N (Fig.5a).

Unfortunately, nothing is known up to now about conditions under which $E_N[v]$ is convex in N for an *interacting* system, although empirically it is convex for electrons in atoms and molecules (which is also confirmed numerically). If $E_N[v]$ is convex in N, then the theory of last section is simply extended by considering in (4.54–4.62) occupation numbers n_i adding up to non-integer N, and densities $n(x)$ integrating to non-integer N.

Nevertheless, as seen from (4.67), also in the general case of interacting particles, $E[v, N]$ for fixed v may be a (piecewise linear) convex function of N even if $E_N[v]$ is not convex. (Cf. Fig.5b. In Chapter 6 the fact will be exploited that $E[v]$ is a *concave* function of v under completely general assumptions.) If $E[v, N]$ were a *strictly* convex function of N for fixed v, then a chemical potential $\mu(N)$ would exist as a strictly monotone function of N, and

$$E[v, N] = \inf_{\hat{\gamma}} \{ \operatorname{tr} (\hat{H}[v] - \hat{N}\mu(N))\hat{\gamma} \}. \tag{4.70}$$

If $E[v, N]$ exists and is convex in N (cf. Section 6.1) then it is piecewise linear between integer N, $\mu(N) = \partial E[v, N]/\partial N$ is monotone and piecewise constant (and so is $N(\mu)$, the map inverse to $\mu(N)$), and it jumps at integer N. In the interaction-free case those jumps are from values ε_{N-1} to values ε_N. ($N(\mu)$ jumps by one at $\mu = \varepsilon_i$ the jumps stacking on each other in cases of orbital degeneracy.) Note that the thermodynamic chemical potential is defined for the dependence of E/V on N/V in the thermodynamic limit $V \to \infty$ for the volume V (where some appropriate assumption on v is to be made, e.g. being assumed periodic), and most of the jumps dissolve in that limit. Also, in the thermodynamic limit, if $E_N[v]$ were not convex, an instability towards a phase-separated state with two different particle densities would appear, which in case of a short-range interaction w would again have a thermodynamic energy corresponding to $E[v, N]$. For Coulomb interaction, the situation is more involved due to its long-range character, and the notion of a chemical potential applies unmodified to situations with local charge neutrality only. Phase separation in this context appears due to restrictions of quantum (coherent) mixtures of macroscopic states. To incorporate this case, $\hat{\gamma}$ in (4.70) comprises phase-separated thermodynamic mixtures.

Now, fix v and consider $E(N)$ for an atom or an ion or a molecule or radical. Of course, in the interacting case, the Kohn-Sham orbital energies

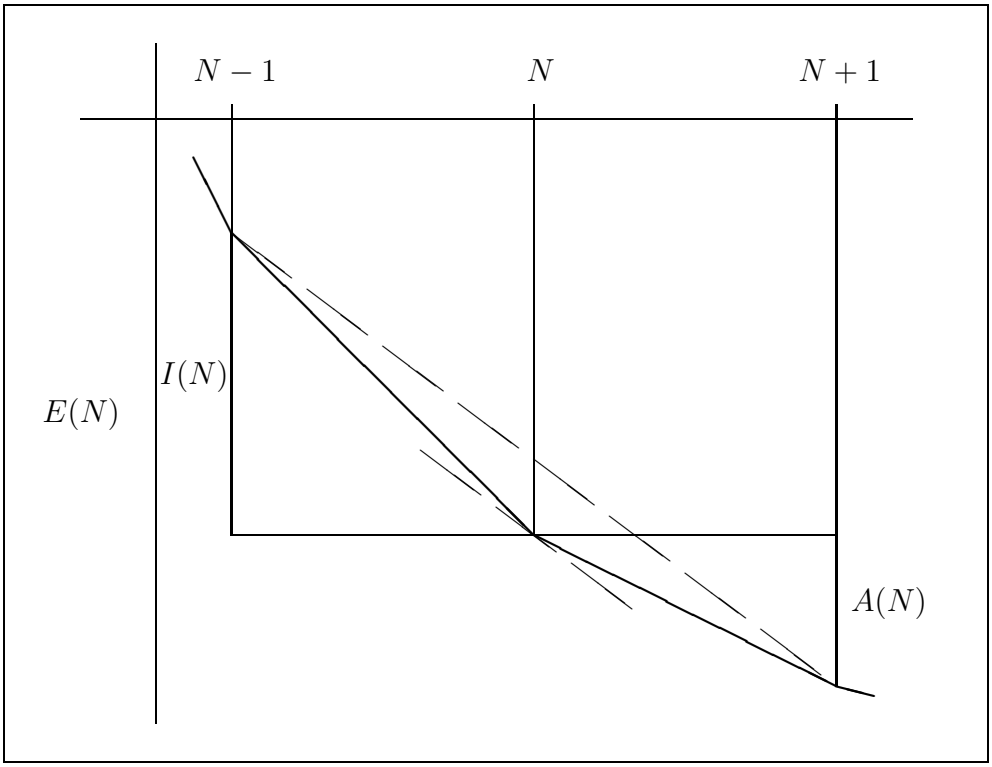

Figure 6: Ionization potential I and electron affinity A for an N-electron species. The slope of the $E(N)$ curve is $\mu(N)$. The negative of the slope of the dashed line is Mulliken's electronegativity, apparently yielding a good value for $-\mu(N)$.

ε_i are now functions of N or of the occupation numbers n_i of (4.62), respectively. Then $A = E(N) - E(N+1)$ is the electron affinity, and from Janak's theorem (4.60) the in principle rigorous relation

$$-A = \int_0^1 dn_{\mathrm{LUMO}}\, \varepsilon_{\mathrm{LUMO}}(n_{\mathrm{LUMO}}) \tag{4.71}$$

follows, where LUMO is the lowest unoccupied Kohn-Sham level (Lowest Unoccupied Molecular Orbital) of the N-particle state, computed self-consistently for varying occupation number of that level with all the remaining occupation numbers fixed (level crossing assumed to be absent). Analogously one finds

$$-I = \int_0^1 dn_{\mathrm{HOMO}}\, \varepsilon_{\mathrm{HOMO}}(n_{\mathrm{HOMO}}) \tag{4.72}$$

for the ionization potential $I = E(N-1) - E(N)$, where HOMO is now the highest occupied Kohn-Sham level of the N-particle state. Mulliken's electronegativity is

$$\chi_M = \frac{1}{2}(I + A), \tag{4.73}$$

since for two species S and T the energy for an electron transition from S to T is $\Delta E = I_S - A_T$ whereas for an electron transition from T to S it is $\Delta E = I_T - A_S$. Both are equal for $I_S + A_S = I_T + A_T$. The whole situation for I, A and χ_M is sketched on Fig.6.

As a reasonable approximation to (4.71, 4.72), Slater's concept of the transition state can be used:

$$-A \approx \varepsilon_{\mathrm{LUMO}}(n_{\mathrm{LUMO}} = 1/2), \quad -I \approx \varepsilon_{\mathrm{HOMO}}(n_{\mathrm{HOMO}} = 1/2). \tag{4.74}$$

A detailed consideration of possible physical meanings of the Kohn-Sham orbital energies ε_i may be found in [Perdew, 1985].

Strictly speaking, (4.74) should even not only be an approximation, because $\partial E[v, N]/\partial N$ should quite generally be constant between two neighboring integer N, and by Janak's theorem it should be given by the Kohn-Sham orbital energy ε_i of the partially occupied orbital, provided $E_N[v]$ is convex in the integer variable N. In approximate versions of the theory in use this piecewise linearity of $E[v, N]$ is, however, never provided.

4.7 Spin Polarization

With a few modifications, this whole text allows for a spin-dependence of the external potential $v(x) = v(\boldsymbol{r}, s) = (v(\boldsymbol{r}, +), v(\boldsymbol{r}, -))$ (the latter notation for the most important spin-half case) with an interaction term

$$\int dx\, nv = \int d^3r\, (n(\boldsymbol{r}, +)v(\boldsymbol{r}, +) + n(\boldsymbol{r}, -)v(\boldsymbol{r}, -)). \tag{4.75}$$

If v is spin-independent, $v(\boldsymbol{r}, +) = v(\boldsymbol{r}, -) = v(\boldsymbol{r})$, then this interaction reduces to

$$\int dx\, nv = \int d^3r\, n(\boldsymbol{r})v(\boldsymbol{r}). \tag{4.76}$$

As declared at the very beginning after (1.7), our notation is always meant to comprise both cases. (Diamagnetic interaction, however, i.e. orbital motion

coupling to an external magnetic field is not covered by this frame.) We used this notation to avoid parallelism in the presentation of the material, because the theories of both cases go formally largely in parallel.

One main difference is that there is no spin dependent analogue to the basic theorem by Hohenberg and Kohn. The spin dependence of an external potential, the external magnetic field, is not a unique function of the ground state spin density any more: for instance for a finite system with discrete levels of total energy, Zeeman energies in a homogeneous external magnetic field shift the levels without (in our non-relativistic approximation (4.1) with spin-orbit interaction and orbital magnetism neglected) changing wave functions *until levels with different total spin in field direction cross.* They must cross because they have different slopes of the Zeeman contributions. At that field the ground state reconstructs. Hence, $n(x)$ is constant in certain intervals of homogeneous magnetic field. Fig.7 shows the case of beryllium atom. The discrete stationary states are eigenstates of the z-component of the total spin, Σ_z, and the interaction energy with a homogeneous $\boldsymbol{B}$-field in z-direction is $-2\mu_{\mathrm{Bohr}}\Sigma_z B$. On the left of Fig.7 the signature of the four-electron state is indicated (main orbital character of the correlated state). For small B, the ground state is $1s^2 2s^2$ with $\Sigma_z = 0$ and the first excited state is $1s^2 2s2p$ with $\Sigma_z = 1$. The next $\Sigma_z = 1$ state $1s2s^2 2p$ is much higher in energy and hence cannot become the ground state even in large B-fields. The lowest $\Sigma_z = 2$ state is $1s2s2p^2$, and it becomes the ground state for larger B. The whole picture is symmetric with respect to $B = 0$, only the right half is sketched in Fig.7. On B-intervals of nonzero lengths the ground state and hence also its spin density does not change with changing field strength.

With inhomogeneous fields (next section) the situation is more involved and not fully explored up to now. In view of this difficulty, von Barth and Hedin [von Barth and Hedin, 1972] modified the proof of the basic theorem by Hohenberg and Kohn and proved directly a one-to-one map of ground state spin densities on ground state wave functions *for non-degenerate ground states* in order to construct the analogue of (4.16). However, now we have at hand the functionals F_{LL} and F_{DM} (and Lieb's F introduced in Section 6.2) which all are uniquely defined without recourse to the basic theorem by Hohenberg and Kohn. Hence, everything of this chapter except the passage from (4.10) to (4.14) of Section 4.1 holds true for the spin dependent case. Implications of the basic theorem by Hohenberg and Kohn for the problem of existence of functional derivatives for $F[n]$ will be discussed in Chapter 6.

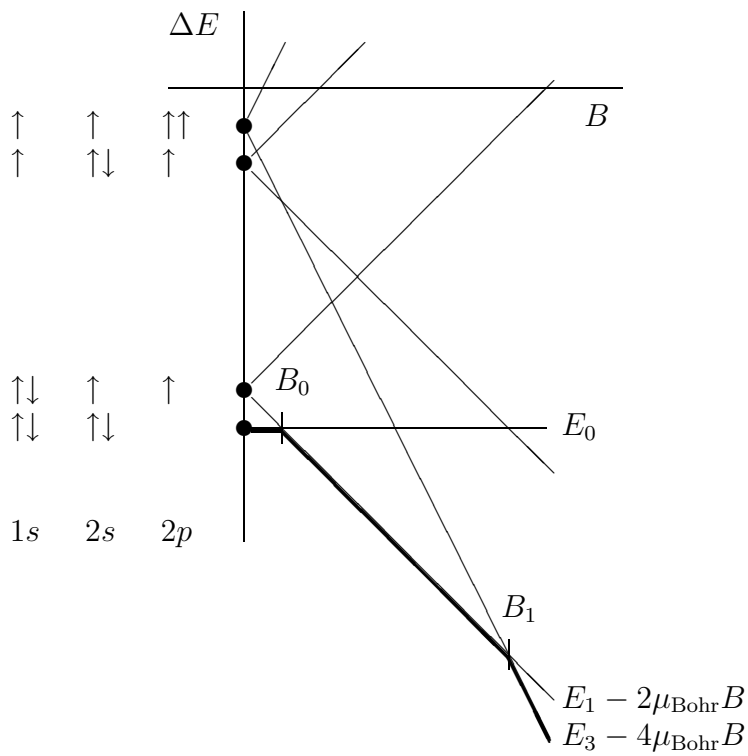

Figure 7: Energy change with field B for the discrete levels of Be atom as explained in the text. The lines with positive slopes correspond to states with all spins reversed. The thick line marks the ground state energy.

Even if the external potential is not spin-dependent, which is always the case if no magnetic field is applied and the hyperfine interaction effects of nuclear spins can be neglected, yet it is rather the rule than the exception that the ground state is spin-polarized. This comes about since an odd number of spin half particles unavoidably has a half-integer total spin, and since the role of spin in the Pauli principle allows for a spin-dependent exchange interaction even when the Hamiltonian does not contain the spin variable. The Hohenberg-Kohn theory, on the other hand, has both variants, one for $n(\boldsymbol{r}, s)$ and $v(\boldsymbol{r}, s)$, and one for $n(\boldsymbol{r}) = n(\boldsymbol{r}, +) + n(\boldsymbol{r}, -)$ and $v(\boldsymbol{r})$. In the latter case, even if the ground state appears to be spin-polarized, v is a unique function of the spatial density $n(\boldsymbol{r})$ alone. (Spontaneous spin polar-

ization always implies degeneracy of the ground state, since the direction of the total momentum in space is arbitrary, it is the classical example of spontaneous symmetry breaking. As we have seen, however, this does not spoil the basic theorem.)

These considerations apply even to a non-interacting system. Let $w \equiv 0$ and $v = -7/r$ for seven electrons. Then the ground state is degenerate; it may have total spin 1/2 or 3/2, and various radial density distributions, because of the 8 degenerate $2s$ and $2p$ spin-orbitals only 5 are occupied. The total spin of the real nitrogen atom is according to Hund's rules 3/2 and is only orientational degenerate. The $O(4)$ symmetry of the interaction-free problem is reduced to $SO(3)$ in the Hamiltonian when Coulomb repulsion of the electrons is introduced. This lifts the degeneracy of $2s$ with $2p$ states (releases conservation of the Pauli-Runge-Lenz vector) and thus fixes the radial density distribution. As we have three electrons in $2p$ states, spin-independence would still suggest a degeneracy of ground states with total spin 1/2 and 3/2. In reality, however, the latter state has a lower energy. This time, the symmetry is broken kinematically: by requirement of the Pauli principle both wavefunctions differ in the number of orbital node surfaces in the nine-dimensional orbital configuration space, although the orbital density $n(\boldsymbol{r})$ may be spherical in both cases, if each of the three spatial $2p$ orbitals is occupied with one electron in an $L = 0$ state (L being the total orbital momentum). In the spin-3/2 state, however, the wavefunction for the relevant spinor component must be zero for $\boldsymbol{r}_1 = \boldsymbol{r}_2$, $\boldsymbol{r}_1 = \boldsymbol{r}_3$, and $\boldsymbol{r}_2 = \boldsymbol{r}_3$ due to antisymmetry, whereas in the spin-1/2 state it has less nodes. These nodes of the *many-particle* wavefunction, which do not cause additional nodes in its single-particle projections, do not increase the kinetic energy; they decrease, however, the Coulomb repulsion energy. Finally, a particularly chosen pure ground state will have a particular spin direction: the symmetry is further broken, this time spontaneously (without a symmetry breaking term in the Hamiltonian or in the constraints on wavefunctions).

At a first glance, one could think that both the spin-3/2 state and the spin-1/2 state of the nitrogen atom considered above will have the same density $n(\boldsymbol{r})$ (spherical symmetric). But of course these densities differ in their radial dependence, although in a tiny difference (cf. Fig.8). They have different total energies and hence different exponential fall-offs. Hartree-Fock theory yields all these details, though approximately, via the spin-dependence of the exchange potential operator. *Spin-independent* Hohenberg-Kohn-Sham theory provides only a local spin-independent po-

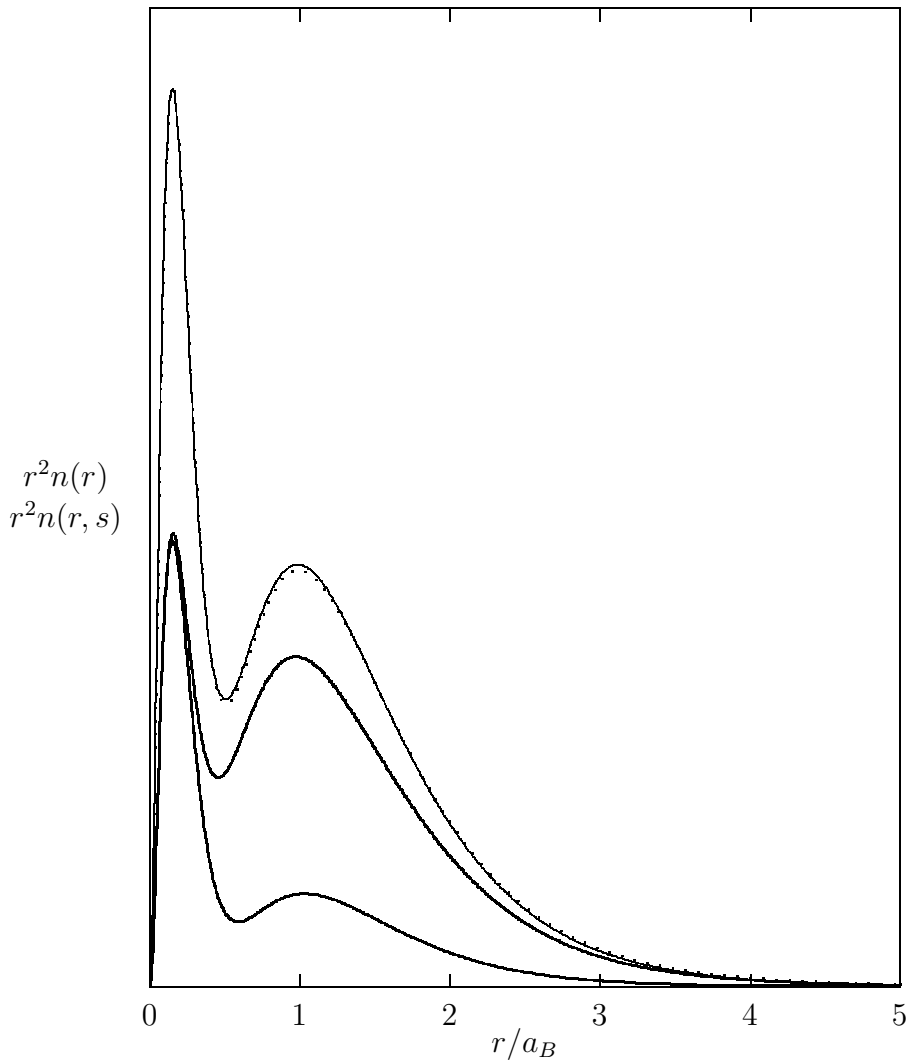

Figure 8: Radial electron densities and spin densities for the nitrogen atom. The upper full line gives the total electron density for the spin 3/2 ground state, and the lower two lines give the spin up and down densities. For comparison, the dotted line very close to the upper full line gives the total density of the lowest spin 1/2 state. (Calculated by L. Steinbeck, unpublished.)

tential $v_{\mathrm{XC}}(\boldsymbol{r}) = \delta E_{\mathrm{XC}}[n]/\delta n(\boldsymbol{r})$. The question arises, how does this theory know about the total spin of the ground state of the nitrogen atom being 1/2 or 3/2? The simple answer is, it does not. Being a theory for $n(\boldsymbol{r})$ and the ground state energy, if $E_{\mathrm{XC}}[n]$ would be exactly known, it would supply

the correct ground state energy and the correct ground state density (with the correct exponential falloff) as the sum over *spin-independent* Kohn-Sham orbital densities, *but it would not know anything about the spin.* It should have become clear from this little excursion that in this variant $E_{\mathrm{XC}}[n(\boldsymbol{r})]$ must be an extremely tricky functional.

It should immediately be clear that the spin dependent variant of the theory provides a much more promising situation, *because now* $E_{XC}[n(x)]$ *is told about the spin of the state through* $n(x)$ comprising the spatial spin density $m(\boldsymbol{r}) = n(\boldsymbol{r},+) - n(\boldsymbol{r},-)$, and it need no further trace tiny differences of spatial density fall-offs. Recall that, since for a (possibly spin-polarized) ground state $n(\boldsymbol{r})$ coming from a spin-independent $v(\boldsymbol{r})$ the latter is uniquely determined by the former, and $m(\boldsymbol{r})$ is determined by that v, e.g. via the N-particle Schrödinger equation, this $m(\boldsymbol{r})$ is also determined by $n(\boldsymbol{r})$ (although in cases of degeneracy not uniquely). These considerations aim at making clear what has not yet been duly stressed in the literature: $E_{\mathrm{XC}}[n(\boldsymbol{r})]$ and $E_{\mathrm{XC}}[n(x)]$ play completely different roles in their respective versions of theory. Both cope with all ground states, spin-polarized or not (since degeneracy of the ground state need not be excluded), of all spin-independent potentials, one knowing only the spatial particle density $n(\boldsymbol{r})$ (and hence it must be by far the more clever one) and one knowing additionally the spatial spin density $m(\boldsymbol{r})$. The latter, less demanded by ground states to potentials $v(\boldsymbol{r})$, copes additionally with ground states to potentials $v(x)$. The exact functionals would of course do equally well for potentials $v(\boldsymbol{r})$, but clearly a guess for the less demanding case is more promising.

We mention in passing that even in the exact versions of both theories the Kohn-Sham orbital energies would come out spin-independent in the one case and spin-dependent in the other, which proves again that the immediate physical meaning of those entities is rather limited.

4.8 Non-Collinear Spin Configurations

We now consider the more general case of an arbitrary possibly not unidirectional and $\boldsymbol{r}$-dependent in magnitude magnetic field $\boldsymbol{B}(\boldsymbol{r})$ coupled to the electron spin only via the Zeeman term. The relativistic coupling to the orbital motion as well as dipole-dipole interaction between the spins are still neglected. This case is covered by a 2×2 spin matrix

$$v_{ss'}(\boldsymbol{r}) = v(\boldsymbol{r})\delta_{ss'} - 2\mu_{\mathrm{Bohr}}\boldsymbol{B}(\boldsymbol{r}) \cdot \boldsymbol{\sigma}_{ss'} \tag{4.77}$$

for the external potential, where $\boldsymbol{\sigma}_{ss'}$ is the vector of the three components (1.10) of the spin operator (1/2 times the Pauli matrices). The potential energy due to this external potential is

$$\sum_{ss'} \int d^3r\, n_{ss'}(\boldsymbol{r}) v_{s's}(\boldsymbol{r}) = \int d^3r \Big(n(\boldsymbol{r}) v(\boldsymbol{r}) - \boldsymbol{m}(\boldsymbol{r}) \cdot \boldsymbol{B}(\boldsymbol{r}) \Big) \tag{4.78}$$

with the spin magnetization density $\boldsymbol{m}$ from (2.34). This case was first considered by [von Barth and Hedin, 1972]. Here we follow the recent analysis by [Eschrig and Pickett, 2001].

Following the route of Section 4.1, let again $v_1 \neq v_2$ be two different potentials having the same ground state Ψ_0. Instead of (4.9), subtraction of the two Schrödinger equations from each other now yields

$$\sum_i \sum_{s'_i} \Delta v_{s_i s'_i}(\boldsymbol{r}_i) \Psi_0(\boldsymbol{r}_1 s_1, \ldots, \boldsymbol{r}_i s'_i, \ldots, \boldsymbol{r}_N s_N) = \\ = \Psi_0(\boldsymbol{r}_1 s_1, \ldots, \boldsymbol{r}_N s_N) \Delta E \tag{4.79}$$

with $\Delta v = v_1 - v_2$, $\Delta E = E[v_1] - E[v_2]$. In this notation, Ψ_0 may be understood as a set of 2^N functions of $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N$ for all combinations of spin indices $s_i = +, -$. (Due to the symmetry of Ψ_0, not all of those functions are independent and not all of them need be non-zero; however, at least one of them must be non-zero.) Contrary to (4.9) where the equations for those functions were decoupled, they are now coupled by the l.h.s. of (4.79), and the reasoning after (4.9) does not apply any more.

In order to decouple the equations, perform an $\boldsymbol{r}$-dependent unitary spin rotation $Q_{ss'}(\boldsymbol{r})$ which diagonalizes Δv:

$$\big[Q(\boldsymbol{r}) \Delta v(\boldsymbol{r}) Q^\dagger(\boldsymbol{r}) \big]_{ss'} = \Delta \tilde{v}_s(\boldsymbol{r}) \delta_{ss'}. \tag{4.80}$$

The ground state Ψ_0 is transformed into the new local spin variables according to

$$\Big[\prod_i Q_{s_i s'_i}(\boldsymbol{r}_i) \Big] \Psi_0(\boldsymbol{r}_1 s'_1, \ldots, \boldsymbol{r}_N s'_N) = \tilde{\Psi}_0(\boldsymbol{r}_1 s_1, \ldots, \boldsymbol{r}_N s_N), \tag{4.81}$$

and (4.79) now reads

$$\sum_i \Delta \tilde{v}_{s_i}(\boldsymbol{r}_i) \tilde{\Psi}_0(\boldsymbol{r}_1 s_1, \ldots, \boldsymbol{r}_N s_N) = \tilde{\Psi}_0(\boldsymbol{r}_1 s_1, \ldots, \boldsymbol{r}_N s_N) \Delta E. \tag{4.82}$$

For one component of $\tilde{\Psi}_0$ with $s_1 = s_2 = \cdots = s_{N_+} = +$, $s_{N_+ +1} = s_{N_+ +2}, \cdots = s_N = -$ (this order of the s_i is without loss of generality due to the antisymmetry of Ψ_0) (4.82) reads

$$\left\{ \sum_{i=1}^{N_+} \Delta\tilde{v}_+(\boldsymbol{r}_i) + \sum_{i=N_+ +1}^{N} \Delta\tilde{v}_-(\boldsymbol{r}_i) \right\} \tilde{\Psi}_0(\boldsymbol{r}_1 +, \ldots, \boldsymbol{r}_N -) \; = \\ = \tilde{\Psi}_0(\boldsymbol{r}_1 +, \ldots, \boldsymbol{r}_N -) \Delta E. \quad (4.83)$$

Now, the reasoning which followed (4.9) applies. The dependence of both sides of (4.83) on $\boldsymbol{r}_1$ requires that $\Delta\tilde{v}_+(\boldsymbol{r})$ is a constant a.e., and likewise by virtue of the dependence on $\boldsymbol{r}_N$, $\Delta\tilde{v}_-(\boldsymbol{r})$ must be (possibly another) constant a.e. The special cases $N_+ = 0$ or $N_+ = N$ need no separate treatment since then one of the functions $\Delta\tilde{v}_\pm(\boldsymbol{r})$ is irrelevant. Hence,

$$\Delta\tilde{v}_+ = C_+, \qquad \Delta\tilde{v}_- = C_-. \quad (4.84)$$

Two separate cases must be considered.

Case A: impure spin states. Suppose that there are at least two components of $\tilde{\Psi}_0$ nonzero with *different* numbers N_+ and $N_- = N - N_+$. Then, $N_+ C_+ + (N - N_+) C_- = \Delta E$ must hold for two different numbers N_+, hence $C_+ = C_- = C$ which implies $\Delta\tilde{v}_+ = \Delta\tilde{v}_-$ and from inverting (4.80), $\Delta v_{ss'} = C\delta_{ss'}$. This means $\Delta\boldsymbol{B} = 0$. Repeating the steps from (4.10) to (4.12) recovers the Hohenberg-Kohn result in the more general case:

$$n_{1ss'} = n_{2ss'} \;\longrightarrow\; \begin{pmatrix} v_1(\boldsymbol{r}) - v_2(\boldsymbol{r}) \equiv C \\ \boldsymbol{B}_1(\boldsymbol{r}) - \boldsymbol{B}_2(\boldsymbol{r}) \equiv 0 \end{pmatrix}. \quad (4.85)$$

The ground state spin density determines the scalar potential up to a constant and the applied magnetic field uniquely. This also implies a non-zero spin susceptibility.

Case B: pure spin states. This case is more involved. Suppose that all non-zero components of $\tilde{\Psi}_0$ belong to the same numbers N_+ and N_-. Such a $\tilde{\Psi}_0$ may be considered as a pure spin state, an eigenstate of the operator $\hat{\Sigma}_z = \sum_i \sigma_{z\, s_i s_i'}$ with the eigenvalue $\Sigma_z = N_+ - N/2 = (N_+ - N_-)/2$. Then C_+ and C_- need not be equal:

$$\Delta\tilde{v} = \begin{pmatrix} C_+ & 0 \\ 0 & C_- \end{pmatrix} = \bar{C}\mathbf{1} - 2\mu_{\text{Bohr}}\bar{B}\sigma_z, \quad (4.86)$$

where $\bar{C} = (C_+ + C_-)/2$ and $-2\mu_{\text{Bohr}}\bar{B} = (C_+ - C_-)/2$. Back transforming according to the inverse of (4.80) gives

$$\Delta v_{ss'}(\boldsymbol{r}) = \bar{C}\delta_{ss'} - 2\mu_{\text{Bohr}}\bar{B}\big[Q^\dagger(\boldsymbol{r})\sigma_z Q(\boldsymbol{r})\big]_{ss'}. \quad (4.87)$$

Now the implication is

$$\Psi_{01} = \Psi_{02} \longrightarrow \begin{pmatrix} v_1(\boldsymbol{r}) - v_2(\boldsymbol{r}) = \bar{C} \\ \boldsymbol{B}_1(\boldsymbol{r}) - \boldsymbol{B}_2(\boldsymbol{r}) = \bar{B}\boldsymbol{e}(\boldsymbol{r}) \end{pmatrix},$$
$$\boldsymbol{e}(\boldsymbol{r}) = 2\,\mathrm{tr}\,[\boldsymbol{\sigma} Q^\dagger(\boldsymbol{r})\sigma_z Q(\boldsymbol{r})], \quad (4.88)$$

where the property $2\,\mathrm{tr}\,[\sigma_\alpha\sigma_\beta] = \delta_{\alpha\beta}$ of the σ-matrices (1.10) was used. *A ground state Ψ_0 which may be transformed into an eigenstate of the z-component of total spin by an $\boldsymbol{r}$-dependent spin rotation determines the external scalar potential up to a constant and the applied magnetic field up to a possibly non-unidirectional but constant in magnitude field contribution.*

A simple example of this highly non-trivial generalization of the basic Hohenberg-Kohn theorem is the Be atom as discussed in the last section. In this case $Q(\boldsymbol{r}) \equiv \mathbf{1}$, and the applied magnetic field is determined up to a unidirectional constant field.

The question arises what are the general conditions for such a ground state which may be transformed into a pure spin state $\tilde{\Psi}_0$. A sufficient condition is the existence of an operator

$$\hat{\mathcal{S}} = \sum_{i=1}^{N} \sum_{t_i t_i'} Q^*_{t_i s_i}(\boldsymbol{r}_i)\sigma_{z,t_i t_i'} Q_{t_i' s_i'}(\boldsymbol{r}_i) \quad (4.89)$$

that commutes with the Hamiltonian $\hat{H} = \hat{T} + \hat{W} + \hat{U}$. Consider the particularly important case with $\boldsymbol{B}_1 \equiv 0$. This yields the behavior of a spontaneously (without applied $\boldsymbol{B}$-field) spin-polarized system when a magnetic field $\boldsymbol{B}_2(\boldsymbol{r})$ is switched on. Since the interaction $\hat{W}$ and in the considered case also $\hat{U}_1$ are spin independent, $\hat{\mathcal{S}}$ will commute with $\hat{H}_1$ if and only if it commutes with $\hat{T}$ which means that the ith item of $\hat{\mathcal{S}}$ must commute with the ith item of $\hat{T}$. This necessitates that Q be $\boldsymbol{r}$-independent, so that Ψ_0 itself is an eigenstate of $\hat{\Sigma}_z$ and hence is a collinear spin state. The second condition in (4.88) reduces to $\boldsymbol{B}_2 = \bar{B}\boldsymbol{e}$ where $\boldsymbol{e}$ points in the spin direction; a turning on of a uniform magnetic field leaves the ground state invariant. Restated: *in the subspace of collinear spontaneous magnetizations, the ground state determines an applied magnetic field only up to some codirectional uniform field. The static $\boldsymbol{q} = 0$ spin susceptibility for a field in magnetization direction is zero.* Apart from the considered sufficient condition for this case there may additionally at most be certain accidental cases. The considered situation is the most general systematic possibility which can appear with a *spontaneous* spin-polarized ground state. Implications on half-metalicity are discussed in [Eschrig and Pickett, 2001].

5 Legendre Transformation

The Legendre transformation as a change of variables of convex functions is a familiar concept in physics. It was introduced into density functional theory by Lieb as a then completely new aspect of the transfer from v-dependences to n-dependences [Lieb, 1983]. Legendre transforms, or conjugate functionals in the modern terminology, are very powerful and very general tools of convex analysis [Zeidler, 1986, volume III, chapter 51] or [Young, 1969, §45]. This chapter introduces a few important mathematical concepts needed for handling conjugate functionals and thus provides a basis to understand the next chapter.

The first section presents the basic idea of a Legendre transform on the most simple intuitive level, but general enough to show some features important in our context although usually not considered in classical physical applications. We are firmly convinced that *working through the content of this first section is the minimum necessary for an understanding of what a Legendre transform really is.* The subsequent sections introduce some notions of Banach space and of duality theory in convex functional analysis, which are necessary for a deeper understanding of Lieb's density functional theory. In order to establish the connection of this theory with the Kohn-Sham equations, the mathematical notion of the functional derivative is considered. The last section of this chapter generalizes the notion of a Lagrange multiplier to (a slightly specialized version of) Fenchel's duality, which might become useful in further developments and applications of density functional theory.

Three different approaches to studying this chapter are possible: those who are not really interested in the mathematical concepts and want only to grasp the formal structure of Lieb's theory may carefully read the first section and cursorily scan through the rest of the chapter before passing to the next, the more physical one. A more careful reading of the present chapter provides a more rigorous introduction to the next one. It finally may be taken as a guide to those parts of convex functional analysis, which are relevant in connection with the mathematics of density functional theory, and then e.g. the content and the bibliography of [Zeidler, 1986] may be used for further studies.

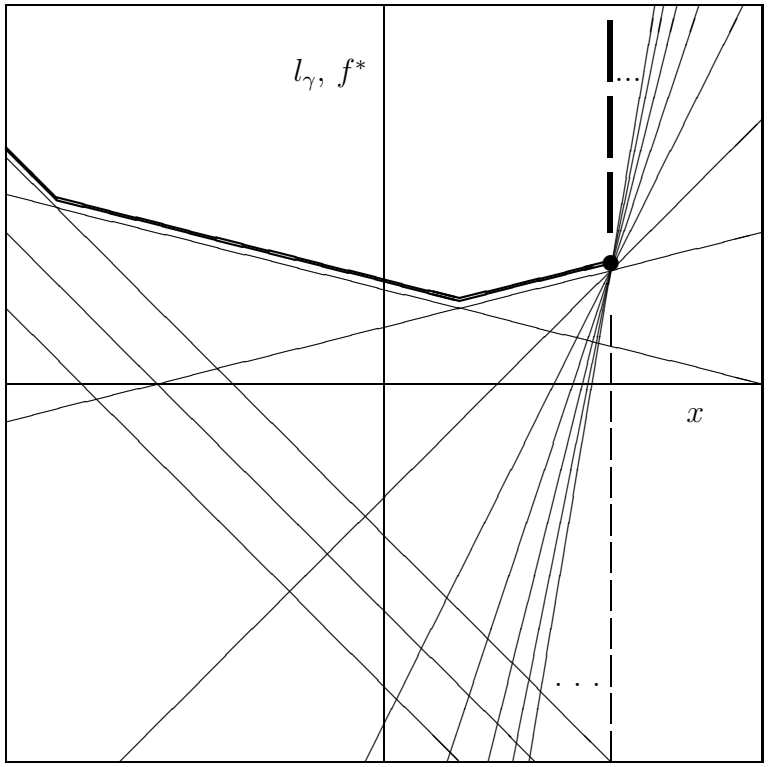

Figure 9: Supremum of affine-linear functions as explained in the text. The family is supposed to contain functions with unlimited positive slopes, all passing through the heavy dot. The supremum $f^*(x)$ (thick line) is equal to plus infinity for x-values right of the heavy dot.

For what follows, the reader is assumed to have an elementary understanding of normed linear spaces (as for example the notion of a Hilbert space).

5.1 Elementary Introduction

Consider a family of affine-linear functions on the real line $\boldsymbol{R}$

$$l_\gamma(x) = c_\gamma x - d_\gamma, \quad \gamma \in \Gamma \neq \emptyset, \tag{5.1}$$

as shown on Fig.9. Γ is some index set, supposed not to be empty. For every $x \in \boldsymbol{R}$, define $f^*(x)$ as the supremum over the family $l_\gamma(x)$. On Fig.9, $f^*(x)$ is shown as a thick line. Apparently, $f^*(x)$ may take on the value $+\infty$ at

least on part of $\boldsymbol{R}$, if c_γ is not bounded on Γ. It cannot take on the value $-\infty$, however, if Γ is not empty. Hence, f^* is *finite from below*:

$$f^*(x) \in \boldsymbol{R} \cup \{+\infty\}. \tag{5.2}$$

Moreover, f^* is *convex*:

$$\begin{aligned} &f^*(cx_1 + (1-c)x_2) \le cf^*(x_1) + (1-c)f^*(x_2) \\ &\qquad \text{for every } c \in [0,1] \text{ and every pair } x_1,\, x_2 \in \boldsymbol{R}, \end{aligned} \tag{5.3}$$

i.e., the graph of the function f^* is never above a chord between two points of that graph. This follows immediately from the affine-linearity of the l_γ's and from the simple property of a supremum of a sum not to exceed the sum of the respective independent suprema: $\sup l_\gamma(cx_1 + (1-c)x_2) = \sup\{cl_\gamma(x_1) + (1-c)l_\gamma(x_2)\} \le c \sup l_\gamma(x_1) + (1-c) \sup l_\gamma(x_2)$. Fig.9 shows also, that f^* need not be continuous: it may jump to $+\infty$. Analyzing the way this jump appears in Fig.9, one easily finds out that f^* takes on the lower value there. The function f^* is *lower semicontinuous*:

$$f^*(\lim x_n) \le \liminf f^*(x_n) \text{ for every converging sequence } (x_n). \tag{5.4}$$

If $\lim x_n = x$, then $\lim_{n'} l_\gamma(x_{n'}) = l_\gamma(x)$ for every subsequence $(x_{n'})$ of the sequence (x_n) because l_γ is continuous. Now, $\sup_\gamma\{\lim_{n'} l_\gamma(x_{n'})\} \le \lim_{n'}(\sup_\gamma l_\gamma(x_{n'}))$, if the limits on both sides exist, is again a general property of the supremum, the proof of which is a simple exercise. (The limes on the right side exists only for properly chosen subsequences; this is the reason for considering subsequences here. Consider as an example the marked dot of Fig.9. For any sequence (x_n) converging to the x-value of that dot, subsequences $(f^*(x_{n'}))$ of the sequence $(f^*(x_n))$ converge only, if all $x_{n'}$ are left of or equal to $\lim x_n$, the limes of $(f^*(x_{n'}))$ in this case being the ordinate value of the dot, or if all $x_{n'}$ are right of $\lim x_n$, the limes in this case being $+\infty$.) By definition, $\liminf(\sup_\gamma l_\gamma(x_n)) = c$ implies that there is a subsequence $(x_{n'})$ of (x_n) for which $\lim(\sup_\gamma l_\gamma(x_{n'})) = c$. Taking this subsequence, all that together yields $f^*(\lim x_n) = \sup_\gamma\{l_\gamma(\lim x_n)\} = \sup_\gamma\{\lim_{n'} l_\gamma(x_{n'})\} \le \lim_{n'}(\sup_\gamma l_\gamma(x_{n'})) = \liminf f^*(x_n)$. (To be precise, what we have used is the definition of sequential lower semicontinuity, whereas lower semicontinuity is defined somewhat differently in general. We will, however, only meet situations where both notions coincide.)

As is seen from Fig.9, not all functions $l_\gamma(x)$ are effective in defining $f^*(x)$. The supremum $f^*(x)$ does not change, if the left lower parallel lines are omitted. Hence, if we define $f(c) = \inf\{\, d_\gamma \mid c_\gamma = c \,\}$ as the infimum of all

negative ordinate sections d_γ over functions l_γ of the family with given slope $c_\gamma = c$ and put the infimum of the empty set equal to $+\infty$ (i.e., $f(c) = +\infty$, if the family $\{l_\gamma\}$ does not contain a function with slope c), then, for every slope c, $l_c(x) = cx - f(c)$ is the upper linear function in the family (5.1), and

$$f^*(x) = \sup_c \{cx - f(c)\}, \quad f : \boldsymbol{R} \to \boldsymbol{R} \cup \{+\infty\}, \quad f \not\equiv +\infty. \tag{5.5}$$

The last assumption on f not to be identically equal to $+\infty$ for all c corresponds to the earlier presupposition of Γ not to be empty. Note however, that the function $f(c)$, for which the above definition of $f^*(x)$ makes sense, is otherwise completely arbitrary and may be as wild as anybody can imagine, nevertheless

$$f^*(x) \in \boldsymbol{R} \cup \{+\infty\}, \qquad f^* \not\equiv +\infty,$$

$$f^* \text{ is convex and lower semicontinuous.} \tag{5.6}$$

$f^*(x)$ as defined by (5.5) is called the *Legendre transform* or the *conjugate function* to $f(c)$. Directly from the definition of f^* the generalized Young inequality

$$f^*(x) + f(c) \geq xc \tag{5.7}$$

follows. The classical Young inequality

$$xc \leq |c|^p/p + |x|^q/q \text{ when } 1/p + 1/q = 1 \tag{5.8}$$

is obtained from $f(c) = |c|^p/p$, for which $f^*(x) = |x|^q/q$ is easily verified by determining the supremum (5.5) (which is a maximum in this case) from the zero of the first derivative with respect to c.

Given x, equality holds in (5.7) for those c for which the affine-linear functions $l_c(x) = cx - f(c)$ pass through $(x, f^*(x))$, with

$$f^*(x') \geq f^*(x) + c(x' - x) \text{ for all } x' \in \boldsymbol{R} \tag{5.9}$$

being valid (cf. Fig.9). Such a c is called a *subgradient* of f^* at x. If it is unique, it is obviously equal to the derivative $c = df^*/dx$. But it need not be unique. At the x-value of the heavy dot of Fig.9 there are infinitely many subgradients to $f^*(x)$ of that figure: the slopes of all lines passing through that point and being otherwise below the graph of $f^*(x')$. Geometrically, a

straight line touching the graph of $f^*(x')$ at x and being otherwise below that graph is a *tangent of support* for f^* at x. A subgradient is the slope of a tangent of support. The *subdifferential* $\partial f^*(x)$ is defined to be the set of all subgradients of f^* at x. The subdifferential is the empty set, $\partial f^*(x) = \emptyset$, if no subgradient exists (e.g. if $f^*(x) = +\infty$ or if $f^*(x)$ would not be convex). With those definitions we have

$$f^*(x) + f(c) = xc \iff c \in \partial f^*(x) \tag{5.10}$$

for all x, $c \in \boldsymbol{R}$.

A simple relation between f and f^* follows directly from the definition (5.5):

$$f_1(c) \le f_2(c) \text{ for all } c \implies f_1^*(x) \ge f_2^*(x) \text{ for all } x. \tag{5.11}$$

It may become very useful for finding bounds.

Now we may consider the conjugate function $f^{**}(c) \stackrel{\text{def}}{=} (f^*)^*(c)$ to $f^*(x)$:

$$f^{**}(c) = \sup_x \{xc - f^*(x)\} \le \sup_x \{f^*(x) + f(c) - f^*(x)\} = f(c). \tag{5.12}$$

In the second relation the inequality (5.7) was used. Due to the complete generality of (5.6) with respect to f, it is clear that f^{**} has all properties (5.6); it is a lower semicontinuous convex minorant to f.

We will show that f^{**} is the *maximal* lower semicontinuous convex minorant of f. To this goal we need a completely natural but extremely important statement on convex functions:

> *For every point below the graph of a lower semicontinuous convex function there is a straight line separating the former from the latter, i.e. for which the graph of the convex function is entirely above the line, and the point is below.*

This is just the simplest variant of the famous Hahn-Banach theorem, which, loosely speaking, half of twentieth century mathematics is based upon. To prepare for the general functional case, we give a formal proof: Consider the convex function $f(c)$, and pick c_0 and $a_0 < f(c_0)$. Pick a so that $a_0 < a <$

$f(c_0)$. For every α, $\beta > 0$,

$$\begin{aligned}
\beta a + \alpha a &= (\alpha+\beta)a \; < \\
&< (\alpha+\beta) f(c_0) \; = \\
&= (\alpha+\beta)\, f\left(\frac{\beta}{\alpha+\beta}(c_0-\alpha) + \frac{\alpha}{\alpha+\beta}(c_0+\beta)\right) \; \leq \\
&\leq (\alpha+\beta)\left(\frac{\beta}{\alpha+\beta} f(c_0-\alpha) + \frac{\alpha}{\alpha+\beta} f(c_0+\beta)\right) \; = \\
&= \beta f(c_0-\alpha) + \alpha f(c_0+\beta),
\end{aligned}$$

where except for elementary algebra only the definition of convexity was used. Division by $\alpha\beta > 0$ and rearrangement of terms yields

$$\frac{1}{\alpha}\left[-f(c_0-\alpha)+a\right] < \frac{1}{\beta}\left[f(c_0+\beta)-a\right].$$

Since α and β are completely independent,

$$x_l \stackrel{\text{def}}{=} \sup_{\alpha}\left[\frac{1}{\alpha}(-f(c_0-\alpha)+a)\right] \leq \inf_{\beta}\left[\frac{1}{\beta}(f(c_0+\beta)-a)\right] \stackrel{\text{def}}{=} x_r.$$

From the lower semicontinuity of f it follows that $\limsup_{\alpha\downarrow 0}[-f(c_0-\alpha)+a] \leq [-f(c_0)+a] < 0$. Hence $x_l < +\infty$; and $-\infty < x_r$ is obtained in the same manner. Geometrically the above relation between x_l and x_r simply means that from the point (c_0, a) lying below the graph of the convex function $f(c)$ the slope of the tangent to the left part of this graph cannot exceed the slope of the tangent to the right. Now, there is x with $x_l \leq x \leq x_r$ and a_1 with $a_0 < a_1 < a$ so that $f(c_0-\alpha) > a_1 - x\alpha$, $f(c_0+\beta) > a_1 + x\beta$ for all α, $\beta > 0$. The graph of the affine-linear function $l(c) = a_1 + x(c-c_0)$ gives the wanted straight line ($f(c) > l(c)$ for all c). Note that x, the slope of the straight line, is always finite. It was necessary to take the point (c_0, a) *below* $f(c_0)$ in order to get through the proof with $f(c)$ taking possibly on the value $+\infty$: The strict inequality was needed to exclude that both bounds x_l and x_r might be simultaneously $+\infty$ or $-\infty$.

Coming to our main goal—the property of f^{**}, suppose that f be lower semicontinuous and convex (and $-\infty < f(c) \leq +\infty$). Pick c_0, and pick $a_0 < f(c_0)$. For $a_0 < a < f(c_0)$, there is $l(c) = a + x_0(c-c_0)$ with $l(c) < f(c)$ everywhere. Hence, for all c, $cx_0 - f(c) < c_0x_0 - a$. The supremum over c of the left side is $f^*(x_0)$. Thus, $f^*(x_0) \leq c_0x_0 - a < c_0x_0 - a_0$, or $a_0 < x_0c_0 - f^*(x_0) \leq f^{**}(c_0)$. Since a_0 may be chosen arbitrarily close to $f(c_0)$,

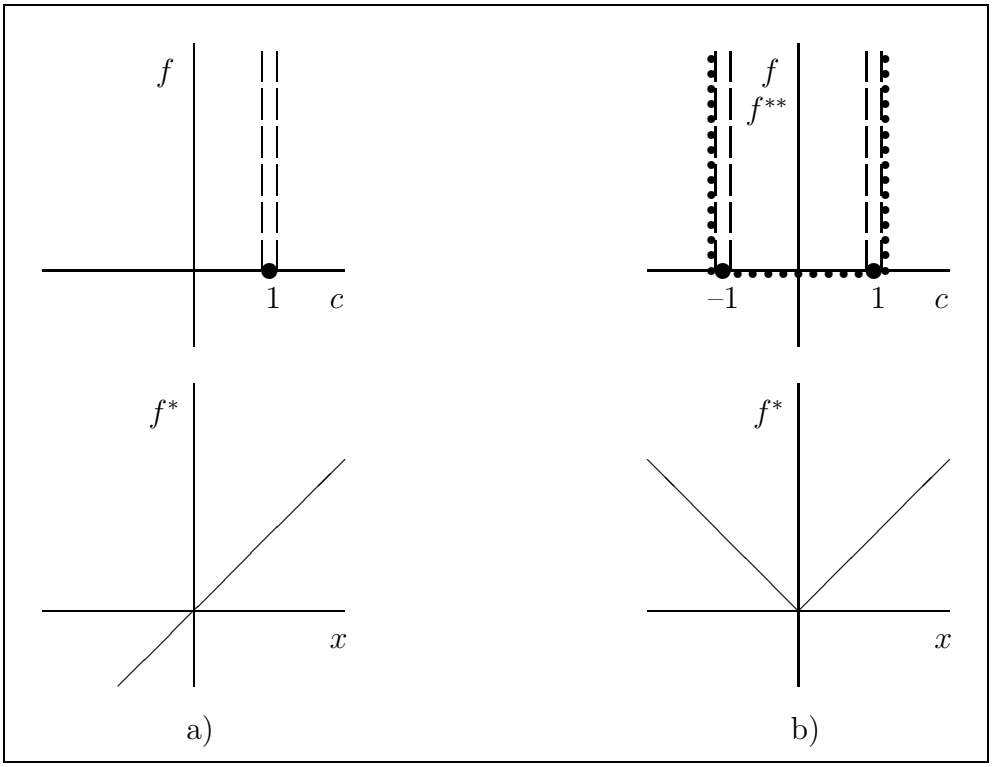

Figure 10: Examples of mutual Legendre transforms. a) $f(c) = 0$ for $c = 1$ and $f(c) = +\infty$ otherwise; $f^*(x) = x$. b) $f(c) = 0$ for $c = -1$, 1 and $f(c) = +\infty$ otherwise (dashed); $f^*(x) = |x|$; $f^{**}(c) = 0$ for $c \in [-1, 1]$ and $f^{**}(c) = +\infty$ otherwise (dotted).

this proves $f(c_0) \leq f^{**}(c_0)$. But we know already $f^{**}(c_0) \leq f(c_0)$, hence, as c_0 was arbitrary, $f^{**}(c) = f(c)$.

Considering again a general function $f(c)$ (not necessarily convex), let $f_0(c) \leq f(c)$ everywhere and f_0 lower semicontinuous and convex. According to (5.11), $f_0^*(x) \geq f^*(x)$ everywhere and hence $f_0^{**}(c) \leq f^{**}(c)$ everywhere. But $f_0^{**}(c) = f_0(c)$, hence $f_0(c) \leq f^{**}(c)$ everywhere: f^{**} is the maximal lower semicontinuous convex minorant to f, i.e. *the lower semicontinuous convex hull* of f, and

$$f^{**} \equiv f \Longleftrightarrow f \text{ is lower semicontinuous and convex.} \tag{5.13}$$

In this latter case f and f^* are called mutual Legendre transforms or mutual conjugate functions.

Let $f(c)$ be strictly convex, twice differentiable at c and $f''(c) > 0$. Then $\partial f(c) = \{f'(c)\}$, and from a relation conjugate to (5.10) $f^*(f'(c)) + f(c) =$

$cf'(c)$. Differentiating this equation a second time one finds $f^{*\prime}(f'(c))f''(c)+f'(c) = f'(c) + cf''(c)$, hence

$$f^{*\prime}(f'(c)) = c. \tag{5.14}$$

This means $\partial f^*(f'(c)) = \{c\}$, that is, $f^*(x)$ is differentiable at $x = f'(c)$ with the derivative equal to c. (Check (5.14) for $f(c) = |c|^p/p$.)

To compare to an elementary well known example, consider the Lagrange function $L(v) = mv^2/2$ of a free particle of mass m and with velocity v, and as its Legendre transform the corresponding Hamilton function $H(p) = \sup_v\{pv - L(v)\}$. The condition for the extremum is $\partial(pv - L)/\partial v = 0$, yielding $p = L'(v) = mv$ and $H(p) = p^2/2m$. The relation (5.14) takes in this case the form $H'(L'(v)) = H'(p) = p/m = v$. Another typical example is given by the inner energy $E(S, V)$ as a function of entropy S and volume V and the free energy $F(T, V)$ as a function of temperature T and volume, both being thermodynamic potentials and E and $-F$ being mutual Legendre transforms for fixed V: $-F(T, V) = \sup_S\{TS - E(S, V)\}$, $E(S, V) = \sup_T\{ST + F(T, V)$. Hence, E is convex in S (and in V) and F is concave in T (and convex in V), and the conditions for the extrema are the well known thermodynamic relations $S = -\partial F/\partial T$ and $T = \partial E/\partial S$. In most textbooks of Thermodynamics you may read $F = E - TS$. But this does not tell you much since it must be understood as $F(T, V) = E(S(T, V), V) - TS(T, V)$ with $S(T, V)$ implicitly given by $T = \partial E/\partial S$. Two simple instructive non-standard examples of mutual Legendre transforms are given on Fig.10.

The material of this section was presented in such a way that little has to be modified, if the real line $\boldsymbol{R}$ is replaced by a functional space and functions are replaced by functionals.

5.2 Prelude on Topology

In order to generalize analysis to functions of (possibly uncountably) infinitely many variables, the notion of a limes has to be put onto a broader basis. A sequence has a limes, if its elements get arbitrarily close to the limiting point as the subscript increases. This is alternatively expressed by saying that every neighborhood of the limiting point contains all but finitely many of the elements of the sequence. A neighborhood of a point is a set containing that point as an inner point; and it is precisely the notion of an open set which relates its points as inner points of the set. This way we are let to the concept of open sets as underlying the notion of a limes.

A set X of which all open subsets are distinguished is called a *topological space*. In order that this definition meets the notion in use, the following properties of the family $\mathcal{T}$ of all open subsets of X are to be added:

1. $\mathcal{T}$ is closed under finite intersections, i.e., if $A,\ B \in \mathcal{T}$, then $A \cap B \in \mathcal{T}$.
2. $\mathcal{T}$ is closed under unions, i.e., if $A_\gamma \in \mathcal{T}$ for all γ of some index set Γ, then $\cup_{\gamma\in\Gamma} A_\gamma \in \mathcal{T}$.
3. $\emptyset \in \mathcal{T}$ and $X \in \mathcal{T}$.
4. Any two different points $x,\ y \in X$ are contained in two disjunct open sets $A,\ B \in \mathcal{T}$: $x \in A,\ y \in B,\ A \cap B = \emptyset$ (Hausdorff).

$\mathcal{T}$ is the topology of the topological space $\{X,\ \mathcal{T}\}$. The family of all complements $X \setminus A$ of open sets A in X forms the family of all closed sets of the topological space. A subfamily $\mathcal{B} \subseteq \mathcal{T}$ is a base of the topology $\mathcal{T}$, if any $T \in \mathcal{T}$ can be obtained as $T = \cup_\gamma B_\gamma$ with $\{B_\gamma\} \subseteq \mathcal{B}$. (We included the Hausdorff property in our definition because we consider only topologies having this property; it guarantees that, if a limes exists, it is unique.)

Take, for example, the set X of all sequences (c_i), $i = 0, 1, 2, \ldots$ of complex numbers for which $\sum_i |c_i|^2 = 1$ (quantum states as expanded into a complete basis) and consider the families $\mathcal{B}_1 = \{B_{n,(z_i,\varepsilon_i)} = \{\,(c_i) \mid |c_i - z_i| < \varepsilon_i \text{ for } i < n,\ c_i \text{ arbitrary for } i \geq n\,\}\}$ (all open cylinders perpendicular to finitely many directions), $\mathcal{B}_2 = \{B_{(z_i),\varepsilon} = \{\,(c_i) \mid \sum_i (c_i - z_i)^2 < \varepsilon^2\,\}\}$ (all open ε-balls), $\mathcal{B}_3 = \{B_{(z_i)} = \{(z_i)\}\}$ (all sets consisting of a single point). They form bases of topologies $\mathcal{T}_1$, $\mathcal{T}_2$, $\mathcal{T}_3$, and for short we denote the corresponding topological spaces by $X_1 = \{X,\ \mathcal{T}_1\}$, $X_2 = \{X,\ \mathcal{T}_2\}$, and $X_3 = \{X,\ \mathcal{T}_3\}$. A point in X is (c_i), a sequence of points in X is $((c_i)_j) \equiv (c_{ij})$. Consider the sequences C_1 and C_2 of points given by $C_1 :\ c_{ij} = \delta_{ij}$, $C_2 :\ c_{0j} = (1 - 1/j)^{1/2}$, $c_{ij} = 1/j$ for $i = 1, \ldots, j$, $c_{ij} = 0$ for $i > j$. C_1 converges to (0) in X_1, because any $B_{n,(0,\varepsilon_i)}$ contains all $(c_i)_j$ for $j \geq n$. It does not converge to (0) nor to any other point in X_2, because $\sum_i (c_{ij} - c_{ik})^2 = 2$ for $j \neq k$. If any point of the sequence C_1 is in some $B_{(z_i),1/2}$ of X_2, then all the remaining points are outside of this ball of radius $1/2$. Each component c_i of C_1 converges individually, because $c_{ij} = 0$ for all $j > i$ and hence $\lim_{j\to\infty} c_{ij} = 0$, and this means convergence in X_1. For every j there is, however, a component c_i (with $i = j$) in C_1, which is away by one from the limiting value zero: no uniform convergence of all components of $(c_i)_j$ (for all i simultaneously), i.e., no convergence in X_2. C_2 converges to

(δ_{i0}) in both X_1 and X_2, as easily seen. None of the two sequences converges in X_3, because if a sequence of points converges to some (z_i) in X_3, then there must be some j_0 so that all $(c_i)_j \in B_{(z_i)}$ for $j > j_0$. But this means that all $c_{ij} = z_i$ for $j > j_0$. (So to say convergence in a finite number of steps.)

We have seen that convergence of a sequence is a weak statement in X_1, a stronger statement in X_2 and the strongest statement in X_3. As is easily seen from the above definition of a base of a topology, $\mathcal{T}_1 \subset \mathcal{T}_2 \subset \mathcal{T}_3$, or, in other words, the family of open sets of X_1 is a subset of the family of open sets of X_2, and that one is a subset of the family of open sets of X_3 (the latter comprising *all* sets of X because every set is a union of points). With respect to their strength, topologies $\mathcal{T}$ are (partially) ordered by the subset relation. In X_3, every subset of X is an open set (and therefore is every subset closed too). Hence this is the strongest of all possible topologies. (It is called the discrete topology because in a sense not discussed here all points are disconnected from each other.) If we, on the other hand, denote by $\mathcal{C}_i$ the set of all converging sequences in $(X, \mathcal{T}_i)$, then $\mathcal{C}_1 \supset \mathcal{C}_2 \supset \mathcal{C}_3$. (Note the reversed order.)

One subtlety of an important issue considered in the next paragraphs requires a generalization of sequences. A directed system is an index set Γ, partially ordered (by $\prec$) in such a way that for every two γ', $\gamma'' \in \Gamma$ there is a common successor $\gamma \in \Gamma$, $\gamma' \prec \gamma$, $\gamma'' \prec \gamma$. A net or generalized sequence (or Moore-Smith sequence) is a set $(x_\gamma)_{\gamma\in\Gamma}$ indexed by a directed system Γ. A section of a net (x_γ) is $(x_\gamma)_{\tilde{\gamma}\prec\gamma}$ for some $\tilde{\gamma}$, i.e. that part of the net with subscripts following $\tilde{\gamma}$. A net $(x_{\gamma'})_{\gamma'\in\Gamma'}$ is a subnet of $(x_\gamma)_{\gamma\in\Gamma}$, if every section of the latter contains all but finitely many points of the former. As for ordinary sequences (being special nets), a *net* (x_γ) *converges* to a limes x in a topological space, if every open set containing x also contains all but finitely many of the x_γ. Unlike an ordinary sequence, a net may consist of uncountably many points, and that is why it is introduced. Nets are needed each time, if the topology of the space X at point x cannot locally be characterized by a countable set of neighborhoods.

One important concept of topology is compactness: A set C of a topological space $\{X, \mathcal{T}\}$ is compact, if every open cover of C, i.e. $C \subseteq \cup_{U\in\mathcal{U}} U$, $\mathcal{U} \subseteq \mathcal{T}$, contains a finite subcover: $C \subseteq \cup_{i=1}^n U_i$, $U_i \in \mathcal{U}$. This is important because of the famous Bolzano-Weierstrass theorem: *A set C of a topological space is compact if and only if every net in C has a convergent subnet.* (The limes of a subnet is a cluster point of the net: every neighborhood of it contains infinitely many points of the net; however, unlike the case of the limes

of the net itself, another infinite number of points may still be outside.) In $\boldsymbol{R}^N$ with its ordinary topology the closed unit ball and every closed bounded set is compact; reversely, every compact set is closed and bounded. It is a major difference for analysis that in infinite-dimensional normed spaces the closed unit ball is not compact. In the above examples, the sequence C_1, though bounded by the closed unit ball of X_2, does not have a subsequence converging in X_2.

In the following we will only meet two types of topological spaces: metric spaces and locally convex spaces. In a metric space a non-negative distance of points is defined for all pairs of points, which is non-zero if the points are different, and which obeys the ordinary triangle inequality. A base of its topology is the family of all ε-balls centered at all points. In a metric space, nets are unnecessary; they may be replaced by sequences in all statements. In a linear metric space the distance $d(x, y)$ is often defined via a norm:

$$d(x, y) = \|x - y\|, \tag{5.15}$$

where a norm is characterized as usual by the properties

$$\|\alpha x\| = |\alpha|\|x\|, \quad \|x + y\| \leq \|x\| + \|y\|,$$

$$\|x\| = 0 \text{ if and only if } x = 0. \tag{5.16}$$

If the last property is abandoned, one speaks of a seminorm.

Unfortunately, normed linear spaces are often a too narrow frame in functional analysis. A locally convex space is by definition a linear topological space whose topology is given by a family of seminorms $\{p_j\}_{j\in J}$ such that

$$x = 0 \text{ if and only if } p_j(x) = 0 \text{ for all } j \in J. \tag{5.17}$$

(This property and likewise the last property (5.16) of a norm guarantees that the so defined topology is Hausdorff.) A base of this topology is

$$B_{n,(j_k),z,\varepsilon} = \{\, x \,|\, p_{j_k}(x - z) < \varepsilon,\ k = 1, \ldots, n \,\}. \tag{5.18}$$

The sets $B_{n,(j_k),z,\varepsilon}$ are absolutely convex, i.e. $x,\, y \in B_{n,(j_k),z,\varepsilon}$ and $\alpha \in [0, 1]$ implies $\alpha x + (1 - \alpha)y \in B_{n,(j_k),z,\varepsilon}$, and $x \in B_{n,(j_k),z,\varepsilon}$ and $|\alpha| \leq 1$ implies $\alpha x + (1 - \alpha)z \in B_{n,(j_k),z,\varepsilon}$. This property is giving the type of spaces its name.

In the special case where the family of seminorms consists of a single norm, we are again left with a metric linear space whose metric is given by

a norm. X_1 of the above given examples of topological spaces is a locally convex space whose topology was not given via a metric. Nevertheless it is metrizable by

$$d((c_i)_j, (c_i)_k) = \sum_{i=1}^{\infty} 2^{-i} \left[\frac{|c_{ij} - c_{ik}|}{1 + |c_{ij} - c_{ik}|} \right]. \tag{5.19}$$

This metric generates the same topology as the family of seminorms $p_l((c_i)_j, (c_i)_k) = |c_{lj} - c_{lk}|$ from which it is derived, but this metric cannot be derived from a single norm. However, countability of the family of seminorms was essential in this construction of the metric.[6] In the general case of a locally convex space, the topology of which is not metrizable, nets are needed to exhaust all possible limiting processes.

Another basic concept for sequences of real numbers one would like to maintain in a general situation is Cauchy's convergence criterion. In a metric space, a sequence (x_i) is called Cauchy, if for every $\varepsilon > 0$ there is an i_ε so that $d(x_i, x_j) < \varepsilon$ for all i, $j > i_\varepsilon$. This is expressed by writing

$$\lim_{i,j\to\infty} d(x_i, x_j) = 0. \tag{5.20}$$

A metric space is called complete, if every Cauchy sequence converges to a point of the space. (Every metric space may be completed by adding all Cauchy sequences, not yet converging in it, just as additional points; the real line is obtained by completing the rational line this way.) A normed linear space complete in the metric of the norm is called a *Banach space*.

In a locally convex space, a net (x_γ) is accordingly called Cauchy, if for every $\varepsilon > 0$ and every seminorm p_j there is a section cut by $\gamma_{\varepsilon,j}$ so that $p_j(x_{\gamma'} - x_{\gamma''}) < \varepsilon$ in the section, i.e. for all γ', $\gamma'' \succ \gamma_{\varepsilon,j}$. A locally convex space is called complete, if every Cauchy net converges. (It is—less demanding—sequentially complete, if every Cauchy sequence converges.)

The basic maps of topological spaces are continuous functions. Recall that a function $f : X \to Y$ is continuous, if the preimage $f^{-1}(A)$ of every open set $A \subseteq Y$ is an open set of X. This is equivalent to saying that for every net (x_γ) converging to x in X the net $(f(x_\gamma))$ converges to $f(x)$ in Y. Indeed, suppose that $(f(x_\gamma))$ does not converge to $f(x)$. This means that

[6] *Complete* metric spaces of that type are called Frechet spaces. They play a fundamental role in the theory of generalized functions as well as in the modern theory of analytic complex functions. The relation between the considered topological linear spaces is {locally convex spaces}⊃{Frechet spaces}⊃{Banach spaces}⊃{Hilbert spaces}⊃{finite-dimensional Euklidian spaces}.

infinitely many points $f(x_\gamma)$ of the net are outside of some open neighborhood U of $f(x)$. Hence, infinitely many points x_γ are outside of $f^{-1}(U) \ni x$ implying that either $f^{-1}(U)$ is not open or (x_γ) does not converge to x. Important examples of continuous functions from a linear topological space onto the real line $\boldsymbol{R}$ with its ordinary topology given by open intervals of real numbers are the norm $x \mapsto \|x\|$ and the seminorm $p_j(x)$. In fact the topologies derived from a norm or a family of seminorms as above may alternatively be defined as the weakest topologies in which the norm or all seminorms, respectively, are continuous.

(A bijective map $f : X \to Y$ is called a homeomorphism, if both f and f^{-1} are continuous functions. If it exists, then both spaces X and Y are equivalent with respect to their topologies. Hence continuous functions introduce into the class of topological spaces the structure of an algebraic category of objects and morphisms; see e.g. [Lang, 1965, §7].)

5.3 Prelude on Lebesgue Integral

In linear functional spaces a norm is usually defined by some integral over a real-valued function. In order to obtain a *complete* metric space with points closest to the ordinary notion of function, a broader notion of integral than that of Riemann is necessary. We will, however, by far not consider the most general case as we need only the Lebesgue integral over real- or complex-valued functions on the real N-dimensional Euclidean space $\boldsymbol{R}^N$. This starts with non-negative real-valued functions $f(\boldsymbol{x})$, where the contribution of function values above some positive number y to the integral is estimated from below by the value y times the measure of the domain on which $f(\boldsymbol{x}) > y$, and this measure has to be defined.

The simplest measurable subset of $\boldsymbol{R}^N$ is an open brick A (open in the topology of the Euclidean metric, hence N-dimensional). Its measure $\mu(A)$ is naturally equal to the product of the edge lengths of A. The next step is the measure of countable unions of disjoint open bricks, defined as

$$\mu\left(\cup_{i=1}^{\infty} A_i\right) \stackrel{\text{def}}{=} \sum_{i=1}^{\infty} \mu(A_i), \qquad A_i \cap A_j = \emptyset \text{ for } i \neq j, \tag{5.21}$$

where the sum may be $+\infty$. We denote the family of countable unions of disjoint open bricks by $\mathcal{A}$.

This suffices to define the Rieman integral, the *Lebesgue measure* is, however, defined for a much broader family of sets called the *Borel sets* and

being defined as the smallest family $\mathcal{B}$ of subsets of $\boldsymbol{R}^N$ with the properties

1. $\mathcal{B}$ contains all open bricks of $\boldsymbol{R}^N$.
2. $\mathcal{B}$ is closed under countable unions,
3. $\mathcal{B}$ is closed under complements in $\boldsymbol{R}^N$,

Hence $\mathcal{B}$ contains countable unions of complements of countable unions of complements of ... of bricks. This comprises practically all *conceivable* subsets of $\boldsymbol{R}^N$. Although the existence of non-Borel subsets of $\boldsymbol{R}^N$ can be proved, there is little more than no idea what they actually look like. For every $B \in \mathcal{B}$,

$$\mu(B) \stackrel{\text{def}}{=} \inf_{B \subseteq A \in \mathcal{A}} \mu(A). \tag{5.22}$$

Here, an essential point is that the infimum is not merely over finite unions of bricks as intuition would suggest, but over countable unions. This Lebesgue measure has all the defining properties of a general regular measure on $\mathcal{B}$:

1. $\mu(\emptyset) = 0$,
2. μ is σ-additive, i.e. for a countable family $\{B_i\}_{i=1}^{\infty} \subseteq \mathcal{B}$ of mutually disjoint sets, $\mu(\cup B_i) = \sum \mu(B_i)$,
3. $\mu(B) = \inf\{\, \mu(O) \,|\, B \subseteq O,\ O \text{ open} \,\}$,
4. $\mu(B) = \sup\{\, \mu(C) \,|\, C \subseteq B,\ C \text{ compact} \,\}$.

A real-valued function f is called measurable, if for every open interval $I = (a, b)$ the set $f^{-1}(I)$ is measurable (i.e. $f^{-1}(I) \in \mathcal{B}$). In this case, $f_+ = \max\{f, 0\}$ and $f_- = \max\{-f, 0\}$ are also measurable. Because of the third property of $\mathcal{B}$, for every closed or semiclosed interval I', $f^{-1}(I')$ is measurable too. Let $n \geq 0$ be integer, then

$$\begin{aligned} \int d^N x\, f_+(\boldsymbol{x}) &\stackrel{\text{def}}{=} \lim_{n\to\infty} \left\{ \sum_{m=0}^{\infty} \frac{m}{2^n}\, \mu\left(f_+^{-1}\left(\left[\frac{m}{2^n}, \frac{m+1}{2^n} \right) \right) \right) \right\}, \\ \int d^N x\, f(\boldsymbol{x}) &\stackrel{\text{def}}{=} \int d^N x\, f_+(\boldsymbol{x}) - \int d^N x\, f_-(\boldsymbol{x}). \end{aligned} \tag{5.23}$$

Because the expression in braces in the first line is non-negative and, as easily seen, not decreasing as a function of n, the limes exists (it may be

$+\infty$). The definition of the second line is only valid if at least one of the two terms on the right side is finite. For a complex-valued function the integral is defined by separate integration over the real and imaginary parts.

The meaning of this definition is to divide the range of f into intervals of width $1/2^n$ and to look for the sets in the domain of f for which the function values fall into that interval. The contribution of this set to the integral is then estimated from below with the help of the Lebesgue measure of the set ($m/2^n$ estimates the function value from below on the set). Call this estimate (the expression in braces) I_n. As was already stated, $I_{n+1} \geq I_n$. If I_n as a function of n is not bounded, then its limes is $+\infty$, and the integral is defined to be $+\infty$. If I_n is bounded, then, as a monotone sequence, it has a finite limes I.

Consider a domain $\Omega \subset \boldsymbol{R}^N$ of finite measure $\mu(\Omega)$, and replace the measure in the expression in braces of (5.23) by $\mu(\Omega \cap f_+^{-1}([m/2^n, (m+1)/2^n)))$. Call the expression in braces now $I_n(\Omega)$, it estimates the integral over the domain Ω from below. Replacing the prefactor $m/2^n$ by $(m+1)/2^n$, now estimates the function value from above. Call this expression $J_n(\Omega)$. Now, $J_{n+1}(\Omega) \leq J_n(\Omega)$ by considering one step of refinement. It is easily seen that $I_n(\Omega) \leq J_n(\Omega) \leq I_n(\Omega) + \mu(\Omega)/2^n$, thus both estimates have the same limes, equal to $\int_\Omega d^N x\, f_+(x)$. For every increasing sequence (Ω_i), $\Omega_i \subseteq \Omega_{i+1}$, $\cup_i \Omega_i = \boldsymbol{R}^N$, the limes of $\int_{\Omega_i} d^N x\, f_+(x)$ is equal to the first line of (5.23) as easily seen. This justifies the definition (5.23).

This Lebesgue integral has all the ordinary properties of an integral and coincides with the Riemann integral for all functions for which the latter is defined. The value of the integral remains unchanged, if the function value $f(\boldsymbol{x})$ is arbitrarily changed for $\boldsymbol{x}$-points forming a set of zero Lebesgue measure in $\boldsymbol{R}^N$. (Recall that the Lebesgue measure in $\boldsymbol{R}^N$ depends by definition on N: e.g. the measure of the set of all points on the x_1-axis of the $\boldsymbol{R}^2$ is zero, whereas the measure in $\boldsymbol{R}^1$ of this set as a set of $\boldsymbol{R}^1$ is infinite; the first is an area with zero extension in one direction, whereas the second is an infinite interval length.) While for the Riemann integral only changes of the function values at nowhere dense points are harmless, there are plenty of dense sets of Lebesgue measure zero, e.g. the set of all rational points of $\boldsymbol{R}^1 = \boldsymbol{R}$. There are even plenty of uncountable sets of Lebesgue measure zero, e.g. the well known Cantor set. This only indicates how far one has to leave comprehension, if one wants to treat all Cauchy sequences of functions as functions (see next section).

A statement is said to hold true *almost everywhere* (a.e.), if it holds true for all points but a set of zero measure.

5.4 Banach Space

As was already stated in Section 5.2, a Banach space is defined as a normed linear space complete in the metric of the norm. Besides this definition, there is one important property which is decisive for a normed linear space to be Banach.

Let X be a normed linear space. Then the following two statements are equivalent:

- *X is a Banach space.*
- *$x_n \in X$ and $\sum_{n=0}^{\infty} \|x_n\| < \infty \quad \Longrightarrow \quad \sum_{n=0}^{\infty} x_n = s \in X$.*

The sequence (x_n) of the second statement is called *absolutely summable.* To prove this theorem, let first X be Banach, and let $\sum \|x_n\| < \infty$. Denote $\sum_{n=0}^{m} x_n$ by s_m. We must prove $\lim s_m = s \in X$. Because of the convergence of the series of norms, there is a $p(\varepsilon)$ for every $\varepsilon > 0$ so that for every $m \geq p(\varepsilon)$ and for every k, $\sum_{n=m+1}^{m+k} \|x_n\| \leq \varepsilon$ and hence $\| \sum_{n=m+1}^{m+k} x_n\| \leq \varepsilon$. This latter relation means $\|s_{m+k} - s_m\| \leq \varepsilon$, that is, (s_m) is Cauchy, and hence it converges to some $s \in X$. Let now the second statement of the theorem be true. This time we have to prove that every Cauchy sequence converges in X. Let (x_n) be Cauchy, this means that for every integer $k \geq 0$ there is a $p(k)$ so that for $m, n \geq p(k)$, $\|x_m - x_n\| \leq 1/2^k$. We choose $p(k+1) > p(k)$ for every k and consider the series $x_{p(0)} + \sum_{k=0}^{\infty}(x_{p(k+1)} - x_{p(k)})$. It converges in X by our assumption, because the corresponding series of norms converges. Its limes equals the limes of the Cauchy sequence (x_n) which hence exists in X. This proves that X is Banach.

The simplest Banach space is $\boldsymbol{R}$ with the norm $\|x\| = |x|$ given by the absolute value. If we were to confine ourselves to this trivial case only, we would have been done with Section 5.1. A simple generalization would be $\boldsymbol{R}^N$. The goal of this whole chapter is to transfer all the content of Section 5.1 to uncomparably much richer structures. It may literally be transfered to any Banach space one can invent, we will, however, only consider $\boldsymbol{L}^p$ spaces, the points of which may be characterized by Lebesgue measurable functions.

Let $p \geq 1$ and let $\mathcal{L}^p(\boldsymbol{R}^N)$ be the set of Lebesgue measurable real- or complex-valued functions $f : \boldsymbol{R}^N \to \boldsymbol{R}$ or $\boldsymbol{C}$ for which

$$\int d^N x \, |f(\boldsymbol{x})|^p < \infty. \tag{5.24}$$

Gather all functions coinciding almost everywhere on $\boldsymbol{R}^N$ with a given $f \in \mathcal{L}^p(\boldsymbol{R}^N)$ into one equivalence class. (If $g(\boldsymbol{x}) = f(\boldsymbol{x})$ a.e. and $h(\boldsymbol{x}) = f(\boldsymbol{x})$ a.e., then $g(\boldsymbol{x}) = h(\boldsymbol{x})$ a.e. since the union of two sets of zero measure has zero measure; hence $g(\boldsymbol{x}) = f(\boldsymbol{x})$ a.e. is indeed an equivalence relation.) The obtained equivalence class is again denoted by f and called a p-summable function (although it is in fact a class of functions). Obviously those equivalence classes are compatible with addition of functions (compatible means that, if we take any two functions out of two given equivalence classes and add them, the result will always be in one and the same equivalence class; this is again true because the union of two sets of zero measure has zero measure). The equivalence classes are also compatible with multiplication of a function by a real or complex number. Furthermore, the condition (5.24) is stable under linear combinations. Hence, p-summable functions form a linear space over the scalar field of real or complex numbers. This space, equipped with the norm

$$\|f\|_p \stackrel{\text{def}}{=} \left(\int d^N x \, |f(\boldsymbol{x})|^p \right)^{1/p}, \quad 1 \le p < \infty \tag{5.25}$$

is a Banach space, the (real or complex) *Lebesgue space* $\boldsymbol{L}^p(\boldsymbol{R}^N)$, or in short $\boldsymbol{L}^p$, of p-summable functions.

To be indeed a norm, (5.25) must have the three properties (5.16), the first of which is trivially fulfilled for every scalar (i.e. constant on $\boldsymbol{R}_N$) α. Starting with Young's inequality (5.8), one has for $f \in \boldsymbol{L}^p$, $g \in \boldsymbol{L}^q$, $1/p + 1/q = 1$ (and hence p, $q > 1$) with $\|f\|_p$, $\|g\|_q \neq 0$

$$\int d^N x \left| \frac{f(\boldsymbol{x})}{\|f\|_p} \frac{g(\boldsymbol{x})}{\|g\|_q} \right| \le \int d^N x \left(\frac{|f(\boldsymbol{x})|^p}{p\|f\|_p^p} + \frac{|g(\boldsymbol{x})|^q}{q\|g\|_q^q} \right) = \frac{1}{p} + \frac{1}{q} = 1,$$

hence

$$\int d^N x \, |f(\boldsymbol{x}) g(\boldsymbol{x})| \le \|f\|_p \|g\|_q \text{ for } 1/p + 1/q = 1. \tag{5.26}$$

This is Hölder's inequality. If at least one of the norms on the right side of the inequality is zero, then $f(\boldsymbol{x})g(\boldsymbol{x}) = 0$ a.e., hence the inequality is valid also in this case. Next, because of $|f+g| \le |f| + |g|$, for f, $g \in \boldsymbol{L}^p$,

$$\int d^N x\, |f(\boldsymbol{x}) + g(\boldsymbol{x})|^p \le$$

$$\le \int d^N x\, |f(\boldsymbol{x})| \cdot |f(\boldsymbol{x}) + g(\boldsymbol{x})|^{p-1} + \int d^N x\, |g(\boldsymbol{x})| \cdot |f(\boldsymbol{x}) + g(\boldsymbol{x})|^{p-1}.$$

Applying to both right hand integrals Hölder's inequality yields

$$\|f+g\|_p^p \le \|f\|_p \| \, |f+g|^{p-1} \, \|_q + \|g\|_p \| \, |f+g|^{p-1} \, \|_q.$$

Since $(p-1)q = p$, the common right hand factor is $\| \, |f+g|^{p-1} \, \|_q = (\int d^N x\, |f+g|^{(p-1)q})^{1/q} = (\int d^N x\, |f+g|^p)^{(1/p)(p/q)} = \|f+g\|_p^{p/q}$ (wherefore it exists on grounds of our presupposition on f and g). Considering finally $p - p/q = 1$, the norm property

$$\|f+g\|_p \le \|f\|_p + \|g\|_p \tag{5.27}$$

is obtained, which in the case of Lebesgue space is called Minkowski's inequality. It holds also true for $p = 1$, which is immediately seen from the definition of the $\boldsymbol{L}^1$-norm. The last property (5.16) of a norm is guaranteed by the convention, in $\boldsymbol{L}^p$ not to make a difference between a function f non-zero on a set of zero measure only and the function $f \equiv 0$.

Finally, the proof of completeness of $\boldsymbol{L}^p$, i.e.

$$f_i \in \boldsymbol{L}^p \text{ and } \lim_{i,j\to\infty} \|f_i - f_j\|_p = 0 \quad \Longrightarrow \quad \lim_{i\to\infty} f_i = f \in \boldsymbol{L}^p, \tag{5.28}$$

may be found in standard textbooks on functional analysis (Riesz-Fisher theorem).

Let f be an equivalence class (in the above sense) of Lebesgue measurable functions on $\boldsymbol{R}^N$ such that $f(\boldsymbol{x})$ is a.e. bounded. This property is apparently stable under linear combinations, and hence these classes form a linear space. Consider

$$\|f\|_\infty \stackrel{\text{def}}{=} \text{ ess sup } |f(\boldsymbol{x})| \stackrel{\text{def}}{=} \inf\{\, c \,|\, |f(\boldsymbol{x})| \le c \text{ a.e.} \,\}. \tag{5.29}$$

This is a norm by the properties of an absolute value and because a union of two sets of zero measure has zero measure. Due to σ-additivity of a measure, even a countably infinite union of sets of zero measure has

zero measure, and hence the presupposition of (5.28) for $p = \infty$ means $\lim_{i,j\to\infty} |f_i(\boldsymbol{x}) - f_j(\boldsymbol{x})| = 0$ for almost all $\boldsymbol{x}$. This implies $\lim f_i(\boldsymbol{x}) = f(\boldsymbol{x})$ for almost all $\boldsymbol{x}$ and some $f(\boldsymbol{x})$, since $\boldsymbol{R}$ or $\boldsymbol{C}$, respectively, is complete. Choose a subsequence $f_{i'}$ of the Cauchy sequence f_i so that $\|f_{i'} - f_{i'+1}\|_\infty < 2^{-i'}$. Then, $1 > \sum_{i'=1}^\infty \|f_{i'} - f_{i'+1}\|_\infty \geq \sum_{i'=1}^\infty |f_{i'}(\boldsymbol{x}) - f_{i'+1}(\boldsymbol{x})| \geq |\sum_{i'=1}^\infty (f_{i'}(\boldsymbol{x}) - f_{i'+1}(\boldsymbol{x}))| = |f_{1'}(\boldsymbol{x}) - f(\boldsymbol{x})|$ for almost all $\boldsymbol{x}$, and hence $\|f\|_\infty < \|f_{1'}\|_\infty + 1$. We have shown that (5.28) extents to $p = \infty$, and the so defined normed space, denoted by $\boldsymbol{L}^\infty(\boldsymbol{R}^N)$ or shortly $\boldsymbol{L}^\infty$, is a Banach space. As is easily seen, even (5.26) extents to the case $p = 1$ and $q = \infty$.

Note that, as another special case, $\boldsymbol{L}^2(\boldsymbol{R}^N)$ is a Hilbert space with the scalar product $(f|g) = \int d^N x\, f^*(\boldsymbol{x}) g(\boldsymbol{x})$. (It contains the usual Hilbert space of quantum mechanics as a subspace.) For this case, (5.26) implies according to $|(f|g)| = |\int d^N x\, f^*(\boldsymbol{x}) g(\boldsymbol{x})| \leq \int d^N x\, |f(\boldsymbol{x}) g(\boldsymbol{x})| \leq \|f\|_2 \|g\|_2$ the well known Schwarz inequality

$$|(x|y)| \leq \|x\|\, \|y\| \tag{5.30}$$

valid in every Hilbert space. (A Hilbert space is just a special case of a Banach space whose norm is given via a scalar product.)

Let $p \leq t \leq p'$ and $\int d^N x\, |f|^p < \infty$, $\int d^N x\, |f|^{p'} < \infty$. Define the complementary sets $A = \{\boldsymbol{x} \,|\, |f(\boldsymbol{x})| \leq 1\}$ and $B = \{\boldsymbol{x} \,|\, |f(\boldsymbol{x})| > 1\}$. Both sets are measurable, and $\int d^N x\, |f|^t \leq \int_A d^N x\, |f|^p + \int_B d^N x\, |f|^{p'} < \infty$. Hence,

$$f \in \boldsymbol{L}^p \text{ and } f \in \boldsymbol{L}^{p'} \quad \Longrightarrow \quad f \in \boldsymbol{L}^t \text{ for } p \leq t \leq p'. \tag{5.31}$$

Note that $\boldsymbol{L}^p \cap \boldsymbol{L}^{p'}$ is again a Banach space with respect to the norm

$$\|f\|_{p\ldots p'} \stackrel{\text{def}}{=} \sup_{p \leq t \leq p'} \|f\|_t. \tag{5.32}$$

The first and third properties (5.16) of a norm are evidently fulfilled for this norm. Furthermore, $\|f_1 + f_2\|_{p\ldots p'} = \sup_{p\leq t\leq p'} \|f_1 + f_2\|_t \leq \sup_{p\leq t\leq p'} (\|f_1\|_t + \|f_2\|_t) \leq \|f_1\|_{p\ldots p'} + \|f_2\|_{p\ldots p'}$. Finally, a sequence which is Cauchy in the norm $\|\cdot\|_{p\ldots p'}$ is evidently Cauchy in every norm $\|\cdot\|_t$ for $p \leq t \leq p'$. Hence it converges to an $f \in \boldsymbol{L}^t$ for every t with $p \leq t \leq p'$.

Next, consider the family of Banach spaces $\boldsymbol{L}^p(\boldsymbol{T}^N)$, $1 \leq p \leq \infty$, on the N-dimensional torus of *finite* total measure L^N, defined by periodic boundary conditions $x_i \equiv x_i + L$, $i = 1, \ldots, N$, cf. (1.28). Let $f \in \boldsymbol{L}^p(\boldsymbol{T}^N)$, i.e., $\int_{\boldsymbol{T}^N} d^N x\, |f|^p < \infty$, and let again $A = \{\boldsymbol{x} \,|\, |f(\boldsymbol{x})| \leq 1\}$ and $B = \{\boldsymbol{x} \,|\, |f(\boldsymbol{x})| > 1\}$. This time, $\int_A d^N x\, |f|^t \leq L^N < \infty$ for every $t \geq 1$, and $\int_B d^N x\, |f|^t \leq \int_B d^N x\, |f|^p$ for every $1 \leq t \leq p$. Hence,

$$f \in \boldsymbol{L}^p(\boldsymbol{T}^N) \quad \Longrightarrow \quad f \in \boldsymbol{L}^t(\boldsymbol{T}^N) \text{ for all } 1 \leq t \leq p. \tag{5.33}$$

Especially, $f \in \boldsymbol{L}^\infty(\boldsymbol{T}^N) \Rightarrow f \in \boldsymbol{L}^t(\boldsymbol{T}^N)$ for all $t \geq 1$.

Finally, consider the linear hull of the spaces $\boldsymbol{L}^q(\boldsymbol{R}^N)$ and $\boldsymbol{L}^{q'}(\boldsymbol{R}^N)$, that is, the space of equivalence classes of Lebesgue measurable functions $f(\boldsymbol{x})$, which may be represented as $f = g + h$, $g \in \boldsymbol{L}^q(\boldsymbol{R}^N)$, $h \in \boldsymbol{L}^{q'}(\boldsymbol{R}^N)$. This space is denoted by $\boldsymbol{L}^q(\boldsymbol{R}^N) + \boldsymbol{L}^{q'}(\boldsymbol{R}^N)$. It is a Banach space with the norm

$$\|f\|_{(qq')} \stackrel{\text{def}}{=} \inf\{\|g\|_q + \|h\|_{q'} \,|\, g + h = f\}. \tag{5.34}$$

The infimum is taken over all possible decompositions of f into a sum $g+h$. Again, the first and the third property (5.16) are evident for this norm. Moreover, $\|f_1 + f_2\|_{(qq')} = \inf\{\|g_1 + g_2\|_q + \|h_1 + h_2\|_{q'} \,|\, g_1 + g_2 + h_1 + h_2 = f_1 + f_2\} \leq \inf\{\|g_1\|_q + \|g_2\|_q + \|h_1\|_{q'} + \|h_2\|_{q'} \,|\, g_1 + g_2 + h_1 + h_2 = f_1 + f_2\} \leq \inf\{\|g_1\|_q + \|h_1\|_{q'} \,|\, g_1 + h_1 = f_1\} + \inf\{\|g_2\|_q + \|h_2\|_{q'} \,|\, g_2 + h_2 = f_2\} = \|f_1\|_{(qq')} + \|f_2\|_{(qq')}$. For every absolutely summable series $\sum \|f_n\|_{(qq')} < \infty$, by definition of the norm (5.34) there is g_n, h_n so that $\|g_n\|_q < 2\|f_n\|_{(qq')}$, $\|h_n\|_{q'} < 2\|f_n\|_{(qq')}$ and hence $\sum \|g_n\|_q < \infty$, $\sum \|h_n\|_{q'} < \infty$. Since $\boldsymbol{L}^q$ and $\boldsymbol{L}^{q'}$ are Banach, it follows $\sum g_n = g \in \boldsymbol{L}^q(\boldsymbol{R}^N)$, $\sum h_n = h \in \boldsymbol{L}^{q'}(\boldsymbol{R}^N)$ and therefore $\sum f_n = \sum(g_n + h_n) = g + h \in \boldsymbol{L}^q(\boldsymbol{R}^N) + \boldsymbol{L}^{q'}(\boldsymbol{R}^N)$. Thus we have shown that every absolutely summable series converges in that space, and hence $\boldsymbol{L}^q(\boldsymbol{R}^N) + \boldsymbol{L}^{q'}(\boldsymbol{R}^N)$ is a Banach space.

Let X and Y be two normed linear spaces, and consider bounded linear operators $\hat{A}: X \to Y$ defined on X and having values in Y so that

$$\|\hat{A}\| \stackrel{\text{def}}{=} \sup_{x \in X,\, x \neq 0} \frac{\|\hat{A}x\|_Y}{\|x\|_X} < \infty. \tag{5.35}$$

It is easily seen that $\hat{A}$ is continuous at $x = 0$ and hence, due to the linearity of X, Y and $\hat{A}$, on all X. Continuity of the linear operator means that the x-points for which $\|\hat{A}x\|_Y < \varepsilon$ must form an open neighborhood of $x = 0$. That is, there must be a $\delta(\varepsilon) > 0$ so that for all x, for which $\|x\|_X < \delta(\varepsilon)$ holds, $\|\hat{A}x\|_Y < \varepsilon$, hence $\|\hat{A}\| \leq \varepsilon/\delta(\varepsilon) < \infty$. Therefore bounded linear operator and continuous linear operator from a normed linear space into a normed linear space mean the same thing. Linear combinations of bounded linear operators are again bounded linear operators, and it is not hard to see that (5.35) has all properties of a norm. Thus, all bounded linear operators from a normed linear space X into a normed linear space Y form again a normed linear space, which is denoted by $\mathcal{L}(X, Y)$, and *which is a Banach space if Y is a Banach space.* (Note that it was not presupposed that X is a Banach space, the comparatively simple proof of this statement can again be found in textbooks or treated as an exercise.)

5.5 Dual Space

A bounded linear functional is a bounded linear operator from a normed linear space X to the real line or complex plane, respectively, depending on whether X is real or complex. We will only consider the real case, the complex case is a straightforward generalization. The (topological) dual to a normed space X is the Banach space

$$X^* \stackrel{\text{def}}{=} \mathcal{L}(X, \boldsymbol{R}). \tag{5.36}$$

(It is *always* a Banach space because $\boldsymbol{R}$ is a Banach space.)

Let $1 \le p \le \infty$ and $1/p + 1/q = 1$. Let $g(\boldsymbol{x}) \in \boldsymbol{L}^q$. From Hölder's inequality (5.26) it is clear that

$$\hat{g}f = (g|f) \stackrel{\text{def}}{=} \int d^N x\, g(\boldsymbol{x}) f(\boldsymbol{x}) \text{ for every } f \in \boldsymbol{L}^p \tag{5.37}$$

defines a bounded linear functional $\hat{g}$ on $\boldsymbol{L}^p$ whose norm $\|\hat{g}\| \le \|g\|_q$ since $|\hat{g}f|/\|f\|_p \le \|g\|_q$. Since $f(\boldsymbol{x}) = \text{sign} g(\boldsymbol{x})\ |g(\boldsymbol{x})|^{q-1} \in \boldsymbol{L}^p$ with $\|g^{q-1}\|_p = \|g\|_q^{q/p}$, equality holds for this f in Hölder's inequality, hence $\|\hat{g}\| = \|g\|_q$.

It is a rather tedious work to show that for $1 \le p < \infty$ (but not for $p = \infty$) every bounded linear functional $\hat{g}$ may be defined by some $g \in \boldsymbol{L}^q$, hence, by identifying $\hat{g}$ with $g(\boldsymbol{x})$ and $\|\hat{g}\|$ with $\|g\|_q$, we have

$$\boldsymbol{L}^{p*} = \boldsymbol{L}^q, \quad 1 \le p < \infty, \quad 1/p + 1/q = 1, \quad \boldsymbol{L}^{\infty *} \supset \boldsymbol{L}^1. \tag{5.38}$$

(For $1 < q \le \infty$ this implies by the way again that $\boldsymbol{L}^q$ is complete.) From (5.38) it follows immediately that

$$\boldsymbol{L}^{p**} = \boldsymbol{L}^p, \quad 1 < p < \infty, \quad \boldsymbol{L}^{1**} \supset \boldsymbol{L}^1, \quad \boldsymbol{L}^{\infty **} \supset \boldsymbol{L}^\infty. \tag{5.39}$$

Moreover, for $1 \le p, p' < \infty$, $(\boldsymbol{L}^p \cap \boldsymbol{L}^{p'})^* = \boldsymbol{L}^q + \boldsymbol{L}^{q'}$.

A Banach space X is said to be reflexive, if $X^{**} = X$. (By considering the bilinearity of $(g|f)$, allowing for an interpretation as a linear functional $\hat{g}$ on X or as a linear functional $\hat{f}$ on X^*, $X^{**} \supseteq X$ is true for every normed linear space, where X^{**} is a Banach space even if X is not; hence, to form the second dual is a way to complete X.)

From the definition of a seminorm by the first two properties (5.16) it is seen that any bounded linear functional $\hat{y}$ defines a seminorm on X by $p_y(x) = |\hat{y}x| = |(y|x)|$. If y runs over all elements of the dual space X^*, then

(5.17) is fulfilled, hence the dual can be used in a given Banach space X to define a weak topology with the base

$$\begin{aligned} B_{n,(y_k),z,\varepsilon} &= \{\, x \,|\, |(y_k|x-z)| < \varepsilon,\ k = 1,\dots,n \,\} \\ &\qquad \text{for every } n,\ y_k \in X^*,\ z \in X,\ \varepsilon > 0. \end{aligned} \tag{5.40}$$

Except for finite-dimensional spaces, this weak topology is strictly weaker than the norm topology. X equipped with this topology is in general merely a locally convex space rather than a Banach space. (The Banach property always refers to a corresponding norm topology.)

Geometrically, a point $y \in X^*$ defines via $(y|x) = 0$ a norm-closed hyperplane in X and via $d_y(x) = (y|x)/\|y\|$ a signed distance of x from that hyperplane. While norm convergency of a series (x_i) means uniform convergency of all those distances from all possible hyperplanes, weak convergency only means independent convergency of those distances. (Cf. the examples of Section 5.2.)

Analogously, the points of X^{**} can be used to introduce the weak topology in X^*. If, however, X is not reflexive, the points of $X \subset X^{**}$ define even a weaker topology in X^*, called the weak* topology, and for this the Banach-Alaoglu theorem says: *The closed unit ball of X^* is weak* compact.* This implies that for a reflexive Banach space (for which the weak and weak* topologies coincide) the closed unit ball is weak compact, or, in other terms, *every bounded net in a reflexive Banach space has a weak converging subnet.*

The main theorem for extremal problems states that *for a real functional F on a topological space X the minimum problem*

$$\min_{x \in A \subseteq X} F[x] = \alpha \tag{5.41}$$

has a solution $x_0 \in A$, $F[x_0] = \alpha$, if F is lower semicontinuous and A is nonempty and compact. We have learned that different topologies may be considered in X. The weaker the topology, the stronger the semicontinuity condition, but the weaker the compactness condition. If X is a reflexive Banach space, these conditions are often realized in the weak topology (without reflexivity in the weak* topology). The proof of the theorem is simple: consider the infimum of the functional on A, pick a corresponding net, and use compactness to select a convergent subnet.

Just for comparison we come back to the trivial case $X = \boldsymbol{R}$ at the end of this section. A linear functional in this case is a linear function $y = cx$ for all $x \in \boldsymbol{R}$ with some $c \in \boldsymbol{R}$. With the ordinary norm $\|x\| = |x|$ in $\boldsymbol{R}$, the norm (5.35) of y is $\|y\| = |c|$. Since $|c| < \infty$, the linear function is always bounded

(in the sense of a linear operator, i.e. of bounded norm), and it is of course continuous. (Non-continuous linear functions have to do with an infinite number of dimensions of space: the slope of the function may unboundedly increase, if one runs through the dimensions.) Hence, $\boldsymbol{R}^* = \boldsymbol{R} = \boldsymbol{R}^{**}$. Weak and norm topologies coincide in $\boldsymbol{R}$. (The same holds true for $\boldsymbol{R}^N$.)

5.6 Conjugate Functionals

In this section we consider a dual pair (X, X^*) of reflexive *real* Banach spaces, $x \in X,\ y \in X^*$. A functional on X is a function $F : X \to \boldsymbol{R}$; in analogy to Section 5.1 we slightly generalize this notion by allowing $F : X \to \boldsymbol{R} \cup \{+\infty\}$. The conjugate functional (or Legendre transform) F^* to F is defined as

$$F^*[y] \stackrel{\text{def}}{=} \sup_{x \in X} \{(y|x) - F[x]\} \text{ for all } y \in X^*. \tag{5.42}$$

Now we are ready to transfer the content of Section 5.1 to this case.

A functional F is convex, if

$$F[cx_1 + (1-c)x_2] \le cF[x_1] + (1-c)F[x_2]$$
$$\text{for every } c \in [0,1] \text{ and every } x_1,\ x_2 \in X. \tag{5.43}$$

Lower semicontinuity is also defined exactly as in Section 5.1, this time using the norm topology for convergence of sequences of points of X. Weak lower semicontinuity is defined using the weak topology in X and eventually replacing sequences by nets. In general there are less weak lower semicontinuous functionals than (norm) lower semicontinuous functionals: the latter comprise the former but not vice versa. However, every *convex* norm lower semicontinuous functional on a Banach space is weak lower semicontinuous. A subgradient of F at x is a bounded linear functional y so that

$$F[x'] \ge F[x] + (y|x' - x) \text{ for all } x' \in X \tag{5.44}$$

(cf. (5.9)). Geometrically, the graph (in $X \times \boldsymbol{R}$ with points $(x', F[x'])$) of the affine-linear functional $H_{x,y}[x'] = F[x] + (y|x' - x)$; x, y fixed in accordance with (5.44), is a hyperplane of support for F at x, a hyperplane touching the graph of $F[x']$ at $(x, F[x])$ and being nowhere above this graph. The subdifferential $\partial F[x]$ is defined again to be the set of all subgradients y of F at x.

Now, for every $F : X \to \boldsymbol{R} \cup \{+\infty\}$,

$$F^*[y] \in \boldsymbol{R} \cup \{+\infty\}, \qquad F^* \not\equiv +\infty,$$

$$F^* \text{ is convex and lower semicontinuous,} \tag{5.45}$$

$$F^*[y] + F[x] \geq (y|x), \tag{5.46}$$

$$F^*[y] + F[x] = (y|x) \iff x \in \partial F^*[y]. \tag{5.47}$$

$$F_1[x] \leq F_2[x] \text{ for all } x \Longrightarrow F_1^*[y] \geq F_2^*[y] \text{ for all } y. \tag{5.48}$$

The relevant issue of the Hahn-Banach theorem is:

> *For every point (in $X \times \boldsymbol{R}$) below the graph of a lower semicontinuous convex functional there is a norm-closed hyperplane separating the former from the latter, i.e. for which the graph of the convex functional is entirely above the hyperplane, and the point is below.*

(The proof consists of transfinite induction through linear independent $y_\gamma \in X^*$, of the step considered in Section 5.1.) From this issue, by the same reasoning as in Section 5.1, $F^{**} = (F^*)^*$ is the lower semicontinuous convex hull of F, and

$$F^{**} \equiv F \iff F \text{ is lower semicontinuous and convex.} \tag{5.49}$$

In this case, F and F^* form a dual pair of lower semicontinuous convex mutually conjugate functionals or of Legendre transforms.

5.7 The Functional Derivative

There are generalizations of partial derivatives and of a gradient to the case of a functional F on a normed linear space X. If there exists a bounded linear functional $F'[x_0] \in X^*$, $x_0 \in X$, so that

$$\lim_{\alpha \to 0} \frac{F[x_0 + \alpha x] - F[x_0]}{\alpha} = (F'[x_0]|x) \text{ for all } x \in X \tag{5.50}$$

with scalar (real or complex) α, then $F'[x_0]$ is called the *G-derivative* (or Gâteaux derivative, directional derivative) of F at x_0. For every fixed nonzero $x \in X$ the number $(F'[x_0]|x)/\|x\|$ generalizes a partial derivative at x_0 in the direction of x. (As a further generalization, the G-derivative is by the same expression (5.50) even defined for functionals on a locally convex space X without a norm.)

If F is a real functional on the real $\boldsymbol{L}^p(\boldsymbol{R}^N) = \boldsymbol{L}^p$ and $f_0 \in \boldsymbol{L}^p$, then $F'[f_0]$, if it exists, is an element of $\boldsymbol{L}^q$, i.e. $F'[f_0] = g \in \boldsymbol{L}^q$ and, for every

$f \in \boldsymbol{L}^p$, $(F'[f_0]|f) = \int d^N x\, g(\boldsymbol{x}) f(\boldsymbol{x})$. (Note that, in this particular case, our notation replaces $x \in X$ and $y \in X^*$ of our general considerations by $f \in \boldsymbol{L}^p$ and $g \in \boldsymbol{L}^q$, respectively, while $\boldsymbol{x}$ means a point of $\boldsymbol{R}^N$.) This situation is often expressed by writing

$$\frac{\delta F[f]}{\delta f(\boldsymbol{x})} = g(\boldsymbol{x}). \tag{5.51}$$

This writing implies without mention that both sides are functionals of $f_0 \in \boldsymbol{L}^p$, the point where the derivative is taken.

If X is a normed linear space and a bounded linear functional $F'[x_0] \in X^*$, $x_0 \in X$ exists so that

$$F[x_0 + x] - F[x_0] = (F'[x_0]|x) + o(\|x\|) \text{ as } x \to 0 \tag{5.52}$$

for all $x \in X$, then $F'[x_0]$ is called the *F-derivative* (or Fréchet derivative, gradient) of F at x_0. (The notation $o(\|x\|)$ means $\lim_{\|x\|\to 0} o(\|x\|)/\|x\| = 0$.) $F'[x_0]$ generalizes the gradient of F at x_0.

Note that while the limes in (5.50) was taken for a fixed x setting a direction in the functional space, (5.52) is demanded *uniformly* for all x. F-derivative and G-derivative coincide at x_0 in a *normed* space, if and only if F' as a function on X with values in X^* is (norm) continuous at x_0. This is exactly the same situation as is already met for functions of a finite number of variables, as for instance functions on $\boldsymbol{R}^2$. For example,

$$F(u,v) = \begin{cases} u^3 v/(u^4 + v^2) & \text{if } (u,v) \neq (0,0), \\ 0 & \text{if } (u,v) = (0,0) \end{cases} \tag{5.53}$$

has at $(u,v) = (0,0)$ both partial derivatives and every directional derivative in the u,v-plane equal to zero, but not continuous: $\partial F/\partial v$ is singular at $(u,v) = (0,0)$ as a function of u. Nevertheless F has a nonzero slope on the curve $v = u^2$, hence there is no gradient at $(0,0)$ [Kolmogorov and Fomin, 1970, chap. X, §1].

A natural generalization replaces $F : X \to \boldsymbol{R}$ by a mapping $F : X \to Y$ from a normed linear space X into a normed linear space Y and the bounded linear functional $F'[x_0]$ by a bounded linear operator $\widehat{F'[x_0]} : X \to Y$. The operator $\widehat{F'[x_0]} \in \mathcal{L}(X,Y)$ is a G-derivative, if

$$\lim_{\alpha\to 0} \frac{F[x_0 + \alpha x] - F[x_0]}{\alpha} = \widehat{F'[x_0]}x \text{ for all } x \in X. \tag{5.54}$$

The limes is understood here in the norm topology of the space Y. It is an F-derivative, if

$$F[x_0 + x] - F[x_0] = \widehat{F'[x_0]}x + o(\|x\|) \text{ as } x \to 0 \tag{5.55}$$

for all $x \in X$.

This generalization allows for the formulation of a general chain rule: Let $G : X \to Y$ and $H : Y \to Z$. Let $F = H \circ G$ be the composite map $F : X \to Z : F[x] = H[G[x]]$. Then

$$\widehat{F'[x_0]} = \widehat{H'[G[x_0]]}\widehat{G'[x_0]} \tag{5.56}$$

is an F-derivative, if the right side expressions exist as F-derivatives. In this equality, $\widehat{F'[x_0]}$ is a linear operator from X into Z, $\widehat{G'[x_0]}$ is a linear operator from X into Y, and $\widehat{H'[G[x_0]]}$ is a linear operator from Y into Z. The last line of (4.31) is to be understood in this sense for a situation, where the linear operators of (5.56) are integral operators.

For a map $F : X \to Y$, the derivative is a map $F' : X \to \mathcal{L}(X, Y)$, i.e., for every $x_0 \in X$, $F'[x_0] \in \mathcal{L}(X, Y)$. Taking again the derivative of F', a map $F'' : X \to \mathcal{L}(X, \mathcal{L}(X, Y))$ is obtained. (Compare the tensor structure of higher derivatives in $\boldsymbol{R}^N$.) For $x_0 \in X$, the image of the map F'' is a bounded linear operator $\widehat{F''[x_0]}$ on X, so that for every $x_1 \in X$ the value $\widehat{F''[x_0]}x_1$ is in $\mathcal{L}(X, Y)$, i.e., it is again a bounded linear operator on X: For every $x_1, x_2 \in X$, the value $\widehat{F''[x_0]}x_1x_2 \in Y$ depends linearly on both x_1 and x_2. For a *functional* $F[x]$, $F''[x_0]$ is a bilinear functional, and in the scalar product notation of linear functionals we write this as $((F''[x_0]|x_1)|x_2) \in \boldsymbol{R}$, in the dual case as $(y_2|(y_1|F^{*''}[y_0]))$, $y_i \in X^*$. Continuing this process, e.g. for a functional $F : X \to \boldsymbol{R}$ and its F-derivatives, the Taylor expansion

$$F[x_0 + x] = F[x_0] + \sum_{k=1}^{n} \frac{1}{k!}(\cdots((F^{(k)}[x_0]|x)|x)\cdots|x) + o(\|x\|^n) \tag{5.57}$$

holds, if the F-derivatives exist. (For $X = \boldsymbol{L}^p$, the k-th F-derivative is an element of $\boldsymbol{L}^q \times \boldsymbol{L}^q \times \cdots \times \boldsymbol{L}^q$ (k factors). On the basis of Fubini's theorem on multiple integrals, this is given by a (symmetric) function $g(\boldsymbol{x}_1 \ldots \boldsymbol{x}_k)$ so that

$$(\cdots(F^{(k)}[f_0]|f)\cdots|f) = \int d^N x_1 \cdots d^N x_k \, g(\boldsymbol{x}_1 \ldots \boldsymbol{x}_k) f(\boldsymbol{x}_1) \cdots f(\boldsymbol{x}_k),$$

where g is a function of f_0.)

As an application let $F^*[y_0]$ be strictly convex, twice differentiable at $y_0 \in X^*$, and $(y_2|(y_1|F^{*\prime\prime}[y_0])) \neq 0$ if y_1, $y_2 \neq 0$. Then, $\partial F^*[y_0] = \{F^{*\prime}[y_0]\}$, and (5.47) reads $F^*[y_0] + F[F^{*\prime}[y_0]] = (y_0|F^{*\prime}[y_0])$. Note that $F^{*\prime}[y_0]$ is a linear functional on X^* and hence an element of X. A further differentiation using the chain rule yields $F^{*\prime}[y_0] + (F'[F^{*\prime}[y_0]]|F^{*\prime\prime}[y_0]) = F^{*\prime}[y_0] + (y_0|F^{*\prime\prime}[y_0])$, hence $(F'[F^{*\prime}[y_0]] - y_0|F^{*\prime\prime}[y_0]) = 0$. Because $F^{*\prime}[y_0]$ is an element of X, $F'[F^{*\prime}[y_0]]$ is defined and is a linear functional on X and hence an element of X^*. With our assumption on $F^{*\prime\prime}$ it follows that

$$F'[F^{*\prime}[y_0]] = y_0. \tag{5.58}$$

This result generalizes (5.14).

5.8 Lagrange Multipliers

Again we do not consider the most general case.

Let $F : X \to \boldsymbol{R} \cup \{+\infty\}$, and let $\hat{G} : X \to Y$ be a continuous linear operator from the Banach space X *onto* the Banach space Y (that is, $\hat{G}$ is supposed to be surjective). Then, $\hat{G}x = y_0$ defines a closed hypersurface in X (with dimension $\dim X - \dim Y$; $Y = \boldsymbol{R}$ would be a simple case). Consider the minimum problem

$$\min\{\, F[x] \,|\, x \in X,\ \hat{G}x = y_0 \,\} = \alpha. \tag{5.59}$$

In this case one says that a bound minimum of $F[x]$ is searched for, where x is bound to the hypersurface defined by the side condition $\hat{G}x = y_0$. (This is in contrast to a free minimum on all X.)

Let $F[x_0] = \alpha$, $\hat{G}x_0 = y_0$, and suppose that $F'[x_0]$ exists as an F-derivative. If $\hat{G}x_1 = 0$, then $\hat{G}(x_0 + x_1) = y_0$ and hence x_1 is an admissible variation by the side condition. Therefore the minimum condition implies $(F'[x_0]|x_1) = 0$. If on the other hand $\hat{G}x_2 \neq 0$, then $(F'[x_0]|x_2)$ is not restricted by the minimum condition because x_2 is not admissible for the minimum search. Hence, the null space of $F'[x_0]$ (i.e. the subspace of X which is nullified by $F'[x_0]$) contains the null space of $\hat{G}$, and consequently there exists a linear functional $\Lambda : Y \to \boldsymbol{R}$, so that $(\Lambda|\hat{G}|x) = (F'[x_0]|x)$ for all $x \in X$. (If there would exist an $x_1 \in X$ with $(F'[x_0]|x_1) \neq 0$, but $\hat{G}x_1 = 0$, then the above Λ could not exist.) Thus we arrive at

$$F'[x_0] - \Lambda \circ \hat{G} = 0 \tag{5.60}$$

as a necessary condition for (5.59). Every solution of (5.59) is thus a solution of

$$F[x] - (\Lambda|\hat{G}|x) \Rightarrow \textit{free} \text{ stationary for all } x \in X. \tag{5.61}$$

Λ of course depends on y_0 and on F and is called a Lagrange multiplier. The commonly known case is $X = \boldsymbol{L}^p$, $Y = \boldsymbol{R}$, $\hat{G} \in \boldsymbol{L}^q$ and $\hat{G}f = \int d^N x\, g(\boldsymbol{x}) f(\boldsymbol{x})$. In this case, $\Lambda : \boldsymbol{R} \to \boldsymbol{R} : r \mapsto \lambda r$ and $(\Lambda|\hat{G}|f) = \lambda \int d^N x\, g(\boldsymbol{x}) f(\boldsymbol{x})$, and the Lagrange multiplier is a number λ.

As a generalization, now let $F : X \to \boldsymbol{R} \cup \{+\infty\}$ be arbitrary, $H : Y \to \boldsymbol{R} \cup \{+\infty\}$ convex and lower semicontinuous, $\hat{G} : X \to Y$ a continuous linear operator, and $x \in X$, $y \in Y^*$. Consider

$$L[x, y] = F[x] + (y|\hat{G}|x) - H^*[y]. \tag{5.62}$$

From the assumptions, the doubly conjugate $H^{**} = H$, $\sup_{y \in Y^*}\{(y|\hat{G}|x) - H^*[y]\} = H^{**}[\hat{G}x] = H[\hat{G}x]$, and hence the inf-sup problem

$$\inf_{x \in X} \sup_{y \in Y^*} L[x, y] = \alpha \tag{5.63}$$

is equivalent to the inf problem

$$\inf_{x \in X} \{F[x] + H[\hat{G}x]\} = \alpha. \tag{5.64}$$

Obviously, $\inf_x L[x, y_0] \le L[x_0, y_0] \le \sup_y L[x_0, y]$ for all x_0, y_0, hence $\sup_y \inf_x L[x, y] \le \inf_x \sup_y L[x, y]$. Generally, from the properties of infima and suprema, (5.63) implies

$$\sup_{y \in Y^*} \inf_{x \in X} L[x, y] = \beta \le \alpha. \tag{5.65}$$

With $(y|\hat{G}|x) = (\hat{G}^\dagger y|x)$ where $\hat{G}^\dagger$ is the operator adjoint to $\hat{G}$, it follows that $\inf_{x \in X}\{F[x] + (\hat{G}^\dagger y|x)\} = -\sup_{x \in X}\{(-\hat{G}^\dagger y|x) - F[x]\} = -F^*[-\hat{G}^\dagger y]$. The sup-inf problem (5.65) is thus equivalent to $\sup_{y \in Y^*}\{-F^*[-\hat{G}^\dagger y] - H^*[y]\} = \beta$ and hence to

$$\inf_{y \in Y^*} \{F^*[-\hat{G}^\dagger y] + H^*[y]\} = -\beta. \tag{5.66}$$

Finally, (5.65) may be written as

$$\inf_{y \in Y^*} \sup_{x \in X} -L[x, y] = -\beta. \tag{5.67}$$

$L[x, y]$ is called a general *Lagrange function*, which mediates a duality between (5.64) and (5.66), called *Fenchel's duality*. Comparing (5.62) with (5.61) one says that in the problem (5.63) equivalent to (5.64) y plays the rôle of an abstract Lagrange multiplier, and that in the problem (5.67) equivalent to (5.66) x plays such a rôle.

The introduction of the Lagrange function (5.62) and of the corresponding inf-sup problems leads naturally to the notion of a *saddle point* of a functional. Let $L[x, y] : A \times B \to \boldsymbol{R}$, and

$$\max_{y \in B} L[x_0, y] = L[x_0, y_0] = \min_{x \in A} L[x, y_0], \quad (x_0, y_0) \in A \times B. \tag{5.68}$$

Then (x_0, y_0) is a saddle point of L with respect to $A \times B$. In this case, $\alpha = \beta$ in the above relations, and the suprema and infima are maxima and minima, respectively. Rather general sufficient conditions for the existence of a saddle point are:

1. X and Y are reflexive Banach spaces, $A \subseteq X$, $B \subseteq Y$, A, B are convex, closed, and non-empty.

2. $L[x, y]$ is convex and lower semicontinuous on A for all fixed $y \in B$, and $-L[x, y]$ is convex and lower semicontinuous on B for all fixed $x \in A$.

3. A is bounded in X, or there exists $y' \in B$ so that $L[x, y'] \to +\infty$ as $\|x\| \to +\infty$. B is bounded in Y, or there exists $x' \in A$ so that $-L[x', y] \to +\infty$ as $\|y\| \to +\infty$.

For more details see e.g. [Zeidler, 1986, volume III].

6 Density Functional Theory by Lieb

The central idea of this chapter is to treat the Hohenberg-Kohn functional $F[n]$ as the Legendre transform to $E[v]$ (in a certain strict sense presented in the subsequent text).

One complication of this approach lies in the fact that the functional spaces of admitted densities n and admitted potentials v, which naturally appear from physical reasoning, are not reflexive. This comes about from the infinity of the physical space $\boldsymbol{R}^3$: in approaching the infimum of energy of an N-particle problem, some of the particles may disappear at infinity. For practically all problems of physical relevance, this difficulty may be separated by first considering a *finite* space (preferably with periodic boundary conditions as introduced in Section 2.7). For getting the required answers, it suffices in most cases to consider a large enough finite periodic volume. Only in special problems the limit with the volume tending to infinity must really be considered afterwards.

The first section of this chapter reformulates the problem of the ground state energy in this specified context and establishes the important convexity properties of the ground state energy. Since dependences on particle numbers are a most important issue of density functional theory, the particle number is treated as a continuous variable from the very beginning. The second section then develops the density functional theory straightforwardly with the use of the machinery of Legendre transforms. A universal density functional $H[n]$, which is most closely related to both Lieb's density functional and approximate functionals in practical use, is obtained for the Hohenberg-Kohn variational principle.

The more technical details and mathematical subtleties are put into the third section which may be skipped in a first reading. One artificial feature of Lieb's functional F is its separate and independent dependence on the total particle number N and on the density n. The gauge invariance of Schrödinger's equation with respect to an additive potential constant puts F to $+\infty$, if the integral over the density n is not equal to N. Our exploiting of Legendre transforms for both the N- and v-dependences of the ground state energy and the above mentioned gauge invariance avoids this feature

from the very beginning and condenses the separate dependences of F on n and N into a single dependence on n of a modified functional $H[n]$ which is the convex hull of all $F[n, N]$. This brings Lieb's theory back to the frame understood in approximative implementations for numerical calculations (as e.g. the local-density approximation) where one uses functionals depending on N only via n.

The last section of this chapter gives a final answer to the question of existence of functional derivatives of density functionals and hence puts the Kohn-Sham equations on a rigorous basis. Finally it summarizes the level the Hohenberg-Kohn-Sham theory has reached as a reliable closed mathematical frame.

In this chapter throughout the general case of spin density functional theory is treated with a possibly non-unidirectional magnetic field and a possibly non-collinear spin polarization.

6.1 The Ground State Energy

In Chapter 4 we dealt with three different definitions of the ground state energy of increasing generality: (4.4), (4.49), and (4.66). In Section 4.5 we saw that the consideration of ensemble states solves problems with level crossings and answers in general the question of occupying Kohn-Sham orbitals with the aufbau principle. In Section 4.6, mixed states with distinct particle numbers led to the convexity of $E[v, N]$ in the real variable N, a very important issue for the following (and for any practical approach to density functional theory as will be discussed later). There are many more reasons to consider most general mixed states as candidates for ground states, some of which were discussed in Section 4.5.

Hence, we leave the particle number N undefined so far and start with the family of Hamiltonians of type (4.1) (with the general potential form (4.77)),

$$\hat{H}[v, M] = -\frac{1}{2}\sum_{i=1}^{M} \nabla_i^2 + \sum_{i=1}^{M} v_{s_i s_i'}(\boldsymbol{r}_i) + \frac{1}{2}\sum_{i \neq j}^{M} w(|\boldsymbol{r}_i - \boldsymbol{r}_j|) \tag{6.1}$$

for all (admissible) external potentials v and all integer particle numbers M. In Section 4.1 we also discussed that given a potential v, not for every particle number M there exists a ground state. This problem is connected with the infinite measure (volume) of the position space $\boldsymbol{R}^3$ which allows unbound particles to disappear at infinity and causes scattering states not

to be normalizable but which has no practical relevance in our context. Therefore we replace the position space $\boldsymbol{R}^3$ by a torus $\boldsymbol{T}^3$ (box with periodic boundary conditions like in (2.82)) with volume $|\boldsymbol{T}^3|$ sufficiently large not significantly to change the considered results. This makes the spectrum of all Hamiltonians discrete and ground states existing and normalizable in any case. In practical calculations for finite systems (atoms, molecules, clusters) the boundary conditions may simply be ignored as long as all particles are bound; in case of extended systems they introduce a discrete mesh of $\boldsymbol{k}$-vectors like in (1.29) and the results have to be converged with the density of $\boldsymbol{k}$-points. The very helpful formal implications on the structure of the theory will be discussed in Section 6.3.

We admit ensemble states of the most general type (4.64):

$$\hat{\gamma} = \sum_K |\Psi_K\rangle g_K \langle\Psi_K|, \quad 0 \le g_K, \quad \sum_K g_K = 1, \tag{6.2}$$

where the pure states $|\Psi_K\rangle$ may be expanded into a fixed orthonormal set of particle number eigenstates:

$$|\Psi_K\rangle = \sum_M |\Phi_K^M\rangle C_M^K, \quad \sum_M |C_M^K|^2 = 1, \quad \hat{N}|\Phi_K^M\rangle = |\Phi_K^M\rangle M. \tag{6.3}$$

The energy expectation value in the general ensemble state (6.2) is

$$\mathrm{tr}\,(\hat{H}\hat{\gamma}) = \sum_{K,M} g_K |C_M^K|^2 \langle\Phi_K^M|\hat{H}[v,M]|\Phi_K^M\rangle, \tag{6.4}$$

where

$$0 \le p_M^K \stackrel{\text{def}}{=} g_K |C_M^K|^2, \quad \sum_{K,M} p_M^K = 1. \tag{6.5}$$

The expectation values for the particle number and particle density are

$$\mathrm{tr}\,(\hat{N}\hat{\gamma}) = \sum_{K,M} p_M^K M, \tag{6.6}$$

and

$$\mathrm{tr}\,(\hat{n}\hat{\gamma}) = \sum_{K,M} p_M^K n_{K,ss'}^M(\boldsymbol{r}) = n_{ss'}(\boldsymbol{r}), \tag{6.7}$$

where $n^M_{K,ss'}(\boldsymbol{r})$ is the spin density matrix of the state $|\Phi^M_K\rangle$. The interaction energy with the external potential

$$\mathrm{tr}\,(\hat{U}\hat{\gamma}) = \sum_{ss'} \int d^3r\, v_{ss'}(\boldsymbol{r}) n_{s's}(\boldsymbol{r}) \stackrel{\mathrm{def}}{=} (v|n). \tag{6.8}$$

is as previously a linear functional of the density n.

Now, the ground state energy as a functional of the external potential v and a function of the real particle number N can be defined as

$$\begin{aligned} E[v,N] &\stackrel{\mathrm{def}}{=} \inf_{\hat{\gamma}} \left\{ \mathrm{tr}\,(\hat{H}\hat{\gamma}) \;\middle|\; \mathrm{tr}\,(\hat{N}\hat{\gamma}) = N \right\} = \\ &= \inf_{\{p^K_M\}} \left\{ \sum_{K,M} p^K_M \langle \Phi^M_K | \hat{H}[v,M] | \Phi^M_K \rangle \;\middle|\; \sum_{K,M} p^K_M M = N \right\}. \end{aligned} \tag{6.9}$$

The M-sums run over all $M = 0, 1, 2, \ldots$ and for each M the K-sums run over arbitrarily many orthonormal M-particle states. (For the sake of completeness we include one abstract state $|\Phi^0\rangle$ (vacuum) with $\langle \Phi^0 | \hat{H}[v,0] | \Phi^0 \rangle = 0$ for all v.)

Caution is needed here since given v and N, (6.9) need not exist even if $\langle \Phi^M_K | \hat{H}[v,M] | \Phi^M_K \rangle$ is bounded below for every fixed M. A simple counterexample is $\langle \Phi^M_0 | \hat{H}[v,M] | \Phi^M_0 \rangle = -\varepsilon M^2$ and $p^0_0 = 1 - N/M_1$, $p^0_{M_1} = N/M_1$. $p^K_M = 0$ else. Clearly the sum in (6.9) diverges for $M_1 \to \infty$. This cannot happen for $w \equiv 0$ since in this case $\langle \Phi^M_K | \hat{H}[v,M] | \Phi^M_K \rangle \geq -\varepsilon_1 M$ where ε_1 is the lowest orbital energy. (Cf. Fig.5a. Recall that $\langle \Phi^M_0 | \hat{H}[v,M] | \Phi^M_0 \rangle$ is convex as a function of M in this case.)

In the whole of this chapter the particle-particle interaction is presupposed repulsive, $w \geq 0$. Then, with v fixed (6.9) is bounded below by the convex function $E^0[v,N]$ and the infimum exists.

Now, fix v and pick two particle numbers N_1 and N_2 (not necessarily integers). There exist sequences $\hat{\gamma}^i_j$, $i = 1,2$; $j = 1,2,\ldots$ with $\mathrm{tr}\,(\hat{N}\hat{\gamma}^i_j) = N_i$ and $\lim_j \mathrm{tr}\,(\hat{H}\hat{\gamma}^i_j) = E[v,N_i]$. Take the sequence $\hat{\gamma}_j = c\hat{\gamma}^1_j + (1-c)\hat{\gamma}^2_j$, $0 \leq c \leq 1$. Obviously $\mathrm{tr}\,(\hat{N}\hat{\gamma}_j) = cN_1 + (1-c)N_2$, $\lim_j \mathrm{tr}\,(\hat{H}\hat{\gamma}_j) = cE[v,N_1] + (1-c)E[v,N_2]$. On the other hand, by definition $E[v, cN_1 + (1-c)N_2] \leq \lim_j \mathrm{tr}\,(\hat{H}\hat{\gamma}_j)$. This proves the convexity of $E[v,N]$ in N for fixed v:

$$E[v, cN_1 + (1-c)N_2] \leq cE[v,N_1] + (1-c)E[v,N_2],\ 0 \leq c \leq 1. \tag{6.10}$$

(By definition (6.9) and the text above, $N \geq 0$ with $E[v,0] = 0$ for all v. One may, however, formally define $E[v,N] = +\infty$ for all $N < 0$ and all v,

which makes $E[v,N]$ defined for all N, convex and lower semicontinuous.) As discussed in Section 4.6, for integer N the energy (6.9) may be below that of (4.49) for the search restricted to particle number eigenstates with $M = N$ (cf. Fig.5b). However, later it will become clear that if one uses one universal density functional $H[n]$ for all densities n integrating up to arbitrary particle numbers N (as one always does in practical implementations) one produces (6.9) instead of (4.49).

Next, we fix N and find

$$\begin{aligned} E[cv_1 + (1-c)v_2, N] \; &= \\ &= \inf_{\hat\gamma} \left\{ c\,\mathrm{tr}\,(\hat H_{v_1}\hat\gamma) + (1-c)\,\mathrm{tr}\,(\hat H_{v_2}\hat\gamma) \,\middle|\, \mathrm{tr}\,(\hat N\hat\gamma) = N \right\} \geq \\ &\geq c \inf_{\hat\gamma} \left\{ \mathrm{tr}\,(\hat H_{v_1}\hat\gamma) \,\middle|\, \mathrm{tr}\,(\hat N\hat\gamma) = N \right\} + \\ &\quad + (1-c) \inf_{\hat\gamma} \left\{ \mathrm{tr}\,(\hat H_{v_2}\hat\gamma) \,\middle|\, \mathrm{tr}\,(\hat N\hat\gamma) = N \right\} = \\ &= cE[v_1,N] + (1-c)E[v_2,N], \quad 0 \leq c \leq 1. \end{aligned} \tag{6.11}$$

In the first equality (4.2) was used, and then the simple fact that the infimum of a sum cannot be lower than the sum of the corresponding independent infima.

In summary:

> *The ground state energy $E[v,N]$ is a convex function of N for fixed v and a concave functional of v for fixed N.*

These simple convexity properties of $E[v,N]$ together with the gauge property (4.6) form the deep logical foundation of density functional theory.

6.2 The Hohenberg-Kohn Variational Principle

Starting with the convexity of $E[v,N]$ in N a Legendre transform $\tilde G[v,\mu]$ may be defined with the pair of transformations

$$\tilde G[v,\mu] = \sup_N \{\mu N - E[v,N]\} \tag{6.12}$$

$$E[v,N] = \sup_\mu \left\{ N\mu - \tilde G[v,\mu] \right\} \tag{6.13}$$

Because of the gauge property (4.6) the expression in braces of (6.12) is $-E[v-\mu,N]$, and hence

$$\tilde G[v,\mu] = \tilde G[v-\mu,0] \stackrel{\text{def}}{=} G[v-\mu] \tag{6.14}$$

has only one functional dependence. The above duality relations simplify to

$$G[v] = -\inf_N E[v,N], \quad E[v,N] = \sup_\mu \{N\mu - G[v-\mu]\}. \tag{6.15}$$

G is convex in v as shown by the following chain of relations ($c \in [0,1]$):

$$\begin{aligned} G[cv_1 + (1-c)v_2] &= -\inf_N E[cv_1 + (1-c)v_2, N] \le \\ &\le -\inf_N \{cE[v_1,N] + (1-c)E[v_2,N]\} \le \\ &\le -c\inf_{N'} E[v_1,N'] - (1-c)\inf_{N''} E[v_2,N''] = \\ &= cG[v_1] + (1-c)G[v_2]. \end{aligned} \tag{6.16}$$

From the first to the second line the convexity of $-E[v,N]$ in v was used, and then again the negative of an infimum of a sum cannot exceed the negative of the sum of the independent infima of the items.

Because $G[v]$ is convex, it can be again back and forth Legendre transformed. This time it is a functional Legendre transformation for which functional spaces have to be specified. We postpone this specification and introduce formally $-n$ as a dual variable to v in order to arrive at common notations:

$$\tilde{H}[-n] = \sup_v \{(-n|v) - G[v]\} \stackrel{\text{def}}{=} H[n] \tag{6.17}$$

$$\begin{aligned} G[v] &= \sup_n \left\{(v|-n) - \tilde{H}[-n]\right\} = \\ &= -\inf_n \{H[n] + (v|n)\}. \end{aligned} \tag{6.18}$$

$(v|n)$ is a real scalar product since v and n are Hermitian spin matrices, hence $(-n|v) = (v|-n) = -(v|n)$.

Inserting the first relation (6.15) into (6.17) yields

$$\begin{aligned} H[n] &= \sup_v \left\{-(n|v) + \inf_N E[v,N]\right\} \le \\ &\le \inf_N \sup_v \{E[v,N] - (n|v)\} = \\ &= \inf_N F[n,N], \end{aligned} \tag{6.19}$$

where the general rule $\sup\inf \le \inf\sup$ (see Section 5.8) was applied, and

$$F[n,N] \stackrel{\text{def}}{=} \sup_v \{E[v,N] - (n|v)\} \tag{6.20}$$

was first introduced by Lieb [Lieb, 1983] as a density functional. (Lieb related it to $E_N[v]$ as defined by (4.4) with the obvious relation $E_N[v] \geq E[v,N]$. The corresponding Lieb functional $F_N[n] \stackrel{\text{def}}{=} \sup_v\{E_N[v] - (n|v)\}$ will shortly be considered below.)

Next we insert (6.18) into the right equation (6.15) and obtain

$$\begin{aligned} E[v,N] &= \sup_\mu \left\{ N\mu + \inf_n \{H[n] + (v-\mu|n)\} \right\} \leq \\ &\leq \inf_n \left\{ H[n] + (v|n) + \sup_\mu \left[N - (1|n)\right] \mu \right\}. \end{aligned} \tag{6.21}$$

Again the rule $\sup\inf \leq \inf\sup$ was applied, and, since μ can be treated as a potential function constant in space, $(\mu|n) = (1|n)\mu$. It will be shown in the next section that in both relations (6.19) and (6.21) the inequalities $\leq$ may be sharpened into equalities. Since the expression under the last supremum is linear in μ, the supremum is either $+\infty$, if $(1|n) \neq N$, or zero, if $(1|n) = N$. Taking then the infimum over all n means just selecting the last case, that is,

$$E[v,N] = \inf_n \{ H[n] + (v|n) \mid (1|n) = N \}. \tag{6.22}$$

Evidently, this is the variational principle by Hohenberg and Kohn, and $H[n]$ is the celebrated density functional in its logically most satisfying context.

As a Legendre transform, $H[n]$ is automatically convex and lower semicontinuous (in the norm and even in the weak topology of the functional space yet to be specified, cf. Section 5.6) and has a non-empty subdifferential $\partial H[n]$ at every n where $H[n]$ is finite. We shall see that this subdifferential is governed by the basic theorem by Hohenberg and Kohn relating v to n.

6.3 The Functionals F, G, and H

The functional dependence of $G[v]$ on v consists of its dependence on $v\mathrm{mod}(\mu)$ and of the dependence on the constant μ resembling a chemical potential. Given $v\mathrm{mod}(\mu)$, the dependence of G on μ can easily be analyzed. The situation is illustrated on Fig.11. On the left panel $E[v,N]$ is plotted against N for fixed v. It is piecewise linear between integer N, $E[v,0] = 0$, and $E[v,N] = +\infty$ for $N < 0$. From ($N = 0$, $E = 0$) it starts out with (negative) slope μ_0. For $0 \leq N \leq 1$, $\mu_0 N - E[v,N] = 0$, and $\mu_0 N - E[v,N] \leq 0$ elsewhere, hence $G[v-\mu_0] = \sup_N\{\mu_0 N - E[v,N]\} = 0$.

In general, for any slope μ, the ordinate section $-G$ of the tangent of support $l(N) = \mu N - G$ defines $G[v - \mu]$. (For any slope $\mu < \mu_0$ this tangent of support passes through the origin, whence $G[v - \mu] = 0$ for $\mu < \mu_0$.) On the right panel of Fig.11 $-G$ is plotted against μ for the sake of simpler comparison with the left panel. G itself is obviously convex (and piecewise linear), and it is not difficult to realize

$$G[v - \mu] = \mu N(\mu) - \sum_{N=0}^{N(\mu)-1} \mu_N, \quad \mu_{N(\mu)-1} \leq \mu \leq \mu_{N(\mu)} \tag{6.23}$$

with

$$\mu_N = E[v, N+1] - E[v, N] = -I_{N+1} = -A_N, \tag{6.24}$$

where I_N is the ionization potential of the N-particle state and A_N is its particle affinity.

Next consider the relation

$$H[n] = \sup_v \inf_N \{E[v, N] - (n|v)\} \leq \inf_N \sup_v \{E[v, N] - (n|v)\}$$

contained in (6.19). The question arises, given n, does $E[v, N] - (n|v)$ as a function of v and N have a saddle point? If yes, then the $\leq$ sign in the above relation can be sharpened into equality (cf. Section 5.8). The answer to the question is yes. Given v (and n), for every potential constant c, $E[v - c, N] = E[v, N] - cN$. Since $E[v, N]$ is convex in N, for $N_n = (n|1)$ one finds a c_n, so that $E[v - c_n, N_n] = E[v, N_n] - (n|1)c_n = \min_N \{E[v, N] - Nc_n\} = \min_N E[v - c_n, N]$. (This c_n just fixes the potential zero at the required value of the chemical potential for $N = N_n$: it is given by the slope of a tangent of support to $E[v, N]$ at $N = N_n$, cf. Fig.11.) Hence, for that v, $\inf_N \{E[v - c_n, N] - (n|v - c_n)\} = E[v - c_n, N_n] - (n|v - c_n) = E[v, N_n] - (n|v)$ and, since v was chosen arbitrarily at the beginning of this consideration, $H[n] \geq E[v, N_n] - (n|v)$ for *every* v, since H is the supremum over all v of the latter v-dependent infimum. The important point is that the only other v-dependent quantity c_n dropped out of the last estimate, and one has $H[n] \geq \sup_v \{E[v, N_n] - (n|v)\}$. Since $H[n]$ is not less than this supremum for $N = N_n$, it is *a fortiori* not less than the infimum of that supremum over all N: $H[n] \geq \inf_N \sup_v \{E[v, N_n] - (n|v)\}$. (6.19) holds with the reversed inequality sign, and hence the indeed universal density functional is

$$H[n] = \inf_N \sup_v \{E[v, N] - (n|v)\} \; = \; \inf_N F[n, N]. \tag{6.25}$$

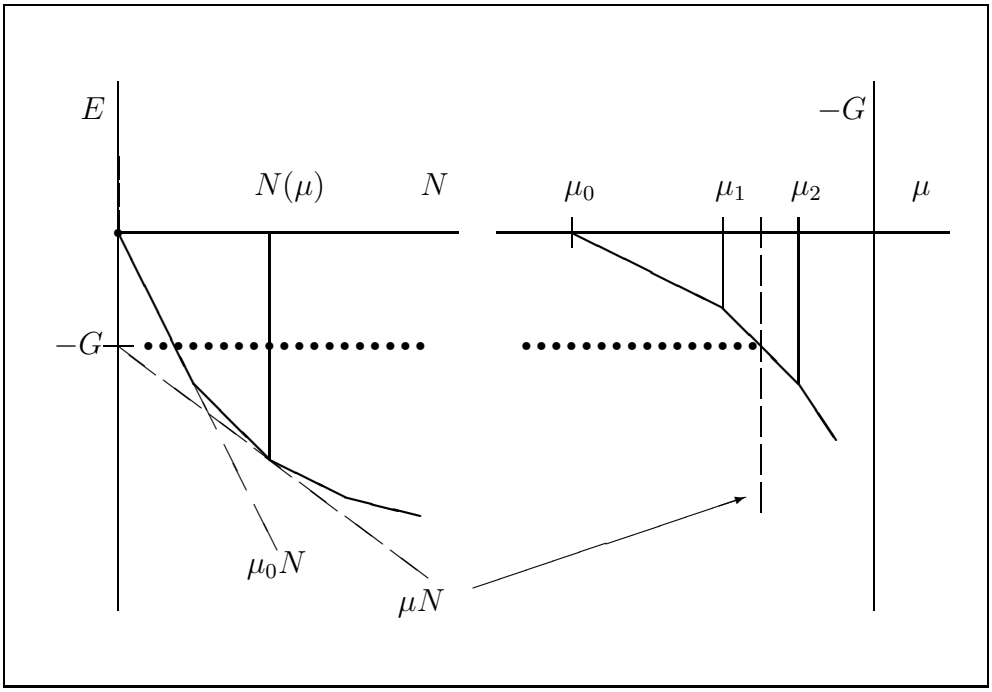

Figure 11: General situation with $E[v, N]$ and $G[v - \mu]$ for $v(x)\mathrm{mod}(\mu)$ fixed as explained in the text. The right panel shows the μ-dependence of $-G[v-\mu]$ according to (6.14) and (6.12).

It is the infimum over N of the density functional by Lieb $F[n, N]$ which latter is defined by (6.20) and depends *independently* on n and N, that is without the connection $N = (n|1)$.

The next question is, what is F in the case $N \neq (n|1)$ or for $n(x)$ negative on a region of non-zero measure (admitted since n varies in a whole functional space, the dual to the space of all potentials). Using the definition (6.20) of F, it is not difficult to show that $F[n, N] = +\infty$ there. To see this, first consider a density $n(x)$ which is negative on a set $\Omega \subset \boldsymbol{T}^3$ of non-zero measure (for the sake of simplicity assumed spin diagonal; the general case reduces to it by a local diagonalization). Take a potential v equal to some positive constant c on Ω and equal to zero on $\boldsymbol{T}^3 \setminus \Omega$. For this potential $v \geq 0$, obviously $E[v, N] \geq 0$ ($E[v, N] \to 0$ for $|\boldsymbol{T}^3| \to \infty$) for every $c > 0$. On the other hand, $-(n|v) = -\int dx\, nv \to +\infty$ for $c \to +\infty$ (recall that n and v are independent in the braces of (6.20)), hence,

$$F[n, N] = +\infty \quad \text{if not } n(x) \geq 0 \text{ a.e.} \tag{6.26}$$

Next consider $N \neq (n|1) = N_n$. Let this time c denote a constant potential equal to the real number c on the whole torus $\boldsymbol{T}^3$: $E[c, N] = E[0, N] + Nc$ and $(n|v) = N_n c$. If $N_n < N$, then the expression in braces of (6.20) tends to $+\infty$ for $c \to +\infty$, in the opposite case $N_n > N$ for $c \to -\infty$, hence,

$$F[n, N] = +\infty \quad \text{if not } (n|1) = N. \tag{6.27}$$

This is just a generalization of the situation sketched in Fig.10a. A density change $\delta n(x)$ on the hypersurface $\int dx\, n = N$ in the functional space of densities is characterized by the subspace $\int dx\, \delta n = 0$. Then, for $v = c$, $\int dx\, \delta n\, c = 0$, i.e., constant potentials $v = c$ are 'perpendicular' to that subspace (they form the annullator of that subspace), and $E[v, N]$ depends linearly on $v = c$. Hence, 'perpendicular' to the hypersurface $\int dx\, n = N$, $F[n, N]$ behaves singularly like $f(c)$ of Fig.10a.

The property (6.27) immediately implies

$$H[n] = \inf_N F[n, N] = F[n, (n|1)], \tag{6.28}$$

and since the inverse Legendre transformation to (6.20) (based on the concavity of $E[v, N]$ in v for fixed N) is

$$E[v, N] = \inf_n \{F[n, N] + (v|n)\} = \inf_{n_N} \{F[n_N, N] + (v|n_N)\} \tag{6.29}$$

with $(n_N|1) = N$. Comparison to (6.28) and (6.22) also proves that the inequality of (6.21) is sharpened into the equality (6.22). Because of (6.28), the property (6.26) is retained for H.

The relation between $F[n, N]$ and $H[n]$ is depicted in Fig.12: $H[n]$ is the convex hull of all $F[n, N]$ for all N. As already mentioned, Lieb defined his functional (6.20) with $E[v]$ from (4.4) for integer N instead of the general $E[v, N]$ from (6.9) for real N. We call the former $E_N[v]$ here to indicate also its N-dependence, and we call the corresponding functional by Lieb $F_N[n]$. Since generally $E[v, N] \leq E_N[v]$, it follows $F[n, N] \leq F_N[n]$. This is also depicted in Fig.12 for an integer value N_3. $H[n]$ by nature relates to $F[n, N]$. One should realize that any explicit dependence of the density functional on the particle number N is a very non-local business (dependence on $\int dx n(x)$) which has never been fully anticipated in approximations to the density functional (except for the kinetic energy part).

It remains to specify the functional spaces for n and v which should be topologically dual to each other (so that the scalar product $(n|v)$ is always finite). In order to define the functional space X for admissible density

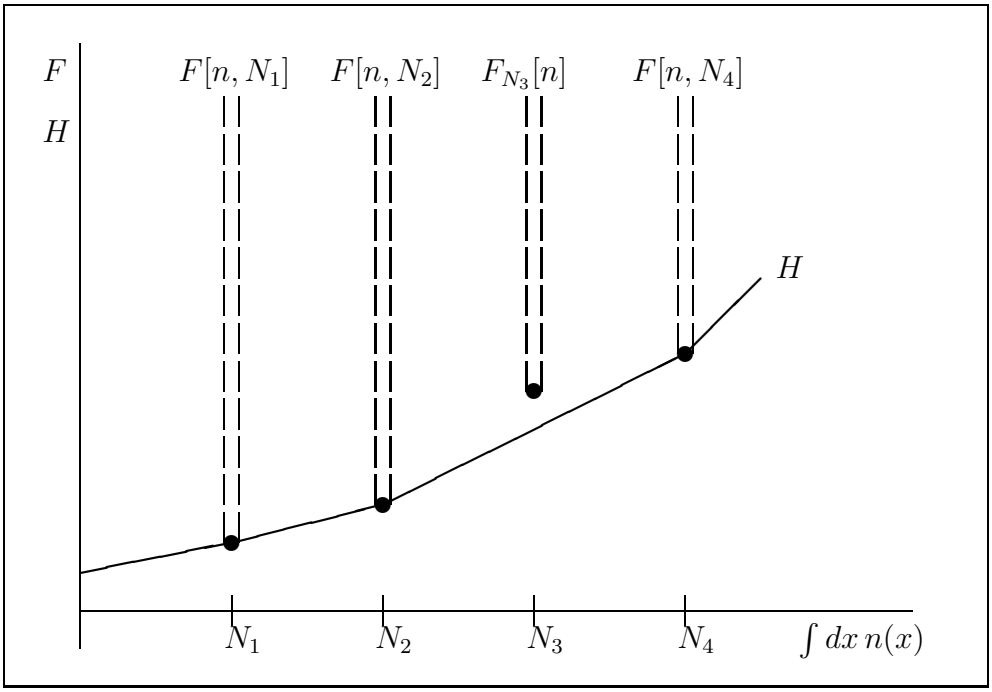

Figure 12: $H[n]$ and $F[n, N_i]$ for several N_i on a path in n-space, 'perpendicular to $\int dx\, n(x) = \text{const.}$' The functional $F[n, N_i]$ is equal to $+\infty$ except for $\int dx\, n(x) = N_i$. H is the convex hull. At N_3, the function $F_{N_3}[n]$ for a case $F_{N_3}[n] > F[n, N_3]$ is sketched.

functions, following Lieb, we first require that the kinetic energy be finite and investigate the consequences on the density of a pure state (ensemble states do not make any difference in this context)

$$\begin{aligned} n(\boldsymbol{r}) &= N \sum_s \int dx_2 \ldots dx_N\, \Psi^*(\boldsymbol{r}s, x_2 \ldots x_N)\Psi(\boldsymbol{r}s, x_2 \ldots x_N) \\ &\stackrel{\text{def}}{=} N\langle\Psi|\Psi\rangle'(\boldsymbol{r}). \end{aligned} \tag{6.30}$$

For every fixed $\boldsymbol{r}$, this expression has all properties of a scalar product, denoted by $\langle\cdot|\cdot\rangle'$. The gradient of the density reads now $\nabla n(\boldsymbol{r}) = N\langle\nabla\Psi|\Psi\rangle' + N\langle\Psi|\nabla\Psi\rangle' = 2N\mathrm{Re}\,\langle\nabla\Psi|\Psi\rangle'$, and its square is

$$[\nabla n]^2 \leq 4N^2|\langle\nabla\Psi|\Psi\rangle'|^2 \leq 4N^2\langle\Psi|\Psi\rangle'\langle\nabla\Psi|\nabla\Psi\rangle' = 4Nn(\boldsymbol{r})\langle\nabla\Psi|\nabla\Psi\rangle'. \tag{6.31}$$

The first inequality just estimates the square of the real part by the square of the absolute value, and the second inequality is an application

of Schwarz' inequality (5.30). The kinetic energy, on the other hand, is $\langle \hat{T} \rangle = (N/2) \int d^3r \, \langle \nabla\Psi | \nabla\Psi \rangle'$, and hence, with $\nabla n^{1/2} = \nabla n / 2n^{1/2}$,

$$\frac{1}{2} \int d^3r \, \left[\nabla n^{1/2} \right]^2 = \int d^3r \, \frac{[\nabla n]^2}{8n} \leq \langle \hat{T} \rangle. \tag{6.32}$$

Therefore, $\langle \hat{T} \rangle < \infty$ implies $\nabla n^{1/2} \in \boldsymbol{L}^2$. (Of course, that does not yet guarantee the opposite: that $\langle \hat{T} \rangle$ is finite for $\nabla n^{1/2} \in \boldsymbol{L}^2$.)

There is a general inequality by Sobolev (see [Lieb, 1983]), in three dimensions estimating the $\boldsymbol{L}^2$-norm of ∇f by the $\boldsymbol{L}^6$-norm of f:

$$\|\nabla f\|_2^2 \geq 3 \left(\frac{\pi}{2} \right)^{4/3} \|f\|_6^2. \tag{6.33}$$

Applying this to $f = n^{1/2}$ and considering $\|n^{1/2}\|_6^2 = \|n\|_3$ and (6.32) yields

$$\frac{3}{2} \left(\frac{\pi}{2} \right)^{4/3} \|n\|_3 \leq \langle \hat{T} \rangle < \infty \quad \Longrightarrow \quad n \in \boldsymbol{L}^3. \tag{6.34}$$

Moreover, of course $N = \int d^3r \, n \Rightarrow n \in \boldsymbol{L}^1$.

Therefore, Lieb considered

$$X = \boldsymbol{L}^3(\boldsymbol{R}^3) \cap \boldsymbol{L}^1(\boldsymbol{R}^3), \quad X^* = \boldsymbol{L}^{3/2}(\boldsymbol{R}^3) + \boldsymbol{L}^\infty(\boldsymbol{R}^3). \tag{6.35}$$

In this case, X^* includes Coulomb potentials as seen in complete analogy to Section 3.1, with use of the decomposition (3.7). Also, by the same argument as in Section 3.1, only with the $\boldsymbol{L}^{5/3}$ replaced by the $\boldsymbol{L}^{3/2}$, the Coulomb pair interaction is finite for $n \in X$. Hence the total Coulomb energy is finite for $n \in X$. If now X comprises all ground states of Coulomb systems, then it follows from the virial theorem (2.81) for Coulomb systems that $E[v, N]$ is finite on X^* in that case. Using the concavity of $E[v, N]$ in v and the estimate of $\langle \hat{T} \rangle$ from (6.34), it has been shown [Lieb, 1983] for *any* interaction w, that $E[v, N]$ (with $N \geq 0$) is finite and hence continuous on X^*, if only $E[v_0]$ is finite for some $v_0 \in X^*$.

From (4.39, 4.41), obviously

$$X \supset \mathcal{J}_N \supset \mathcal{A}_N. \tag{6.36}$$

In the spin-dependent case, each matrix element of $n_{ss'}(\boldsymbol{r})$ and $v_{ss'}(\boldsymbol{r})$ must be in X and X^*, respectively.

Unfortunately, although $X^{**} \supset X$, since $\boldsymbol{L}^1(\boldsymbol{R}^3)$ is not reflexive, $X^{**} \neq X$ from (6.35), and X and X^* would not be mutually dual. This is not

a problem any more when a torus $\boldsymbol{T}^3$ of finite volume $|\boldsymbol{T}^3|$ is considered instead of the $\boldsymbol{R}^3$. In this case,

$$X = \boldsymbol{L}^3(\boldsymbol{T}^3), \quad X^* = \boldsymbol{L}^{3/2}(\boldsymbol{T}^3), \tag{6.37}$$

and we need not further require $n \in \boldsymbol{L}^1$, since we know from (5.33) that the $\boldsymbol{L}^3(\boldsymbol{T}^3)$ is a subspace of the $\boldsymbol{L}^1(\boldsymbol{T}^3)$. (We need not further be concerned with the behavior of functions for $\boldsymbol{r} \to \infty$.) In this case, X and X^* are reflexive, and (6.22, 6.25) hold on grounds of the general theory. One further advantage is that this approach covers also the cases of extended systems.

Of course, replacing the $\boldsymbol{r}$-space $\boldsymbol{R}^3$ by the torus $\boldsymbol{T}^3$ of finite volume with periodic boundary conditions also modifies the sets $\mathcal{A}_N$, $\mathcal{A}_N^0$ and $\mathcal{J}_N$ (as well as $\mathcal{V}_N$). Here we consider quantum mechanics on a torus $\boldsymbol{T}^3$, hence X and X^* denote the reflexive Banach spaces (6.37).

For $v \in \boldsymbol{L}^{3/2}(\boldsymbol{T}^3) \subseteq \mathcal{V}_N$ we have

$$\begin{aligned}
E_N[v] &= \min_{n \in \mathcal{A}_N} \{F_{\mathrm{HK}}[n] + (n|v)\} = \\
&= \min_{n \in \mathcal{J}_N} \{F_{\mathrm{LL}}[n] + (n|v)\} = \\
&= \min_{n \in \mathcal{J}_N} \{F_{\mathrm{DM}}[n] + (n|v)\} = \\
&= \min_{n \in \mathcal{J}_N} \{F_N[n] + (n|v)\} \qquad (6.38) \\
E[v, N] &= \inf_{n \in X} \{F[n, N] + (n|v)\} = \\
&= \inf_{n \in X} \{H[n] + (n|v) \mid (n|1) = N\}. \qquad (6.39)
\end{aligned}$$

Since on a torus the N-particle ground state always exists (and therefore $\boldsymbol{L}^{3/2}(\boldsymbol{T}^3) \subseteq \mathcal{V}_N$), the minimizing density also exists in (6.38). Yet, for the general ensemble state the existence of a minimizing $\hat{\gamma}$ cannot be guaranteed in (6.9) although as regards physics, it will not exist for very exotic v and w only. So far we retained the inf sign in (6.39). However, even if there is no limiting $\hat{\gamma}$ for a series $\{\hat{\gamma}_j\}$ which leads to the infimum of (6.9), the corresponding series of densities $\{n_j\}$ may still converge yielding a minimum in (6.39). This consideration is continued in the next section. In view of (6.36), the above relations imply

$$\begin{aligned}
&n \in \mathcal{A}_N \\
&\quad \Downarrow \\
&F_{\mathrm{HK}}[n] = F_{\mathrm{LL}}[n] = F_{\mathrm{DM}}[n] = F_N[n] \geq F[n, N] = H[n]. \qquad (6.40)
\end{aligned}$$

For $n \in \mathcal{J}_N$ we found $F_{\rm LL} \geq F_{\rm DM}$ in Section 4.5. As a conjugate functional, F_N is convex, whereas $F_{\rm LL}$ is not. Moreover, one can show that $F_{\rm DM} = F_N$ on $\mathcal{J}_N$, and both are equal to the convex hull of $F_{\rm LL}$ on $\mathcal{J}_N$ [Lieb, 1983]. In summary,

$$n \in \mathcal{J}_N \quad \Longrightarrow \quad F_{\rm LL}[n] \geq F_{\rm DM}[n] = F_N[n] \geq F[n,N] = H[n]. \tag{6.41}$$

Only $F_N[n]$, $F[n,N]$, and $H[n]$ are defined on a whole linear space X.

This completes the rigorous density functional theory for the ground state. The version outlined here is particularly appropriate to shed light on the general character of the Hohenberg-Kohn variational principle: Except from the concavity of $E[v,N]$ in v and its convexity in N (and its simple gauge property) the relations (6.22, 6.25) by themselves do not contain any particular physics. This type of theory gains physical content only via constructive expressions for $H[n]$ (or one of the alternative density functionals).

6.4 The Kohn-Sham Equation

A crucial point for deriving Kohn-Sham equations is the existence of the employed functional derivatives. Moreover, the point of solution of the differential equation (as a point in the functional space) must be an inner point of the range of differentiability of the considered quantities (cf. Fig.3a). Formally, Euler's equation for the problem (6.22) is

$$-\frac{\delta H}{\delta n} = v - \mu, \tag{6.42}$$

where μ is the Lagrange multiplier of the side condition. As a Legendre transform, $H[n]$ has a non-empty subdifferential for all n for which it is finite and for which the supremum (6.17) is a maximum. This follows immediately from the general property (5.47) of Legendre transforms. It is naturally conjectured that due to the compactness of the basic space $\boldsymbol{T}^3$ all infima and suprema of this chapter are minima and maxima in cases they are finite.

Indeed, $\|n\|_3 \leq (2/3)(2/\pi)^{4/3}\langle\hat{T}\rangle$ from (6.34), and hence, for a given $v \in X^*$, the search in (6.39) may be restricted to $\|n\|_3 \leq C$ for a sufficiently large positive $C[v]$ taken from an upper estimate of the kinetic energy for that case. The ball $\{n \,|\, \|n\|_3 \leq C\}$ of radius C is weak compact in X by the Banach-Alaoglu theorem and by the reflexivity of X, and hence the main theorem for extremal problems, reported at the end of Section 5.5, guarantees the existence of a solution $n_0(x)$, $E[v] = H[n_0] + \int dx\, n_0 v$, of the

last variational problem of (6.39). Furthermore, $n_0 \in \mathcal{J}_N$. Apart from exotic cases, there is even $n_0 \in \mathcal{A}_N$ by the very definition of $\mathcal{A}_N$, and hence the infima in (6.39) might be replaced by minima for every $v \in X^* = \boldsymbol{L}^{3/2}(\boldsymbol{T}^3)$.

Since H is finite on $\mathcal{J}_N$ of (4.39), it has a non-empty subdifferential for $n \in \mathcal{J}_N$ (even if n is not an inner point of $\mathcal{J}_N$ in the topology of X; such densities, for instance having nodes so that indefinite density functions exist in every neighborhood of n, have been discussed in the literature). Moreover, a convex function on a normed space has a derivative (G-derivative), if and only if the subdifferential consists a unique element of the dual space. (For finite dimensions this is easily realized.) Hence,

> *if $v - \mu \in X^*$ is uniquely defined in the theory for some given $n \in X$ for which $H[n]$ is finite, then $H[n]$ has there a functional derivative equal to $\mu - v$.*

This is the place where the basic theorem by Hohenberg and Kohn (Section 4.1) plays its important role in the theory. (The proof given there goes through with ensemble states of degenerate ground states; certain problems with external magnetic fields as discussed in Sections 4.7 and 4.8 remain.) At least for scalar external potentials, v is uniquely defined by $n \in \mathcal{J}_N$ up to a constant in space, and this constant, μ, is uniquely defined by the particle number N except possibly for integer values of N, where it can make a finite jump (cf. (6.23), (6.24) and Fig.11). At least for the spin-independent theory and for impure spin states,

> the functional derivative $\delta H/\delta n \in X^*$ exists and is equal to $\mu - v$ for non-integer $N = (n|1)$. For integer N it may jump by a finite value, constant in $\boldsymbol{r}$-space.

This establishes the rigorous basis for the Kohn-Sham theory in the form formally developed in Section 4.5. Define a kinetic energy functional for non-interacting reference systems as

$$K[n] \stackrel{\text{def}}{=} \begin{cases} T_{\mathrm{DM}}[n] & for \quad n \in X, n(x) \geq 0 \\ +\infty & elsewhere \end{cases} \tag{6.43}$$

(note that by definition of X, T_{DM} of (4.57) is well determined for $n(x) \geq 0$), and define E_{XC} by

$$H[n] = K[n] + E_{\mathrm{H}}[n] + E_{\mathrm{XC}}[n], \quad n \in X = \boldsymbol{L}^3(\boldsymbol{T}^3). \tag{6.44}$$

Now, E_{XC} can also be defined on the whole X: it is defined for $n(x) \geq 0$ by the above relation, and it can be anything elsewhere, where both H and K

are $+\infty$. (It is left to the reader's choice to understand the letters H and K as a tribute to Hohenberg and Kohn.) Substituting this representation of $H[n]$ into (6.22) and combining the density variation again with the orbital variation of (4.57) into one single step leads to the Kohn-Sham equation exactly in the way as previously derived at the end of Section 4.5, that is, to

$$\left(-\frac{\nabla^2}{2} + v_{\text{eff}}\right)\phi_i = \phi_i\,\varepsilon_i, \quad v_{\text{eff}} \stackrel{\text{def}}{=} v + v_{\text{H}} + v_{\text{XC}}, \tag{6.45}$$

$$v_{\text{XC},ss'}(\boldsymbol{r}) \stackrel{\text{def}}{=} \frac{\delta E_{\text{XC}}[n]}{\delta n_{s's}(\boldsymbol{r})}, \tag{6.46}$$

$$\varepsilon_1 \le \varepsilon_2 \le \cdots, \tag{6.47}$$

$$n_i = 1 \text{ for } \varepsilon_i < \varepsilon_N,\ 0 \le n_i \le 1 \text{ for } \varepsilon_i = \varepsilon_N,\ n_i = 0 \text{ for } \varepsilon_i > \varepsilon_N. \tag{6.48}$$

$$n_{ss'}(\boldsymbol{r}) = \sum_i \phi_i(\boldsymbol{r}s) n_i \phi_i^*(\boldsymbol{r}s'), \quad N = \sum_s \int d^3r\, n_{ss}(\boldsymbol{r}), \tag{6.49}$$

$$E[v,N] = \sum_i n_i\varepsilon_i - E_{\text{H}}[n] - \sum_{ss'} \int d^3r\, n_{ss'} v_{\text{XC},s's} + E_{\text{XC}}[n], \tag{6.50}$$

Also with respect to these equations no representability problems are left.

It is not difficult to convince himself that $K[n]$ as defined in (6.43) coincides with $H[n]$ as defined in (6.25) for $w \equiv 0$. Both have jumps of the derivative by finite constant values in $\boldsymbol{r}$-space at integer values of N. In the interaction free case $w \equiv 0$ those jumps (from one HOMO to the next as N increases) are exactly modeled in the construction (4.57). It is naturally expected that the heights of those jumps depend on the interaction w, hence, from (6.44), $v_{\text{XC}} = \delta E_{\text{XC}}[n]/\delta n$ should also jump by a constant spin matrix, independent of $\boldsymbol{r}$, if N runs through an integer [Perdew and Levy, 1983, Sham and Schlüter, 1983, Lannoo *et al.*, 1985]. (Pure spin states may lead to spin-dependent jumps.)

Orbital variation expressions of the general type (4.57) are a convenient tool to model non-local functional dependences on the density n. Put

$$H[n] = K[n] + L[n], \tag{6.51}$$

$$K[n] = \min_{\phi_i, n_i} \left\{ k[\phi_i, n_i] \;\middle|\; \sum_i \phi_i n_i \phi_i^* = n,\ 0 \le n_i \le 1 \right\}, \tag{6.52}$$

$$L[n] = \int d^3r n(\boldsymbol{r}) l(n_{ss'}(\boldsymbol{r}), \nabla n_{ss'}(\boldsymbol{r}), \ldots) \tag{6.53}$$

with a suitable chosen orbital functional k and a suitable function $l(n,\ldots)$ one derives Kohn-Sham type equations

$$\big(\hat{k} + v + v_L\big)\phi_i = \phi_i \varepsilon_i \tag{6.54}$$

with an orbital operator

$$\hat{k}: \quad \frac{\delta k}{\delta \phi_i^*} = \hat{k}\phi_i n_i, \tag{6.55}$$

and a local Kohn-Sham potential

$$v_L = \frac{\delta L}{\delta n} \tag{6.56}$$

as well as the aufbau principle (6.47) to (6.49).

The restriction to the torus $\boldsymbol{T}^3$ of finite volume brought it about, that Eqs. (6.25, 6.43, 4.57, 6.44–6.50) provide a mathematically well defined closed theory. There is no N-representability problem of potentials (unknown set $\mathcal{V}_N$) and no v-representability problem of densities (unknown set $\mathcal{A}_N$) any more, because those unknown sets do not figure any more in the theory, and there is no problem of the existence of functional derivatives with respect to densities left. There are also no restrictions of applicability of the Kohn-Sham equations to be considered.

If one is interested in N-particle eigenstates as ground states and in their energies $E_N[v]$ instead of density matrix ground states with energies $E[v,N]$, the former states may not be found and their energies may only be estimated from below, if $E_N[v]$ is not convex as a function of N. (Note that the reason for a failure of convexity is always interaction: $E_N^0[v]$ is convex for every potential v.) On the other hand, $E[v,N]$ is always concave in v and convex in N by its very definition (4.66) as a thermodynamic quantity (grand canonical energy). For fixed v, it is the convex hull of $E_N[v]$.

7 Approximative Variants

The density functional theory presented in Chapters 4 and 6 is a rigorous theory on the safe basis of many-body quantum theory. It is, however, not explicit, so that practical computations cannot directly be based upon it. On the other hand, Thomas-Fermi theory, although originally very naïve, could be understood as a crude approximation to density functional theory in the above sense, with the advantage of being explicit. The use of the Kohn-Sham trick in handling the kinetic energy and treating the exchange and correlation energy functional in the spirit of Thomas and Fermi leads to the explicit local density approximation, which is the basis or at least the starting point of nearly all explicit approximative variants of density functional theory in use. The aim of this last chapter of Part I is to link the material presented in the previous ones to practical applications.

So far, and very likely also in future, we do not have a systematic access to the rigorously defined density functional $H[n]$ the theory is based upon. Hence we have to model it and to probe the models by comparison to phenomenology. This situation is not principally different from other many-particle approaches where either models of sufficiently simple Hamiltonians are used (in quantum field theory) or the wave function is modeled (for instance in Hartree-Fock or Gutzwiller approaches). It is important to realize that all numerical density functional results provided so far are *model results* although many of them like for instance the local density approximation do not contain any adjustable parameter, whence they are often termed *ab-initio*.

The basis of the local density approximation for the exchange and correlation energy functional is the theory of the homogeneous electron liquid. This is a most important model system, which of course does not exist in nature, but which can nowadays theoretically be treated with extremely high precision. The results and parameterizations of that theory are reviewed in the first section. Subsequently, the local density approximation is introduced, and explanation is given how and why it works, and in which respect it does not work.

All practical density functional approaches from Thomas-Fermi up to the most sophisticated current ones may be classified with respect to the

manner of splitting the basic density functional $H[n]$ according to (6.44) or more generally according to (6.51)–(6.53) into a non-local part expressed through an orbital variation expression and a part subject to some local density approximation. Except for the Hartree term, the non-local part is usually only implicitly given by an orbital variation expression. This classification is carried through in Section 7.3 of the present chapter.

As two rather successful examples related to that classification, in the last two sections of this chapter we consider the self-interaction correction to the local density approximation and the LDA+U approach. For other approaches beyond the original local density approximation, particularly those based on gradient expansions, we refer the reader to the existing literature. (See, for example [Dreizler and Gross, 1990]).

7.1 The Homogeneous Electron Liquid

The homogeneous electron liquid is an important reference model system, which, however, does not exist in nature. Nevertheless, due to sophisticated theoretical approaches as well as computer simulations, many of its properties are known to an uncomparably high precision. In the literature it is sometimes called the homogeneous electron gas; however, we reserve this name for the interaction-free case (1.41–1.45).

The homogeneous electron liquid is a variant of the Coulomb system (2.69), only with the point charges of the nuclei replaced by a homogeneous, i.e. constant in space, positive charge density

$$n_+ = \left[\frac{4\pi r_s^3}{3}\right]^{-1}, \tag{7.1}$$

where, as in (3.30), the density parameter r_s is the radius of a sphere containing one charge quantum. Hence its Hamiltonian is (formally)

$$\begin{aligned}\hat{H}_{\text{hom}} &= -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - \sum_{i=1}^{N}\int d^3r\,\frac{n_+}{|\boldsymbol{r}_i-\boldsymbol{r}|} + \\ &\quad +\frac{1}{2}\sum_{i\neq j}^{N}\frac{1}{|\boldsymbol{r}_i-\boldsymbol{r}_j|} + \frac{1}{2}\int d^3r\,d^3r'\,\frac{n_+^2}{|\boldsymbol{r}-\boldsymbol{r}'|}.\end{aligned} \tag{7.2}$$

This Hamiltonian is formal because the figuring integrals diverge. As in

Section 2.7, it has to be replaced by

$$\begin{aligned}\hat{H}_{\alpha L,\mathrm{hom}} &= -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - \sum_{i=1}^{N}\int_{\boldsymbol{T}^3} d^3r\, n_+\, v_{\alpha L}(\boldsymbol{r}_i-\boldsymbol{r}) + \\ &\quad +\frac{1}{2}\sum_{i\neq j}^{N} v_{\alpha L}(\boldsymbol{r}_i-\boldsymbol{r}_j) + \frac{1}{2}\int_{\boldsymbol{T}^3} d^3r\, d^3r'\, n_+^2\, v_{\alpha L}(\boldsymbol{r}-\boldsymbol{r}') \quad (7.3)\end{aligned}$$

with

$$N = n_+|\boldsymbol{T}^3| = \left[\frac{4\pi r_s^3}{3}\right]^{-1} L^3. \quad (7.4)$$

Afterwards,

$$\varepsilon(r_s) = (E/N)(r_s) = \lim_{L\to\infty}\frac{1}{N}\lim_{\alpha\downarrow 0} E_{\alpha L}(r_s) \quad (7.5)$$

is to be considered as the ground state energy per electron in dependence on the density parameter r_s.

The external potential $v(\boldsymbol{r}_i) = -\int_{\boldsymbol{T}^3} d^3r\, n_+\, v_{\alpha L}(\boldsymbol{r}_i-\boldsymbol{r})$, produced by n_+ and contained in the second term of (7.3), is the Yukawa interaction energy of the electron at $\boldsymbol{r}_i$ with a homogeneous charge distribution on the torus, visualized as a periodically repeated cube, and hence is independent of $\boldsymbol{r}_i$. It is a potential constant, and N times this constant is exactly canceled by the sum of the spatial average over the electron-electron interaction (third term of (7.3) averaged over all positions $\boldsymbol{r}_i$ and $\boldsymbol{r}_j$) and the likewise constant last item of (7.3). This cancellation prevails in the α-limit (where the individual terms tend to infinity), and hence the Hamiltonian in this limit, but still at finite L, may be given in momentum representation (1.34) as

$$\begin{aligned}\langle\, q_1\ldots q_N\,|\hat{H}|\, q_1'\ldots q_N'\,\rangle &= \frac{1}{2}\sum_i \boldsymbol{k}_i^2 \prod_j \delta_{q_jq_j'} + \\ &+\frac{1}{2}\sum_{i\neq j}\frac{\delta_{\boldsymbol{k}_{\bar{\imath}}+\boldsymbol{k}_{\bar{\jmath}},\boldsymbol{k}_i'+\boldsymbol{k}_j'}}{L^3}\Big[\delta_{s_{\bar{\imath}}s_i'}\delta_{s_{\bar{\jmath}}s_j'}w_{|\boldsymbol{k}_{\bar{\imath}}-\boldsymbol{k}_i'|} - \\ &\qquad -\delta_{s_{\bar{\imath}}s_j'}\delta_{s_{\bar{\jmath}}s_i'}w_{|\boldsymbol{k}_{\bar{\imath}}-\boldsymbol{k}_j'|}\Big]\,(-1)^{\bar{\mathcal{P}}}\prod_{l(\neq i,j)}\delta_{q_{\bar{l}}q_l'} \quad (7.6)\end{aligned}$$

with

$$w_{|\boldsymbol{k}|=0} = 0, \quad w_{|\boldsymbol{k}|\neq 0} = \frac{4\pi}{L^3|\boldsymbol{k}|^2}, \quad (7.7)$$

and the $\boldsymbol{k}$-vectors given by (1.29).

Note that deviating from our general form (4.1) the Hamiltonian (7.3) (and likewise (2.86)) contains a constant term, which is finite for non-zero α only. This term together with a constant external potential is just combined with the long-range part of the electron-electron interaction to modify the latter in the way given in (7.7). After this little trick, α may be let go to zero, and with (7.6) we are back in our frame of theory, with a Coulomb potential screened by the strict charge neutrality of the torus $\boldsymbol{T}^3$. For an L large compared to the diameter of an exchange and correlation hole, this latter screening concerns only the Hartree energy and leaves the exchange and correlation energy unaltered. This is the general way density functional theory applies to extended Coulomb systems. The L-limit then is always understood to be taken at the very end, because otherwise all density functionals would be infinite and moreover the constancy constraint for the particle number could not be formulated any more.

Spin-polarized ground states may be considered at any value of the density parameter r_s, eventually by adding a spin-polarizing spatially constant potential term to (7.6).

The simplest case is of course again the interaction-free homogeneous electron gas. Its ground state density is constant and equal to n_+ at every r_s-value. We relate the electron density n, its spin components $n(+)$ and $n(-)$, and the degree of spin polarization ζ of a collinear spin situation (spin-density matrix diagonal with respect to a global z-direction, cf. (2.33–2.37)) according to

$$n = n(+) + n(-), \quad \zeta = \frac{n(+) - n(-)}{n}, \tag{7.8}$$

$$n(+) = \frac{n}{2}(1+\zeta), \quad n(-) = \frac{n}{2}(1-\zeta) \tag{7.9}$$

and have for the kinetic energy in Hartree units (cf. (1.45, 1.43 and 7.1))

$$\begin{aligned} t(n,\zeta) = \frac{T(n,\zeta)}{N} &= \frac{3}{10}\left(\frac{9\pi}{4}\right)^{2/3} \frac{(1+\zeta)^{5/3} + (1-\zeta)^{5/3}}{2r_s(n)^2} = \\ &= \frac{1.1049}{r_s(n)^2}\,\frac{1}{2}\left[(1+\zeta)^{5/3} + (1-\zeta)^{5/3}\right]. \end{aligned} \tag{7.10}$$

The minimum is at $\zeta = 0$ for every r_s, that is, without an external magnetic field the homogeneous electron gas is not spin-polarized.

In lowest order many-body perturbation theory the exchange energy of the homogeneous electron liquid is obtained as

$$\begin{aligned}\varepsilon_{\mathrm{X},1}(n,\zeta) &= -\frac{3}{4\pi}\left(\frac{9\pi}{4}\right)^{1/3}\frac{(1+\zeta)^{4/3}+(1-\zeta)^{4/3}}{2r_s(n)} = \\ &= -\frac{0.4582}{r_s(n)}\,\frac{1}{2}\left[(1+\zeta)^{4/3}+(1-\zeta)^{4/3}\right]. \end{aligned} \tag{7.11}$$

This contribution dominates in the polarization dependence, therefore, following a suggestion of von Barth and Hedin [von Barth and Hedin, 1972], the polarization dependence of the exchange and correlation energy E_{XC} of (2.65) of the homogeneous electron liquid is generally interpolated between the paramagnetic ($\zeta = 0$) and the saturated ferromagnetic ($\zeta = 1$) cases according to

$$\varepsilon_{\mathrm{XC}}(n,\zeta) = \varepsilon_{\mathrm{XC}}(n,0) + [\varepsilon_{\mathrm{XC}}(n,1) - \varepsilon_{\mathrm{XC}}(n,0)]f(\zeta), \tag{7.12}$$

where

$$f(\zeta) = \frac{(1+\zeta)^{4/3}+(1-\zeta)^{4/3}-2}{2^{4/3}-2} \tag{7.13}$$

is the von Barth-Hedin interpolation function. From (7.8), $-1 \le \zeta \le 1$. The interpolation function $f(\zeta)$ is shown on Fig.13; it is symmetric in ζ and so is $\varepsilon_{\mathrm{XC}}$:

$$f(-\zeta) = f(\zeta), \quad \varepsilon_{\mathrm{XC}}(n,-\zeta) = \varepsilon_{\mathrm{XC}}(n,\zeta). \tag{7.14}$$

The exchange and correlation energy does of course not depend on whether the spin-polarization is up or down.

The functions $\varepsilon_{\mathrm{XC}}(n,0)$ and $\varepsilon_{\mathrm{XC}}(n,1)$ for the homogeneous ground state are obtained from numerical total energy calculations by taking the difference of the total energy and the kinetic energy of the interaction-free case (7.10) at the same values n and ζ. The presently most accurate results distinguish two regions of r_s: For small values of r_s (high density), many-body perturbation theory with partial summations of diagrams yields in the paramagnetic case

$$\varepsilon_{\mathrm{XC}}(n,0) = -\frac{0.4582}{r_s} + 0.03109\ln r_s + B + Cr_s\ln r_s + Dr_s + \cdots \tag{7.15}$$

The first term was already given in (7.11), the second term, whose coefficient is $(1-\ln 2)/\pi^2$, was first obtained in [Macke, 1950], and numerical

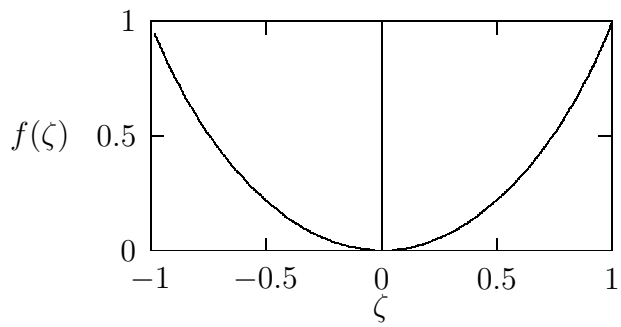

Figure 13: Spin polarization interpolation function $f(\zeta)$ of (7.13).

values of the next terms may be found in [Carr and Maradudin, 1964]. The logarithmic terms arise from partial summations of diagrams diverging due to the long-range part of the Coulomb interaction. The ζ-dependence of the so-called random-phase contributions dominating in (7.15) is given by the scaling relation [Hedin, 1965, Misawa, 1965]

$$\varepsilon_{\mathrm{XC}}^{\mathrm{RPA}}(n,1) = \frac{1}{2}\varepsilon_{\mathrm{XC}}^{\mathrm{RPA}}(2^4 n, 0), \tag{7.16}$$

which is consistent with (7.11). For the total energy of the homogeneous electron liquid at intermediate and large values of r_s, numerical quantum Monte Carlo results exist [Ceperly and Alder, 1980].

Perdew and Zunger [Perdew and Zunger, 1981] proposed an interpolation of those numerical results with Padé approximants in $r_s^{1/2}$ for $r_s > 1$, and recommended to use the high-density many-body perturbation expansion for $0 < r_s \leq 1$ with coefficients C and D of (7.15) fixed in such a way that $\varepsilon_{\mathrm{XC}}(n,\zeta)$ and its first derivative with respect to n is continuous at $r_s = 1$. The result is

$$\begin{aligned} \varepsilon_{\mathrm{XC}}(n,0) &= -\frac{0.4582}{r_s} + 0.03109 \ln r_s - \\ &\quad - 0.0480 + 0.0020 r_s \ln r_s - 0.0116 r_s, \end{aligned} \tag{7.17}$$

$$\begin{aligned} \varepsilon_{\mathrm{XC}}(n,1) &= -\frac{0.5773}{r_s} + 0.01555 \ln r_s - \\ &\quad - 0.0269 + 0.0007 r_s \ln r_s - 0.0048 r_s \end{aligned} \tag{7.18}$$

for $0 < r_s \leq 1$, and

$$\varepsilon_{\rm XC}(n,0) = -\frac{0.4582}{r_s} - \frac{0.1423}{1+1.0529\sqrt{r_s}+0.3334r_s}, \tag{7.19}$$

$$\varepsilon_{\rm XC}(n,1) = -\frac{0.5773}{r_s} - \frac{0.0843}{1+1.3981\sqrt{r_s}+0.2611r_s} \tag{7.20}$$

for $1 < r_s$. There are alternative parameterizations essentially reproducing the same results.[7]

7.2 The Local Density Approximation

The spin density of an inhomogeneous system in a collinear situation

$$n(x) = \begin{pmatrix} n(\boldsymbol{r},+) \\ n(\boldsymbol{r},-) \end{pmatrix} = n(\boldsymbol{r}) \begin{pmatrix} (1+\zeta(\boldsymbol{r}))/2 \\ (1-\zeta(\boldsymbol{r}))/2 \end{pmatrix} \tag{7.21}$$

may be expressed by the two functions $n(\boldsymbol{r})$ and $\zeta(\boldsymbol{r})$. Since n is the particle number per unit volume, and $\varepsilon_{\rm XC}(n,\zeta)$ is the exchange and correlation energy per particle of the homogeneous electron liquid, $n\varepsilon_{\rm XC}(n,\zeta)$ is the energy per unit volume. Hence, in the spirit of the Thomas-Fermi theory,[8]

$$E_{\rm XC}[n] \approx E_{\rm XC}^{\rm LDA}[n] \stackrel{\rm def}{=} \int d^3r\, n(\boldsymbol{r})\, \varepsilon_{\rm XC}(n(\boldsymbol{r}),\zeta(\boldsymbol{r})) \tag{7.22}$$

may be taken as an approximation for the exchange and correlation energy functional of an inhomogeneous system, defined by (6.44). The Kohn-Sham exchange and correlation potential (6.46) is then approximated by

$$\begin{aligned} v_{\rm XC}(\boldsymbol{r},\pm) &\approx \frac{\partial}{\partial n(\boldsymbol{r},\pm)}\left\{n(\boldsymbol{r})\varepsilon_{\rm XC}(n(\boldsymbol{r}),\zeta(\boldsymbol{r}))\right\} = \\ &= \varepsilon_{\rm XC}(n(\boldsymbol{r}),\zeta(\boldsymbol{r})) + n(\boldsymbol{r})\frac{\partial\varepsilon_{\rm XC}}{\partial n}(n(\boldsymbol{r}),\zeta(\boldsymbol{r})) \pm \\ &\quad \pm(1\mp\zeta(\boldsymbol{r}))\frac{\partial\varepsilon_{\rm XC}}{\partial\zeta}(n(\boldsymbol{r}),\zeta(\boldsymbol{r})) \stackrel{\rm def}{=} \\ &\stackrel{\rm def}{=} v_{\rm XC\pm}^{\rm LDA}(n(\boldsymbol{r}),\zeta(\boldsymbol{r})). \end{aligned} \tag{7.23}$$

[7] The presently most precise fit is that of J. P. Perdew and Y. Wang, Phys. Rev. **B45**, 13244–13249 (1992).

[8] For gradient corrections see J. P. Perdew, K. Burke and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865–3868 (1996); Errata *ibid.* **78**, 1396 (1997).

Here, we used the rule

$$\frac{\partial}{\partial n(\pm)} = \frac{\partial n}{\partial n(\pm)}\frac{\partial}{\partial n} + \frac{\partial \zeta}{\partial n(\pm)}\frac{\partial}{\partial \zeta} \;=\; \frac{\partial}{\partial n} \pm \frac{(1 \mp \zeta)}{n}\frac{\partial}{\partial \zeta}, \tag{7.24}$$

which immediately follows from (7.8).

With use of the expressions (7.12, 7.17–7.20), $v_{\mathrm{XC}\pm}^{\mathrm{LDA}}$ is continuous for all $n > 0$ and all ζ, and

$$v_{\mathrm{XC}\pm}^{\mathrm{LDA}}(n(\boldsymbol{r}), \zeta(\boldsymbol{r})) = v_{\mathrm{XC}\mp}^{\mathrm{LDA}}(n(\boldsymbol{r}), -\zeta(\boldsymbol{r})) \tag{7.25}$$

as it should be. In a non-collinear spin situation these formulas may be used after a local spin rotation which diagonalizes $n_{ss'}(\boldsymbol{r})$ so that the local z-axis for the spin points in the direction of $\boldsymbol{e}_m$ of (2.37). The potential (7.25) may then be back rotated to global spin variables.

The main reasons, why this scheme works surprisingly well even for strongly inhomogeneous systems as atoms, molecules and solids, can be analyzed in the following manner [Gunnarsson *et al.*, 1979]: According to (2.67), the exchange and correlation energy (as a physical number, not a density functional) of any ground state of a Coulomb system may be expressed via

$$E_{\mathrm{XC}} = \int d^3r\, n(\boldsymbol{r})\epsilon_{\mathrm{XC}}(\boldsymbol{r}), \tag{7.26}$$

$$\epsilon_{\mathrm{XC}}(\boldsymbol{r}) = \frac{1}{2}\int d^3r'\, \frac{h_{\mathrm{KS}}(\boldsymbol{r}', \boldsymbol{r})}{|\boldsymbol{r}' - \boldsymbol{r}|} \tag{7.27}$$

in terms of the Kohn-Sham exchange and correlation hole $h_{\mathrm{KS}}(\boldsymbol{r}', \boldsymbol{r})$ seen by an electron at $\boldsymbol{r}$ and distributed over the $\boldsymbol{r}'$-space. With (2.68), these are exact relations. Furthermore, since the general sum rule (2.51) holds independent of the coupling constant λ, from (2.68) it also follows that

$$\int d^3r'\, h_{\mathrm{KS}}(\boldsymbol{r}', \boldsymbol{r}) = -1. \tag{7.28}$$

All this is especially also valid for the homogeneous electron liquid, where $\epsilon_{\mathrm{KS}}(\boldsymbol{r})$ is constant in $\boldsymbol{r}$ and equal to $\varepsilon_{\mathrm{XC}}(n, \zeta)$, and $h_{\mathrm{KS}}(\boldsymbol{r}', \boldsymbol{r})$ depends isotropically on the vector $\boldsymbol{r}' - \boldsymbol{r}$: $h_{\mathrm{KS}}(\boldsymbol{r}', \boldsymbol{r}) = h_{\mathrm{hom}}(|\boldsymbol{r}' - \boldsymbol{r}|; n, \zeta)$ with $4\pi \int_0^\infty dy\, y^2\, h_{\mathrm{hom}}(y; n, \zeta) = -1$.

In the general inhomogeneous case we expand the angular dependence on this vector according to

$$h_{\mathrm{KS}}(\boldsymbol{r}', \boldsymbol{r}) \stackrel{\mathrm{def}}{=} \tilde{h}(\boldsymbol{r}' - \boldsymbol{r}, \boldsymbol{r}) = \tilde{h}(\boldsymbol{y}, \boldsymbol{r}), \tag{7.29}$$

$$\tilde{h}(\boldsymbol{y},\boldsymbol{r}) = \sqrt{4\pi}\sum_L h_L(y,\boldsymbol{r})Y_L(\hat{\boldsymbol{y}}) \tag{7.30}$$

with spherical harmonics Y_L and $L=(l,m)$, and have the exact relations

$$\epsilon_{\mathrm{XC}}(\boldsymbol{r}) = \frac{1}{2}\int_0^\infty dy\, 4\pi y^2\,\frac{h_0(y,\boldsymbol{r})}{y} = 2\pi\int_0^\infty dy\, y\, h_0(y,\boldsymbol{r}), \tag{7.31}$$

$$4\pi\int_0^\infty dy\, y^2\, h_0(y,\boldsymbol{r}) = -1, \tag{7.32}$$

which are not influenced by the anisotropy of $\tilde{h}(\boldsymbol{y},\boldsymbol{r})$ as a function of $\boldsymbol{y}$. *Because of the isotropy of the Coulomb interaction* $1/|\boldsymbol{r}'-\boldsymbol{r}|$*, the exchange and correlation energy depends on the isotropic part of the exchange and correlation hole only.*

The local density approximation now means

$$\epsilon_{\mathrm{XC}}(\boldsymbol{r}) \approx \varepsilon_{\mathrm{XC}}(n(\boldsymbol{r}),\zeta(\boldsymbol{r})), \tag{7.33}$$

which is equivalent to putting

$$h_0(y,\boldsymbol{r}) \approx h_0^{\mathrm{LDA}}(y,\boldsymbol{r}) \stackrel{\mathrm{def}}{=} h_{\mathrm{hom}}(y;n(\boldsymbol{r}),\zeta(\boldsymbol{r})). \tag{7.34}$$

The shape of the complete exchange and correlation hole $\tilde{h}(\boldsymbol{y},\boldsymbol{r})$ of an atom or molecule or solid has usually very little in common with $h_{\mathrm{hom}}(y;n(\boldsymbol{r}),\zeta(\boldsymbol{r}))$, however, its isotropic part $h_0(y,\boldsymbol{r})$ may still be very close to (7.34), which closeness gets further support by the sum rule, and for these reasons the local density approximation works.

One defect, however, which definitely cannot be remedied within the frame of the local density approximation, is connected with the dependence on particle number. As we know, for a given $\boldsymbol{r}$, the potential $v_{\mathrm{XC},ss'}(\boldsymbol{r})$ is a *functional* of the functions $n_{ss'}(\boldsymbol{r}')$, or likewise of $n(\boldsymbol{r}')$, $\zeta(\boldsymbol{r}')$ and the spin polarization direction $\boldsymbol{e}_m(\boldsymbol{r}')$. Let us write for a moment $v_{\mathrm{XC}}[n,\zeta,\boldsymbol{e}_m]$ in order to make this functional dependence explicit. We know further, that this potential v_{XC} may change discontinuously, if $n(\boldsymbol{r}')$ is changed in such a way that $N=\int d^3r'\,n(\boldsymbol{r}')$ moves through an integer. The local density approximation replaces this functional dependence by a *local* dependence according to (7.23), with possibly an additional spin rotation:

$$v_{\mathrm{XC}}[n,\zeta,\boldsymbol{e}_m](\boldsymbol{r}) \approx v_{\mathrm{XC}}^{\mathrm{LDA}}(n(\boldsymbol{r}),\zeta(\boldsymbol{r}),\boldsymbol{e}_m(\boldsymbol{r})). \tag{7.35}$$

The numbers $n(\boldsymbol{r})$, $\zeta(\boldsymbol{r})$, $\boldsymbol{e}_m(\boldsymbol{r})$ at position $\boldsymbol{r}$ do not contain any more the information on what N is, and hence, at a given $\boldsymbol{r}$, $v_{\mathrm{XC}}^{\mathrm{LDA}}$ does not depend

any more on N; it depends on $n_{ss'}(\boldsymbol{r})$ at that considered value of $\boldsymbol{r}$ only (and that dependence as given by (7.23) is continuous everywhere).

In order not to overload the analysis, in the rest of this chapter the collinear spin case with $\boldsymbol{e}_m$ pointing in the global z-direction is considered only. Non-collinearity introduces apart from the additional local spin rotations no new aspects into the following considerations.

Let N be an integer, and let $\bar{n}(x)$ be a density with $\int dx\, \bar{n}(x) = N$. Let further $\delta n(x)$ be an increment so that $\int dx\, \delta n(x) = \delta N \overset{>}{\to} 0$. Then,

$$v_{\mathrm{XC}}[\bar{n} + \delta n] = v_{\mathrm{XC}}[\bar{n} - \delta n] + \Delta_{\mathrm{XC}}[\bar{n}], \tag{7.36}$$

and it was shown in Section 6.4 that Δ_{XC}, if non-zero, is a constant in space (equal to $\Delta(v_0^H - v_0^K)$ in the notation of Section 6.4), i.e. independent of $\boldsymbol{r}$, the argument of $v_{\mathrm{XC}}(\boldsymbol{r})$. Analogously,

$$\left.\frac{\delta T[n]}{\delta n(x)}\right|_{\bar{n}+\delta n} = \left.\frac{\delta T[n]}{\delta n(x)}\right|_{\bar{n}-\delta n} + \Delta_T[\bar{n}] \tag{7.37}$$

(where Δ_T is equal to Δv_0^K of Section 6.4), this latter discontinuity, however, is automatically implemented in the Kohn-Sham equations via the replacement of $\delta T[n]/\delta n(x)$ by (4.30, 4.20) and is contained in the differences of subsequent Kohn-Sham levels ε_i of (6.47) computed from the *same* potential v_{eff} via (6.45). Since Δ_{XC} is a constant, if it is non-zero, it results just in a shift of the Kohn-Sham levels by that value Δ_{XC}, too, but this shift must be inserted into the Kohn-Sham equations as a potential shift of v_{XC}.

Now, fixing the external potential $v(x)$ and solving hypothetically the exact Kohn-Sham equations for varying N yields ground state densities $\bar{n}[v, N]$. According to Section 4.6, the first excitation energy of the system, the excitation gap Δ, is $I - A$, and in an exact theory this is strictly expressed by the difference between the lowest unoccupied and the highest occupied Kohn-Sham levels. However, the former is to be calculated for $N + 0$ and the latter for $N - 0$. According to our discussion, the former must be calculated with the potential constant Δ_{XC} included (a positive Δ_{XC} enhances the $\varepsilon_{\mathrm{HOMO}}$ and hence reduces A), so that

$$\Delta = I(N - 0) - A(N + 0) = \Delta_T[\bar{n}[v, N]] + \Delta_{\mathrm{XC}}[\bar{n}[v, N]]. \tag{7.38}$$

The constant Δ_{XC} appears as a gap correction to the LDA excitation gap. The local density approximation, even with the use of (4.71, 4.72) (which only in an exact theory should be equal to (4.74)), introduces two gap errors:

$$\Delta_{\mathrm{XC}}^{\mathrm{LDA}} = 0, \quad \Delta_T^{\mathrm{LDA}} = \Delta_T[\bar{n}^{\mathrm{LDA}}]. \tag{7.39}$$

The latter error is caused by the LDA-error of the ground state density. It is still on debate, which of the errors is larger in which situation.

7.3 Generations of Kohn-Sham Type Equations

The expression (7.22) is the prototype of a local density approximation (6.53), which latter we, in view of Section 4.7, always understand as the local spin-density approximation.

The most consequent, or, as one could say, zeroth generation of such an approximation would try an expression

$$H[n] \approx H^{(0)}[n] = \int d^3r\, n(\boldsymbol{r})\, h(n(\boldsymbol{r}), \zeta(\boldsymbol{r})), \tag{7.40}$$

for the density functional H itself, which, given a suitable function h for the homogeneous case, would replace the usual Kohn-Sham equations by a zeroth generation of equations reading (cf. (6.42) and the text thereafter)

$$\mu - v(\boldsymbol{r}, \pm) = \frac{\delta H^{(0)}}{\delta n(\boldsymbol{r}, \pm)} = h + n\frac{\partial h}{\partial n} \pm (1 \mp \zeta)\frac{\partial h}{\partial \zeta}. \tag{7.41}$$

This would be a set of two transcendent (i.e. non-linear) equations connecting $n(\boldsymbol{r})$ and $\zeta(\boldsymbol{r})$ *locally* with the two components $v(\boldsymbol{r}, \pm)$ of the external potential. For a Coulomb system, however, such a scheme would completely fail because of the long-range character of the Coulomb interaction. (It could be tried for systems with short-range interactions.)

For a Coulomb system, the simplest non-local term to be singled out of the density functional $H[n]$ is the classical Coulomb energy: the Hartree term containing the self-interaction. (It is written as a nonlocal density functional here, but it can likewise be written as an orbital expression and hence it fits into the general scheme (6.51)–(6.53); $E_H[n]$ behaves indifferent between (6.52) and (6.53).) Hence, the first generation of approximation for the density functional would read

$$\begin{aligned} H[n] \approx H^{(1)}[n] \;&=\; \frac{1}{2}\int d^3r d^3r'\, n(\boldsymbol{r}')w(|\boldsymbol{r}' - \boldsymbol{r}|)n(\boldsymbol{r}) \;+ \\ &+ \int d^3r\, n(\boldsymbol{r})\,(\; t(n(\boldsymbol{r}), \zeta(\boldsymbol{r})) + \varepsilon_{\mathrm{XC}}(n(\boldsymbol{r}), \zeta(\boldsymbol{r}))\;)\,. \end{aligned} \tag{7.42}$$

The remaining LDA integral (second line) contains the kinetic and the exchange and correlation energy functionals. This generation replaces the

Kohn-Sham equations by the Thomas-Fermi equations. See (3.3), where the non-polarized case $\zeta = 0$ was considered and the XC-term was neglected, but an exchange-only term was introduced in (3.46). Due to the non-local Hartree-term of the density functional, the spin density is now connected with the external potential via an integral equation of the type (3.47):

$$\frac{5}{3} C_{\mathrm{F}}\, n^{2/3}(\boldsymbol{r})(1 \pm \zeta(\boldsymbol{r}))^{2/3} = [\mu - v(\boldsymbol{r}, \pm) - v_{\mathrm{H}}(\boldsymbol{r}) - v_{\mathrm{XC}}(\boldsymbol{r}, \pm)]_{+}, \quad (7.43)$$

where $n(\boldsymbol{r})$ enters the integral expression (3.13) of the Hartree potential $v_{\mathrm{H}}(\boldsymbol{r})$. Although this generation accounts for the long-range Coulomb effects, it suffers from many deficiencies due to the neglect of non-localities of the kinetic energy functional. These deficiencies were discussed in Chapter 3.

The next, the second generation consequently replaces the kinetic energy functional by the non-local implicit variational expression (4.57), but retains the local approximation for the exchange and correlation energy functional:

$$\begin{aligned} H[n] \approx H^{(2)}[n] \;&=\; T_{\mathrm{DM}}[n] \;+\; \frac{1}{2}\int d^3r d^3r'\, n(\boldsymbol{r}')w(|\boldsymbol{r}' - \boldsymbol{r}|)n(\boldsymbol{r}) \;+ \\ &+ \int d^3r\, n(\boldsymbol{r})\varepsilon_{\mathrm{XC}}(n(\boldsymbol{r}), \zeta(\boldsymbol{r})). \end{aligned} \quad (7.44)$$

Introducing an orbital variational expression like (4.57) as a density functional, as finally pointed out by [Levy, 1982], is the proper way leading to the ordinary Kohn-Sham equations (6.45) having the form of effective single-particle Schrödinger equations. This second generation implies the local density approximation for exchange and correlation as discussed in the previous section.

As a next step, to avoid self-interaction problems, one could combine the Hartree energy with the exchange energy into an ensemble state Hartree-Fock type functional in accord with (6.52):

$$\begin{aligned} K^{\mathrm{HF}}[n] \stackrel{\mathrm{def}}{=} \min \Bigg\{ \sum_i n_i \int dx\, \phi_i^* \left(-\frac{\nabla^2}{2}\right) \phi_i \;+\; \frac{1}{2}\sum_{ij} n_i n_j \, * \\ * \int dx dx'\, \phi_i^*(x)\phi_j^*(x') w(|\boldsymbol{r}' - \boldsymbol{r}|)\, (\phi_j(x')\phi_i(x) - \phi_i(x')\phi_j(x)) \;\Bigg| \\ \Bigg| \sum_i \phi_i n_i \phi_i^* = n,\; 0 \le n_i \le 1 \Bigg\}. \end{aligned} \quad (7.45)$$

Of course, this corresponds to the ordinary Hartree-Fock expression only, if the density matrix comes from a pure-state single-determinant or from an open shell pure state composed by linear-combining a finite number of determinants. The main advantage of this expression is that contrary to the Hartree functional $E_{\mathrm{H}}[n]$ (first expression of (7.42)) it is exactly self-interaction free. $K^{\mathrm{HF}}[n]$ as defined above is a density functional. Hence, the difference between $H[n]$ and this expression is again a density functional, which we call a correlation energy functional and for which we introduce an LDA expression by writing

$$H[n] \approx H^{(3)}[n] = K^{\mathrm{HF}}[n] + \int d^3r\, n(\boldsymbol{r})\varepsilon_{\mathrm{C}}(n(\boldsymbol{r}), \zeta(\boldsymbol{r})). \tag{7.46}$$

Given a suitable function ε_{C} for the homogeneous case, the variation analogous to (4.26) and (4.60) would lead to Hartree-Fock like equations

$$\left(-\frac{\nabla^2}{2} + \hat{v}_{\mathrm{eff}}(x)\right)\phi_i(x) = \phi_i(x)\,\varepsilon_i, \quad \hat{v}_{\mathrm{eff}} \stackrel{\mathrm{def}}{=} v + v_{\mathrm{H}} + \hat{v}_{\mathrm{X}} + v_{\mathrm{C}} \tag{7.47}$$

with the Hartree potential v_{H} defined as previously, and

$$(\hat{v}_{\mathrm{X}}\,\phi_k)(x) = -\sum_j n_{N;j} \int dx'\, \phi_j^*(x') w(|\boldsymbol{r}' - \boldsymbol{r}|)\phi_k(x')\phi_j(x), \tag{7.48}$$

$$v_{\mathrm{C}}(\boldsymbol{r}, \pm) = \varepsilon_{\mathrm{C}} + n\frac{\partial \varepsilon_{\mathrm{C}}}{\partial n} \pm (1 \mp \zeta)\frac{\partial \varepsilon_{\mathrm{C}}}{\partial \zeta}. \tag{7.49}$$

The density itself would be given as in (6.47–6.49).

If ε_{C} is neglected, this generation reduces to the ordinary Hartree-Fock theory, in the same manner as, if $\varepsilon_{\mathrm{XC}}$ is neglected, the first generation (7.42, 7.43) reduces to the ordinary Thomas-Fermi theory. Note, however, that with the correlation term preserved, the right-hand side of (7.46) does not yield the splitting into Hartree-Fock and correlation energies commonly used in connection with the ordinary Hartree-Fock theory in the literature, where the Hartree-Fock part is defined as the total energy of the Hartree-Fock solution and the correlation part as the difference to the exact total energy. This must be kept in mind when comparing numerical results. The point is that *the orbitals involved are different* in both approaches. Likewise, the first term of (7.45) does not yield the T-functional (4.57), since the minimizing orbitals are different. *The splitting into energy contributions in density functional theory is always different from the corresponding splitting in other approaches; only the total energy itself may be compared.*

Another comment concerns the appearance of an exchange potential operator $\hat{v}_{\mathrm{X}}$ in equation (7.47). This potential is non-local (integral operator) but not orbital dependent. (The integral kernel is the same for all orbitals.) Hence there is no problem with choosing the orbital solutions of (7.47) orthonormal to each other. Although the exact theory of Section 6.4 resulted in a local effective potential (cf. (6.45), $v_{\mathrm{eff}} \in X^*$), for the *different* approach $H[n] = K_{\mathrm{HF}}[n] + E_{\mathrm{C}}[n]$ with a potentially also exact $E_{\mathrm{C}}[n]$, the orbital variation (7.45) introduces a non-local potential operator in the same manner as the orbital variation (4.57) earlier introduced a non-local kinetic energy operator.

Of course, there must be a drawback to this third generation of Kohn-Sham type theory, otherwise it would be much more widely used. The main drawback comes from the strong cancellation between exchange and correlation energy contributions at short distances. (Exchange and Coulomb correlations both keep particles apart from each other, hence exchange partly replaces classical correlation.) As can be seen from (7.17) and (7.11), $\varepsilon_{\mathrm{XC}}$ is generally smaller than ε_{X} (for $r_s \approx 5$ by a factor of two), and, less obviously but more importantly, $v_{\mathrm{XC}}[n]$ is in a more local manner depending on n than is $v_{\mathrm{C}}[n]$ (cf. the discussion in [Hedin and Lundqvist, 1971]). This can be understood on the basis of the discussion of the pair correlation function $g(r)$ in Section 2.5, stating that $g(r)$ resulting from exchange *and* correlation is more short-range than that of exchange or correlation only.

Self-interaction corrected density functionals and the so-called LDA+U approach (the latter if properly introduced) are within this frame in between $H^{(2)}[n]$ and $H^{(3)}[n]$. Another approach related to the content of this section is the use of the so-called optimized effective potential method within density functional theory, cf. [Grabo and Gross, 1995].

7.4 The Self-Interaction Correction

A particular situation where the LDA of Section 7.2 fails is a situation with spatially well separated particles.

Recall that for a ground state density $n(x)$ corresponding to an external potential $v(x)$ the total ground state energy is given by

$$E = T[n] + E_{\mathrm{H}}[n] + E_{\mathrm{XC}}[n] + \int dx\, n(x) v(x). \tag{7.50}$$

In Section 2.5 we already discussed situations where E_{XC} is of a purely formal nature. Consider a single hydrogen atom and assume the spin of the electron

in upward direction. Then

$$n(\boldsymbol{r},+) = |\phi(\boldsymbol{r})|^2, \quad n(\boldsymbol{r},-) = 0. \tag{7.51}$$

The exact relation

$$E_{\mathrm{XC}}[n = |\phi|^2, \zeta = 1] = -E_{\mathrm{H}}[|\phi|^2] = -\frac{1}{2}\int d^3r d^3r' \, \frac{|\phi(\boldsymbol{r})|^2 |\phi(\boldsymbol{r}')|^2}{|\boldsymbol{r}-\boldsymbol{r}'|} \tag{7.52}$$

is obtained from a Kohn-Sham exchange and correlation hole

$$h_{\mathrm{KS}}(\boldsymbol{r}',\boldsymbol{r}) = -|\phi(\boldsymbol{r}')|^2, \tag{7.53}$$

which is independent of the particle position $\boldsymbol{r}$. The LDA exchange and correlation hole, on the other hand, is equal to

$$h_{\mathrm{KS}}^{\mathrm{LDA}}(\boldsymbol{r}',\boldsymbol{r}) = h_{\mathrm{hom}}(|\boldsymbol{r}'-\boldsymbol{r}|, n = |\phi|^2, \zeta = 1), \tag{7.54}$$

which leads to

$$E_{\mathrm{XC}}^{\mathrm{LDA}}[n = |\phi|^2, \zeta = 1] \neq -E_{\mathrm{H}}[|\phi|^2] \tag{7.55}$$

and causes an error of the LDA of about 5 percent of the hydrogen energy.

If two hydrogen atoms are at sufficiently separated positions $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ so that their electron clouds do not overlap, then the Hartree energy consists of three contributions: the physically real Coulomb interaction of the electron clouds around atom 1 and 2, respectively, which screens the proton potentials, and the two self-interaction energies of these clouds. The latter again must be canceled by E_{XC} which is not provided by the LDA.

To remedy this defect at least approximately, one replaces the orbital variation expression (6.52) for the Hartree energy functional by the self-interaction corrected (SIC) functional [Perdew and Zunger, 1981], see also [Svane, 1995]

$$\begin{aligned} K^{\mathrm{SIC}}[n] \stackrel{\mathrm{def}}{=} \min \Bigg\{ & \sum_i n_i \int dx \, \phi_i^* \left(-\frac{\nabla^2}{2}\right) \phi_i + \\ & + \sum_{i \neq j} \frac{n_i n_j}{2} \int dx dx' \, |\phi_i(x)|^2 w(|\boldsymbol{r}-\boldsymbol{r}'|) |\phi_j(x')|^2 \; - \\ & - \sum_i n_i E_{\mathrm{XC}}^{\mathrm{LDA}}[n = |\phi_i|^2, \zeta = 1] \;\Bigg|\; \sum_i \phi_i n_i \phi_i^* = n, \; 0 \leq n_i \leq 1 \Bigg\}. \end{aligned} \tag{7.56}$$

The basic density functional $H[n]$ is then approximated by

$$H[n] \approx H^{\mathrm{SIC}}[n] \stackrel{\mathrm{def}}{=} K^{\mathrm{SIC}}[n] + E_{\mathrm{XC}}^{\mathrm{LDA}}[n]. \tag{7.57}$$

In the cases considered above this expression just extracts the self-interaction part from the last contribution and puts it together with the self-interaction corrected Hartree term into the orbital variation term $K^{\mathrm{SIC}}[n]$. If on the other hand the Kohn-Sham orbitals of an N-particle system are extended over a volume $\sim N$, one has $|\phi_i|^2 \sim 1/N$, and, since both E_{H} and E_{XC} depend on the density in higher than first order, the sum over the N terms in braces of (7.56) tends to zero for $N \to \infty$. For extended states of an extended system the self-interaction correction vanishes.

Use of this density functional in the Hohenberg-Kohn-Sham variation leads to a Kohn-Sham like equation

$$\left[-\frac{\nabla^2}{2} + v(x) + v_{\mathrm{H}}(\boldsymbol{r}) + v_{\mathrm{XC},i}^{\mathrm{SIC}}(x)\right] \phi_i(x) = \phi_i(x)\varepsilon_i, \tag{7.58}$$

where v_{H} is the Hartree potential as previously including the self-interaction (but therefore orbital independent). In contrast to the ordinary Kohn-Sham equation (6.45) it contains an orbital-dependent potential term in the exchange and correlation potential,

$$\begin{aligned} v_{\mathrm{XC},i}^{\mathrm{SIC}}(x) &= v_{\mathrm{XC}}^{\mathrm{LDA}}[n,\zeta](x) - \\ &\quad -n_{N;i}\left(\int d^3r' \, \frac{|\phi_i(\boldsymbol{r}')|^2}{|\boldsymbol{r}-\boldsymbol{r}'|} - v_{\mathrm{XC}}^{\mathrm{LDA}}[|\phi_i(x)|^2, 1](x)\right). \end{aligned} \tag{7.59}$$

For a discussion of the consequences of the non-orthogonality of the orbitals resulting from (7.59), which is usually small and was neglected in (7.58), and for further discussion see [Perdew and Zunger, 1981].

The self-interaction correction leads to substantial improvements in many applications. Obviously, (7.57) is to be placed somewhere between the second and the third generation in the context of last section. This scheme would be consistent, if the solutions $\phi_i(x)$ of (7.59) would come out uniquely and orthogonal to each other. This self-interaction correction cannot, of course, remedy the defect discussed at the end of Section 7.2, because that defect is present in delocalized situations, where self-interaction corrections disappear. Gap corrections need another essentially nonlocal contribution to $E_{\mathrm{XC}}[n]$ to be explicitly introduced, which perhaps might be found along similar lines, such as, for example, a screened exchange term. (Cf. [Seidl *et al.*, 1996].)

7.5 The LDA+U Approach

The models for the density functional considered so far were all free of adjustable parameters. This is an important feature which bears the potential of broad applicability and predictive power. Nevertheless, there are areas of very successful application in the sense of producing results in good agreement with phenomenology (of course provided the numerics is accurate enough for such issues; this is unfortunately a weak point of quite a number of publications even today), and there are cases of failure. To the latter belong so called correlation insulators or Mott insulators, solids which would be metals in a mean-field treatment of the crystal potential but which develop an excitation gap for particle-hole excitations due to strong local electron-electron correlation. This case, in particular in the strong correlation limit, is with certain success treated with the LDA+U approach which has seemingly more arbitrariness in modeling and is at least up to now often not free of adjustable parameters, yet it fits perfectly in the above considered scheme and hence it is shortly presented here. For recent surveys see [Anisimov *et al.*, 1997, Eschrig *et al.*, 2003a]. As was stated earlier, we are considering the spin density theory only in this chapter. What we call LDA here is in the literature often called LSDA to make the difference to the spin-independent case explicit.

The first step is to specify local strongly correlated l_c-shells (typically d or f-shells of orbitals $|\boldsymbol{R}m\sigma)$ centered at lattice sites $\boldsymbol{R}$ and chosen as angular momentum eigenstates with azimuth quantum numbers $m = -l_c, \ldots, l_c$ and spin quantum number $\sigma = +, -$ with respect to a chosen and possibly site dependent spin quantization axis while the independent choice of the orbital quantization axis is irrelevant as the so-called rotational invariant treatment of LDA+U will be presented here. It is important to distinguish these correlated model orbitals from the Kohn-Sham orbitals. At center $\boldsymbol{R}$, the screened on-site Coulomb matrix elements of the correlated orbitals are

$$(m_1 m_2|\tilde{w}|m_3 m_4), \quad \tilde{w} \approx \tilde{w}(|\boldsymbol{r} - \boldsymbol{r}'|), \quad \sigma_1 = \sigma_3,\, \sigma_2 = \sigma_4, \tag{7.60}$$

where the screening depends on the orbital occupation but is assumed (as a reasonable approximation) rotational invariant. As a consequence, the SO_3 transformation properties of the matrix elements are

$$\begin{aligned}(m_1 m_2|\tilde{w}|m_3 m_4) = \sum_{m_1' m_2' m_3' m_4'} & U^{\dagger}_{m_1 m_1'}(\hat{O}) U^{\dagger}_{m_2 m_2'}(\hat{O}) * \\ & * (m_1' m_2'|\tilde{w}|m_3' m_4') U_{m_3' m_3}(\hat{O}) U_{m_4' m_4}(\hat{O}),\end{aligned} \tag{7.61}$$

where $\hat{O}$ is any rotation of the $\boldsymbol{r}$-space and the U-matrices yield the relevant SO_3 representation:

$$\begin{aligned} U^\dagger(\hat{O})U(\hat{O}) &= 1 = U(\hat{O})U^\dagger(\hat{O}), \\ \int d\hat{O}\, U_{m_1m_2}(\hat{O})U^\dagger_{m_3m_4}(\hat{O}) &= \frac{1}{2l+1}\,\delta_{m_1m_4}\delta_{m_2m_3}. \end{aligned} \tag{7.62}$$

In the last orthogonality relation, $d\hat{O}$ is Haar's invariant measure in the parameter space of the SO_3, $\int d\hat{O} = 1$.

From these fundamental representation properties the important sum rules for the matrix elements follow: Use unitarity of $U(\hat{O})$ and integrate over $d\hat{O}$ (which leaves the l.h.s. unchanged as it is independent of $\hat{O}$ and $\int d\hat{O} = 1$) to obtain

$$\begin{aligned} &\sum_{m_1}(m_1m_2|\tilde{w}|m_1m_4) = \\ &= \sum_{m_1}\int d\hat{O} \sum_{m_1'm_2'm_3'm_4'} U^\dagger_{m_1m_1'}(\hat{O})U^\dagger_{m_2m_2'}(\hat{O})\ (m_1'm_2'|\tilde{w}|m_3'm_4')\cdot \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \cdot U_{m_3'm_1}(\hat{O})U_{m_4'm_4}(\hat{O}) = \\ &= \int d\hat{O} \sum_{m_1'm_2'm_3'm_4'} \delta_{m_1'm_3'}U^\dagger_{m_2m_2'}(\hat{O})(m_1'm_2'|\tilde{w}|m_3'm_4')U_{m_4'm_4}(\hat{O}) = \\ &= \frac{1}{2l+1}\sum_{m_1'm_2'm_4'}(m_1'm_2'|\tilde{w}|m_1'm_4')\delta_{m_2m_4}\delta_{m_2'm_4'} = \\ &= \frac{\delta_{m_2m_4}}{2l+1}\sum_{m_1'm_2'}(m_1'm_2'|\tilde{w}|m_1'm_2')\ = \delta_{m_2m_4}(2l+1)\ U. \end{aligned} \tag{7.63}$$

The last equation is the definition of the Coulomb integral U (not to be confused with the transformation $U(\hat{O})$). In the same manner,

$$\begin{aligned} &\sum_{m_1}(m_1m_2|\tilde{w}|m_3m_1) = \\ &= \frac{\delta_{m_2m_3}}{2l+1}\sum_{m_1'm_2'}(m_1'm_2'|\tilde{w}|m_2'm_1')\ = \delta_{m_2m_3}(U+2lJ) \end{aligned} \tag{7.64}$$

is obtained which additionally defines the exchange integral J. The first result (7.63) is intuitively obvious: after summation over m_1 and integration over $\boldsymbol{r}$ in the matrix element, no angular dependence is left except the orthogonality $(m_2|m_4) = \delta_{m_2m_4}$ since no direction is any more distinguished. The second result (7.64) is less obvious but nevertheless true.

Expansion of the interaction function into spherical harmonics,

$$\begin{aligned}\tilde{w}\big(|\boldsymbol{r}_1-\boldsymbol{r}_2|\big) &= \tilde{w}\Big((r_1^2+r_2^2-2r_1r_2\cos\theta)^{1/2}\Big) = \\ &= \sum_{l=0}^{\infty}\tilde{w}_l(r_1,r_2)\,P_l(\cos\theta) = \sum_{l=0}^{\infty}\tilde{w}_l(r_1,r_2)\frac{4\pi}{2l+1}\sum_{m=-l}^{l}Y_{lm}(\hat{\boldsymbol{r}}_1)Y_{lm}^*(\hat{\boldsymbol{r}}_2)\end{aligned} \tag{7.65}$$

leads to Slater's analysis

$$\begin{aligned}(m_1m_2|\tilde{w}|m_3m_4) &= \sum_{l=0}^{2l_i}\tilde{F}_l\,a_l(m_1m_2m_3m_4),\\ \tilde{F}_l &= \iint_0^{\infty}dr_1dr_2\,\big(r_1R_c(r_1)\big)^2\big(r_2R_c(r_2)\big)^2\tilde{w}_l(r_1,r_2)\\ &\approx \iint_0^{\infty}dr_1dr_2\,\big(r_1R_c(r_1)\big)^2\big(r_2R_c(r_2)\big)^2\frac{r_<^l}{r_>^{l+1}}\ \text{for } l>0,\\ a_l(m_1m_2m_3m_4) &= \frac{4\pi}{2l+1}\sum_{m=-l}^{l}(Y_{l_cm_1}|Y_{lm}|Y_{l_cm_3})(Y_{l_cm_4}|Y_{lm}|Y_{l_cm_2})^*.\end{aligned} \tag{7.66}$$

Here, R_c is the radial part of the correlated orbitals, and the second line for $\tilde{F}_l$ holds for the unscreened Coulomb interaction which for $l>0$ is a reasonable approximation since intraatomic screening is effective only for the s-component of the interaction.

Now, from $\sum_m Y_{lm}(\boldsymbol{r})Y_{lm}^*(\boldsymbol{r}) = P_l(1)(2l+1)/4\pi$ and

$$\begin{aligned}\sum_{m_1}a_l(m_1m_2m_1m_2) = &= \frac{4\pi}{2l+1}\left[\sum_{m_1}(Y_{l_cm_1}|Y_{l0}|Y_{l_cm_1})\right](Y_{l_cm_2}|Y_{l0}|Y_{l_cm_2})^* = \\ &= \sqrt{4\pi}\frac{2l_c+1}{2l+1}\,\delta_{l0}\,(Y_{l_cm_2}|Y_{l0}|Y_{l_cm_2})^* = (2l_c+1)\delta_{l0}\end{aligned}$$

it follows immediately that

$$U = \tilde{F}_0. \tag{7.67}$$

Furthermore,

$$\begin{aligned}
\sum_{m_1m_2} a_l(m_1m_2m_2m_1) &= \frac{4\pi}{2l+1}\sum_{\substack{m_1m_2\\ m}}(Y_{l_cm_1}|Y_{lm}|Y_{l_cm_2})(Y_{l_cm_2}|Y^*_{lm}|Y_{l_cm_1}) = \\
&= \frac{4\pi}{2l+1}\iint d\Omega_1 d\Omega_2\Big(\sum_{m_1} Y_{l_cm_1}(\boldsymbol{r}_2)Y^*_{l_cm_1}(\boldsymbol{r}_1)\Big)* \\
&* \Big(\sum_{m_2} Y_{l_cm_2}(\boldsymbol{r}_1)Y^*_{l_cm_2}(\boldsymbol{r}_2)\Big)\Big(\sum_{m} Y_{lm}(\boldsymbol{r}_1)Y^*_{lm}(\boldsymbol{r}_2)\Big) = \\
&= \frac{(2l_c+1)^2}{(4\pi)^2}\iint d\Omega_1 d\Omega_2\big[P_{l_c}(\cos\theta_{12})\big]^2 P_l(\cos\theta_{12}) = \\
&= \frac{(2l_c+1)^2}{4\pi}\int d\Omega\big[P_{l_c}(\cos\theta)\big]^2 P_l(\cos\theta) = \\
&= (2l_c+1)^2\begin{pmatrix} l_c & l & l_c \\ 0 & 0 & 0\end{pmatrix}^2
\end{aligned}$$

and hence

$$\begin{aligned}
\sum_{m_1m_2}(m_1m_2|\tilde{w}|m_2m_1) &= (2l_c+1)^2\sum_{l=0}^{2l_c}\tilde{F}_l\begin{pmatrix} l_c & l & l_c \\ 0 & 0 & 0\end{pmatrix}^2 = \\
&= (2l_c+1)(U+2l_cJ).
\end{aligned} \tag{7.68}$$

Eqs. (7.67) and (7.68) relate the Coulomb and exchange integrals U and J to Slater's (screened) integrals $\tilde{F}_l$. U and J are often treated as parameters; a correct approach would try to calculate independently the $\tilde{F}_l$s.

Next, the occupation matrix $\tilde{n}$ of the correlated shell is to be determined with the Kohn-Sham orbitals ϕ_i and their occupation numbers n_i:

$$\tilde{n}_{mm'\sigma} \stackrel{\text{def}}{=} \sum_i(\boldsymbol{R}m\sigma|\phi_i\rangle n_i\langle\phi_i|\boldsymbol{R}m'\sigma). \tag{7.69}$$

Here on the l.h.s. and in the following we drop the site vector $\boldsymbol{R}$ in order not to overload the notation. At each site (and possibly for each spin direction independently) the occupation matrix may be m-diagonalized by introducing local orbital coordinates:

$$\tilde{n}_{mm'\sigma} = \tilde{U}^{(\sigma)}_{m\mu}\,\tilde{n}_{\mu\sigma}\,\tilde{U}^{(\sigma)*}_{m'\mu}. \tag{7.70}$$

Averages over a correlated shell,

$$\tilde{n}_\sigma = \frac{1}{2l_c+1}\sum_\mu \tilde{n}_{\mu\sigma}, \quad \tilde{n} = \frac{1}{2}\big(\tilde{n}_+ + \tilde{n}_-\big), \tag{7.71}$$

are used later on.

The LDA+U density functional is obtained by taking $k[\phi_i, n_i]$ in (6.52) in the form

$$k = t + e^{\mathrm{H}} + e^{U} \tag{7.72}$$

with

$$t + e^{\mathrm{H}} = \sum_i n_i \langle \phi_i | \hat{t} | \phi_i \rangle + \sum_{ij} \frac{n_i n_j}{2} \langle \phi_i \phi_j | \frac{1}{|\boldsymbol{r}' - \boldsymbol{r}|} | \phi_i \phi_j \rangle \tag{7.73}$$

as in the LDA. There have essentially two versions of e^U been introduced [Czyżyk and Sawatzky, 1994]: the 'around the mean field' version (AMF),

$$\begin{aligned} e^{U,\mathrm{AMF}} &= \frac{1}{2} \sum_{\boldsymbol{R}\sigma\mu\mu'} \Big\{ (\mu_\sigma \mu'_{-\sigma} | \tilde{w} | \mu_\sigma \mu'_{-\sigma}) (\tilde{n}_{\mu\sigma} - \tilde{n}_\sigma)(\tilde{n}_{\mu'-\sigma} - \tilde{n}_{-\sigma}) + \\ &\quad + \big[(\mu_\sigma \mu'_\sigma | \tilde{w} | \mu_\sigma \mu'_\sigma) - (\mu_\sigma \mu'_\sigma | \tilde{w} | \mu'_\sigma \mu_\sigma) \big] (\tilde{n}_{\mu\sigma} - \tilde{n}_\sigma)(\tilde{n}_{\mu'\sigma} - \tilde{n}_\sigma) \Big\} \\ &= \frac{1}{2} \sum_{\boldsymbol{R}\sigma\mu\mu'} \Big\{ (\mu_\sigma \mu'_{-\sigma} | \tilde{w} | \mu_\sigma \mu'_{-\sigma})\ \tilde{n}_{\mu\sigma} \tilde{n}_{\mu'-\sigma} + \\ &\qquad + \big[(\mu_\sigma \mu'_\sigma | \tilde{w} | \mu_\sigma \mu'_\sigma) - (\mu_\sigma \mu'_\sigma | \tilde{w} | \mu'_\sigma \mu_\sigma) \big]\ \tilde{n}_{\mu\sigma} \tilde{n}_{\mu\sigma} \Big\} \\ &\quad - \frac{1}{2} \sum_{\boldsymbol{R}\sigma} \Big\{ U\big(N - \tilde{n}_\sigma\big) - J\Big(N_\sigma - \tilde{n}_\sigma\Big) \Big\} N_\sigma, \\ N_\sigma &= \sum_\mu \tilde{n}_{\mu\sigma} = (2l+1)\tilde{n}_\sigma, \end{aligned} \tag{7.74}$$

and the 'atomic limit' version (AL),

$$\begin{aligned} e^{U,\mathrm{AL}} &= \frac{1}{2} \sum_{\boldsymbol{R}\sigma\mu\mu'} \Big\{ (\mu_\sigma \mu'_{-\sigma} | \tilde{w} | \mu_\sigma \mu'_{-\sigma})\ \tilde{n}_{\mu\sigma} \tilde{n}_{\mu'-\sigma} + \\ &\qquad + \big[(\mu_\sigma \mu'_\sigma | \tilde{w} | \mu_\sigma \mu'_\sigma) - (\mu_\sigma \mu'_\sigma | \tilde{w} | \mu'_\sigma \mu_\sigma) \big]\ \tilde{n}_{\mu\sigma} \tilde{n}_{\mu\sigma} \Big\} \\ &\quad - \frac{1}{2} \sum_{\boldsymbol{R}} \Big\{ UN\big(N-1\big) - J \sum_\sigma N_\sigma \Big(N_\sigma - 1\Big) \Big\}, = \\ &= e^{U,\mathrm{AMF}} + \frac{1}{2} \sum_{\boldsymbol{R}\sigma} (U - J)(1 - \tilde{n}_\sigma) N_\sigma. \end{aligned} \tag{7.75}$$

In (7.74), in the second equality the sum rules (7.63, 7.64) have been used for a better comparison of both versions. N is the number of electrons occupying the whole correlated shell, N_σ is that for one spin sort.

The first two lines of (7.75) and of the second form of (7.74) can be considered as the mean-field expectation value of a Hartree-Fock operator of type (1.74); the third line contains the 'double counting correction', it subtracts from the L-part (6.53), for which the LDA expressions of Section 7.2 are used, what is believed to be contained there already of the contributions to the first two lines. In (7.74) it is assumed that the LDA works already well if there is no orbital polarization in the correlated shell, that is, if $\tilde{n}_{\mu\sigma} = \tilde{n}_\sigma$. It is most easily seen from the first form of (7.74) that it vanishes in this case. Therefore the AMF version may also be called an orbital polarization functional because it corrects the LDA for orbital polarization of the correlated shell which is assumed to be better represented by the Hartree-Fock like expression. The AL version is even more *ad hoc*.

For the derivation of Kohn-Sham equations it is essential that in the spirit of (6.51–6.53) the variational quantities are as previously the Kohn-Sham orbitals ϕ_i and their occupation numbers n_i. The orbitals $|\boldsymbol{R}m\sigma)$ of the correlated shells and the corresponding values of $\tilde{F}_l$ or U and J are *model quantities* but not variational. (They are supposed to model $H[n]$ through (6.51–6.53) in the vicinity of certain densities n. In this sense, those model quantities may be different in different vicinities in the functional n-space, but they all are supposed to approximate locally in the functional n-space the same universal functional $H[n]$.) The e^U-functional now leads according to (6.55) to an orbital-dependent potential operator acting on ϕ_i,

$$\frac{1}{n_i}\frac{\delta}{\delta\phi_i^*}e_U = \sum_{\boldsymbol{R}\mu\sigma}\frac{\partial e^U}{\partial\tilde{n}_{\mu\sigma}}\frac{1}{n_i}\frac{\delta\tilde{n}_{\mu\sigma}}{\delta\phi_i^*} = \sum_{\boldsymbol{R}\mu\sigma}|\boldsymbol{R}\mu\sigma)\frac{\partial e^U}{\partial\tilde{n}_{\mu\sigma}}(\boldsymbol{R}\mu\sigma|\phi_i\rangle. \tag{7.76}$$

The U-potentials of the two considered versions are

$$\begin{aligned}\frac{\partial e^{U,\mathrm{AMF}}}{\partial\tilde{n}_{\mu\sigma}} = \sum_{\mu'}\Big\{&(\mu_\sigma\mu'_{-\sigma}|\tilde{w}|\mu_\sigma\mu'_{-\sigma})(\tilde{n}_{\mu'-\sigma}-\tilde{n}_{-\sigma})+\\ &+\big[(\mu_\sigma\mu'_\sigma|\tilde{w}|\mu_\sigma\mu'_\sigma)-(\mu_\sigma\mu'_\sigma|\tilde{w}|\mu'_\sigma\mu_\sigma)\big](\tilde{n}_{\mu'\sigma}-\tilde{n}_\sigma)\Big\}\end{aligned} \tag{7.77}$$

and

$$\frac{\partial e^{U,\mathrm{AL}}}{\partial\tilde{n}_{\mu\sigma}} = \frac{\partial e^{U,\mathrm{AMF}}}{\partial\tilde{n}_{\mu\sigma}} - (U-J)\Big(\tilde{n}_\sigma - \frac{1}{2}\Big). \tag{7.78}$$

One characteristic feature of this latter U-potential is that in case of an isolated shell it moves the occupied states downward by $(U-J)/2$ and the

unoccupied states upward by $(U-J)/2$ independent of the shell occupation. By way of contrast, the center of the AMF spin subshell potential split moves up with increasing subshell occupation.

There are other versions of LDA+U functionals possible. Finding the best LDA+U model for $H[n]$ and the best parameterization of those models is presently an active field of exploration. It should be noticed that the so-called LDA+DMFT approach (DMFT stands for dynamical mean field theory) is an approach to the self-energy of the single-particle Green function, and, though it uses similar ways of modeling, logically it belongs to another although related context.

Part II:
RELATIVISTIC THEORY

8 A Brief Introduction to Quantum Electrodynamics

Due to the zero (or at least up to now immeasurably small) rest mass of the photon, electrodynamics, the theory of time-dependent electromagnetic fields, is an intrinsically relativistic theory and, as is well known, had served as the prototype of relativistic physics. Quantum electrodynamics (QED) is the relativistic quantum theory of electrically charged particles, mutually interacting by exchange of photons. If this system is additionally subject to a (static) external field, an inhomogeneous situation is obtained to which density functional theory may apply. The standard situation is again that of electrons (and possibly appearing positrons) moving in the adiabatic field produced by nuclei in atoms, molecules or solids. Since we will focus on that situation, in our context the particle field will exclusively be the electron-positron field.

The magnitude of relativistic effects in atoms is easily estimated: the total energy of all electrons of an atom with nuclear charge Z is of the order Z^2 (natural atomic units used). It is essentially the energy of the electrons in the $1s$-orbital which is large compared to that of higher orbitals. In non-relativistic treatment a kinetic energy of the same order of magnitude is obtained from the virial theorem. The kinetic energy per electron is hence of the order of Z. Since the electron mass is equal to unity in our natural unit system, the average electron velocity is of the order $\sqrt{Z}$. The velocity of light in natural atomic units is $c \approx 137$. (It is equal to the inverse of the fine structure constant.) Hence the average relativistic mass correction is $\Delta m/m \sim (v/c)^2 \sim Z/137^2$. Since the average kinetic energy per electron was $\sim Z$, the average energy correction per electron is $\Delta\varepsilon \sim (Z/137)^2$Hartree. This amounts to about 1meV for hydrogen, about 1eV for silver, say, and up to 10eV for actinides.

To prepare for a relativistic density functional theory, in this chapter we browse through quantum electrodynamics, just to recall some basic ideas, notions, and relations necessary for our goal. We start with classical electrodynamics, introduce the covariant notation, the Lagrange formalism, relativistic mechanics, and finally introduce the quantized photon field, the Dirac field, and their interaction.

8.1 Classical Electrodynamics

Electrodynamics connects the dynamics of electric fields $\boldsymbol{E}$ and magnetic fields $\boldsymbol{H}$ with the dynamics of their sources and vortices: electric charges and currents. In an infinite space causality demands that outgoing waves of electromagnetic field are correlated with the motion of charges and with currents in the past only, and that incoming waves are correlated with the motion of charges and with currents in the future only.

The basic equations of classical electrodynamics are Maxwell's equations:

$$\epsilon_0 \nabla \boldsymbol{E} = -en, \qquad \nabla \times \boldsymbol{H} = \epsilon_0 \dot{\boldsymbol{E}} - e\boldsymbol{j}, \tag{8.1}$$

$$\nabla \times \boldsymbol{E} = -\mu_0 \dot{\boldsymbol{H}}, \qquad \nabla \boldsymbol{H} = 0. \tag{8.2}$$

The equations in the first line identify the electrical charge density $-en$ (n being the particle density of electrons with charge $-e$, minus that of positrons with charge $+e$) to be the source density and the electric current density $-e\boldsymbol{j}$ ($\boldsymbol{j}$ being the particle current of electrons minus that of positrons) together with the 'displacement current' $\epsilon_0 \dot{\boldsymbol{E}}$ to be the vortex density of the electromagnetic field.

Note that although we have four field-creating components $(n, \boldsymbol{j})$, only three of them are independent, since the first two of Maxwell's equations provide charge conservation

$$\dot{n} + \nabla \boldsymbol{j} = 0 \tag{8.3}$$

and hence determine only three independent components of the electromagnetic field. The remaining three field components are determined by the second group (8.2) of Maxwell's equations, of which again only three are independent because the right one ensures that both sides of the left one have zero divergence.

Besides the charge quantum e, the vacuum permittivity ϵ_0 and the vacuum permeability μ_0 appear as coefficients in Maxwell's equations. Causality adds boundary conditions for the fields to those equations, but we will not consider them here.

To fulfill the second group (8.2) of Maxwell's equations, electromagnetic potentials $(U, \boldsymbol{A})$ are introduced for the fields according to

$$\mu_0 \boldsymbol{H} = \nabla \times \boldsymbol{A}, \qquad \boldsymbol{E} = -\dot{\boldsymbol{A}} - \nabla U. \tag{8.4}$$

Since four potentials were introduced to fulfill three independent equations, there is still a free choice of one of the potential components. The Lorentz

gauge

$$\frac{1}{c^2}\dot{U} + \nabla \boldsymbol{A} = 0, \qquad \frac{1}{c^2} = \epsilon_0 \mu_0 \tag{8.5}$$

is the most symmetric of all choices, transforming the first group (8.1) of Maxwell's equations into

$$\Box U = -\frac{en}{\epsilon_0}, \quad \Box \boldsymbol{A} = -\mu_0 e \boldsymbol{j}, \quad \Box = \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \nabla^2. \tag{8.6}$$

The d'Alembert operator $\Box$ uncovers the full symmetry of the Maxwell theory.

The connection between the electrodynamic potential and the mechanical Hartree potential is $v_{\mathrm{H}} = -eU$, and that between the electrodynamic charge and the electrostatic coupling constant is $\lambda = e^2/4\pi\epsilon_0$. (In electrostatics one often defines units by putting $\epsilon_0 = 1/4\pi$; this was done in Part I.) Poisson's equation (3.14) is now the static version of the first equation (8.6).

8.2 Lorentz Covariance

D'Alembert's operator suggests Minkowski's geometry of space-time:

$$x^\mu = (ct, x, y, z), \qquad x_\mu = g_{\mu\nu} x^\nu = (ct, -x, -y, -z), \tag{8.7}$$

$$g_{\mu\nu} = \begin{pmatrix} 1 & & & 0 \\ & -1 & & \\ & & -1 & \\ 0 & & & -1 \end{pmatrix} = g^{\mu\nu}, \tag{8.8}$$

$$\partial_\mu = \frac{\partial}{\partial x^\mu}, \quad \partial^\mu = \frac{\partial}{\partial x_\mu}, \quad \partial_\mu \partial^\mu = \Box, \tag{8.9}$$

where summation over repeated indices, one upper and one lower, (tensor contraction) is understood. Besides the fundamental or (pseudo-)metric tensor $g_{\mu\nu}$, which determines the indefinite metric (or pseudo-metric) of the Minkowski space, there is a fundamental four-form (Levi-Civita pseudotensor)

$$\epsilon^{\mu\nu\rho\sigma} = \epsilon_{\mu\nu\rho\sigma}, \quad \epsilon^{\mathcal{P}(0123)} = (-1)^{\mathcal{P}}, \ \epsilon^{\mu\nu\rho\sigma} = 0 \text{ otherwise}, \tag{8.10}$$

where $\mathcal{P}$ means permutation. It determines the invariant measure (volume) of a parallel epiped spanned by any four four-vectors $(dx^\mu)_i$, $i = 1, \ldots, 4$ or $(dx_\mu)_i$, $i = 1, \ldots, 4$ to be

$$\begin{aligned} dV &= \epsilon_{\mu\nu\rho\sigma}(dx^\mu)_1(dx^\nu)_2(dx^\rho)_3(dx^\sigma)_4 = \\ &= \epsilon^{\mu\nu\rho\sigma}(dx_\mu)_1(dx_\nu)_2(dx_\rho)_3(dx_\sigma)_4 \end{aligned} \tag{8.11}$$

in that affine Minkowski space. (If the dx^μ are differentials of coordinate transformation functions, the above expression for dV produces automatically the corresponding Jacobian.) Every totally antisymmetric rank-4 tensor is proportional to that fundamental four-form. Every totally antisymmetric rank-n tensor (n-form, $n \leq 4$) can be expressed as the contraction of a totally antisymmetric rank-$(4-n)$ tensor with the fundamental four-form.

In this language the four-current density (free of four-sources because of (8.3)) becomes

$$j^\mu = (cn, j_x, j_y, j_z), \qquad \partial_\mu j^\mu = 0, \tag{8.12}$$

and the four-field tensor is

$$F^{\mu\nu} = \begin{pmatrix} 0 & E_x/c & E_y/c & E_z/c \\ -E_x/c & 0 & \mu_0 H_z & -\mu_0 H_y \\ -E_y/c & -\mu_0 H_z & 0 & \mu_0 H_x \\ -E_z/c & \mu_0 H_y & -\mu_0 H_x & 0 \end{pmatrix}, \tag{8.13}$$

that is,

$$F^{0i} = E_i/c, \quad F^{ik} = \mu_0 \epsilon^{ikl} H_l, \tag{8.14}$$

where as in the whole following text Greek super-(sub-)scripts refer to Minkowski's four-quantities and Latin ones to three-dimensional spatial tensors. The fundamental three-form ϵ^{ikl} is defined in complete analogy to (8.10). The antisymmetric four-field tensor is made up of an antisymmetric three-tensor being dual (via ϵ) to the (pseudo-)three-vector $\boldsymbol{H}$ of the magnetic field (axial vector), and framed by the three-vector $\boldsymbol{E}$ of the electric field (polar vector).

Maxwell's equations are cast into

$$\partial_\nu F^{\mu\nu} = -\mu_0 e j^\mu, \qquad \epsilon_{\mu\nu\rho\sigma}\partial^\nu F^{\rho\sigma} = 0. \tag{8.15}$$

The four-potential (four-source free by the Lorentz gauge) is

$$A^\mu = (U/c, A_x, A_y, A_z), \qquad \partial_\mu A^\mu = 0, \tag{8.16}$$

$$F^{\mu\nu} = \partial^\nu A^\mu - \partial^\mu A^\nu. \tag{8.17}$$

Check the equivalence to (8.4).[9]

The whole theory is now covariant under transformations

$$x'^\mu = Q^\mu_\nu x^\nu, \quad Q^\mu_\nu g_{\mu\rho} Q^\rho_\sigma = g_{\nu\sigma}, \tag{8.18}$$

where the right relation is the isometry condition for the transformation Q in Minkowski's indefinite metric. E.g. the special Lorentz transformation is

$$Q^\mu_\nu = \begin{pmatrix} \frac{1}{\sqrt{1-v^2/c^2}} & \frac{-v/c}{\sqrt{1-v^2/c^2}} & 0 & 0 \\ \frac{-v/c}{\sqrt{1-v^2/c^2}} & \frac{1}{\sqrt{1-v^2/c^2}} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{8.19}$$

It connects two inertial systems, one moving with velocity v in x-direction relative to the other. Another special case of (8.18) is an orthogonal transformation of the three-space (rotation by some Euler angles with coordinate origin fixed).

8.3 Lagrange Formalism

A dynamical system in physics is characterized by a Lagrange function L of its dynamical degrees of freedom $f(t)$, so that the dynamics minimizes the action S with $f(t_1)$ and $f(t_2)$ fixed:

$$\delta S = \delta \int_{t_1}^{t_2} dt\, L = 0. \tag{8.20}$$

[9]The topologically interested reader may note that Maxwell's theory is a case of de Rham's cohomology: If ω is a differentiable form (n-form), then the exterior differential $d\omega$ generalizes the rot-operation of three dimensions and creates an $(n+1)$-form (a coboundary to ω). $d^2 = 0$ generalizes rot rot $= 0$. The generalized integral theorem reads $\int_V d\omega = \int_{\partial V} \omega$, where ∂V is the n-dimensional oriented closed boundary ($\partial^2 = 0$, i.e., the closed boundary itself has no boundary) of the $(n+1)$-dimensional manifold V. In this frame, the four-potential A is a one-form and the four-field $F = dA$ is its coboundary. The second group of Maxwell's equations, $dF = 0$, is a simple case of the general rule $d^2 = 0$. More generally, a closed manyfold with vanishing boundary is a cycle, and a form with vanishing coboundary is a cocycle. The second group of Maxwell's equations states that F is a cocycle, and hence, in a manyfold which is homotopically contractible into a point, it may be given as the coboundary of a potential. (Cf. [Schwarz, 1994].)

This variation yields in the well known way equations of motion

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{f}} = \frac{\partial L}{\partial f}. \tag{8.21}$$

(Provided the functional $S[f]$ has a unique functional derivative $S' = \partial L/\partial f - (d/dt)(\partial L/\partial \dot{f})$.)

If the dynamical degrees of freedom are fields $f(\boldsymbol{r})$ in three-space, the action is expressed in terms of a Lagrange density $\mathcal{L}$ according to

$$S = \int_{t_1}^{t_2} dt \int d^3r\, \mathcal{L}, \tag{8.22}$$

and the equations of motion adopt the form

$$\partial_\mu \frac{\partial \mathcal{L}}{\partial(\partial_\mu f)} = \frac{\partial \mathcal{L}}{\partial f}. \tag{8.23}$$

(Cf. e.g. [Itzykson and Zuber, 1980].)

The Lagrange density of a free electromagnetic field, i.e. a field in the absence of sources j^μ, is, up to a constant factor α to be determined at the end of this section,

$$\begin{aligned} \mathcal{L} &= \alpha F^{\mu\nu} F_{\mu\nu} = \\ &= 2\alpha\left[(\partial^\nu A^\mu)(\partial_\nu A_\mu) - (\partial^\mu A^\nu)(\partial_\nu A_\mu)\right] = \\ &= 2\alpha(\partial^\nu A^\mu)(\partial_\nu A_\mu) - 2\alpha\partial^\mu(A^\nu \partial_\nu A_\mu) = \\ &= \mathcal{L}' + \text{ four-divergence.} \end{aligned} \tag{8.24}$$

The parentheses limit the action of the differential operators in these expressions. The third expression for $\mathcal{L}$ is equal to the second because of the Lorentz gauge of the four-potential. An additive four-divergence term of $\mathcal{L}$ consists of a sum of a three-divergence and a time-derivative. The three-divergence does not contribute to L and to the equations of motion, respectively, if the fields vanish at infinity in space (or if a finite space without boundary is considered, as, e.g., a three-dimensional torus). An additive time-derivative to L does also not contribute to the equations of motion, because it adds only a constant to the action S. From (8.24) we find

$$\partial_\nu \frac{\partial \mathcal{L}'}{\partial(\partial_\nu A_\mu)} = 4\alpha\partial_\nu\partial^\nu A^\mu = 4\alpha\Box A^\mu, \tag{8.25}$$

and

$$\frac{\partial \mathcal{L}'}{\partial A_\mu} = 0, \tag{8.26}$$

since the Lagrange density does not depend on the four-potential itself but only on its derivatives. Hence, the field equations of motion take the form

$$\Box A^\mu = 0, \tag{8.27}$$

thus justifying (8.24). Recall that we have essentially used the Lorentz gauge of the four-potential in deriving this result.

The covariant formalism uncovers the full symmetry of a dynamical system. A particular physical state of the system will in general of course not have this symmetry. The only requirement is that the state obtained by a symmetry transformation from a given one is again a physical state of that same system. Particular problems are usually most effectively solved using coordinates adapted to the symmetry of the state of interest only.

The transition from the Lagrange form of the equations of motion to the Hamilton form is provided by a Legendre transformation, which destroys the symmetry of space-time because it distinguishes time (and it distinguishes energy against momentum). Hence the Hamilton formalism is not covariant. The canonical momentum density of the field is[10]

$$\Pi^\mu = \frac{\partial \mathcal{L}}{\partial(\partial_0 A_\mu)} = 4\alpha(\partial^0 A^\mu - \partial^\mu A^0) = 4\alpha F^{\mu 0}. \tag{8.28}$$

The Hamilton density is the Legendre transform of the Lagrange density according to $\mathcal{H} = \sup_{(\partial_0 A_\mu)}(\Pi^\mu \partial_0 A_\mu - \mathcal{L})$ (where the supremum condition leads to (8.28)). We find

$$\begin{aligned} \mathcal{H} &= 4\alpha\left(F^{\mu 0}\partial_0 A_\mu - \frac{1}{4}F^{\mu\nu}F_{\mu\nu}\right) = \\ &= 4\alpha\left(F^{\mu 0}F_{\mu 0} - \frac{1}{4}F^{\mu\nu}F_{\mu\nu}\right) + 4\alpha F^{i0}\partial_i A_0 = \\ &= \mathcal{H}' + 4\alpha\partial_i(F^{i0}A_0). \end{aligned} \tag{8.29}$$

Both sides of the last equality are equal up to a term containing $\partial_i F^{i0} = \partial_\mu F^{\mu 0}$ as a factor, which is zero in the absence of field sources j^0 by (8.15)

[10] Note $\Pi^0 \equiv 0$; only three components of momentum correspond to the three independent components of A^μ.

(or likewise by (8.27) under the Lorentz gauge). The three-divergence term does not contribute to the integral of the Hamilton density over the space (for fields vanishing at infinity or for space without boundary). The final result is

$$\mathcal{H}' = \frac{1}{2}\left(\epsilon_0 \boldsymbol{E}^2 + \mu_0 \boldsymbol{H}^2\right), \tag{8.30}$$

if we put $4\alpha = -1/\mu_0$. This reproduces the well known expression for the energy density of an electromagnetic field.

8.4 Relativistic Kinematics

It was Einstein's essential step on top of the achievements of Lorentz and Poincaré to realize that just *kinematics*, i.e. the role of space and time in physics, had to be refined in order to bring electrodynamics and mechanics together. In a natural way, this led Minkowski to his introduction of the four-tensor formalism.

The invariant line element of Minkowski's geometry is (with $d\boldsymbol{r} = \boldsymbol{v}dt$)

$$ds^2 = dx^\mu dx_\mu = c^2 dt^2 - d\boldsymbol{r}^2 = c^2 dt^2 (1 - v^2/c^2). \tag{8.31}$$

In the rest frame of that line element one has

$$v = 0: \quad ds = c\,dt: \quad ds/c = \text{ 'proper time interval'}. \tag{8.32}$$

This suggests the introduction of a four-velocity $u^\mu = dx^\mu/(\text{proper time interval})$:

$$u^\mu = c\,dx^\mu/ds = (c, \boldsymbol{v})/\sqrt{1 - v^2/c^2}, \qquad u^\mu u_\mu = c^2, \tag{8.33}$$

whose spatial components correspond to the ordinary velocity.

Energy and momentum conservation follow from invariance under time and spatial translations, respectively. Energy and momentum must hence together form a four-vector. It is given the dimension of a momentum by replacing the energy E by E/c and is called the four-momentum. For $v/c \ll 1$, its spatial components must be $\boldsymbol{p} = m\boldsymbol{v}$, hence

$$p^\mu = m\,u^\mu = (E/c, \boldsymbol{p}), \quad E^2/c^2 - p^2 = p^\mu p_\mu = m^2 u^\mu u_\mu = m^2 c^2, \tag{8.34}$$

where m is the rest mass of the particle under consideration, i.e. the mass in the limit $v/c \ll 1$. The last chain of relations (8.34) yields

$$E = c\sqrt{m^2 c^2 + p^2} \tag{8.35}$$

as the relativistic dispersion relation between energy and momentum of a free particle.

This may also be obtained from the action

$$\begin{aligned} S &= -\int ds\, mc = -\int dt\, mc^2 \sqrt{1 - v^2/c^2}, \\ L &= -mc^2 \sqrt{1 - \dot{\boldsymbol{r}}^2/c^2}, \end{aligned} \tag{8.36}$$

since $\int ds$ is (up to constant factors defining units of physical quantities) the only invariant action which can be formed for a free particle, represented by a world line with line element ds. With $\boldsymbol{p} = \partial L/\partial \dot{\boldsymbol{r}}$ and $E = H = \boldsymbol{p} \cdot \dot{\boldsymbol{r}} - L$ we arrive again at (8.35).

8.5 Relativistic Mechanics

For a particle with charge $-e$, moving in an external (given) electromagnetic field, the most simple invariant action, which can be formed, is

$$\begin{aligned} S &= \int (-mc\, ds + eA_\mu dx^\mu) \; = \\ &= \int dt\, (-mc^2 \sqrt{1 - v^2/c^2} + eU - e\boldsymbol{A} \cdot \boldsymbol{v}) \; = \; \int dt\, L. \end{aligned} \tag{8.37}$$

Constant factors again define units. A Legendre transformation yields the canonical momentum

$$\boldsymbol{p} = \frac{\partial L}{\partial \boldsymbol{v}} = \frac{m\boldsymbol{v}}{\sqrt{1 - v^2/c^2}} - e\boldsymbol{A} \tag{8.38}$$

and the Hamilton function

$$H = \boldsymbol{p} \cdot \boldsymbol{v} - L = c\sqrt{m^2c^2 + (\boldsymbol{p} + e\boldsymbol{A})^2} - eU, \tag{8.39}$$

which describes the motion of a charged particle with charge $-e$ in given external electric and magnetic fields, characterized by their potentials U and $\boldsymbol{A}$.

As a next step one would consider charged particles and fields as a closed dynamical system, this type of theory, however, has to fight against a severe defect, because the field created by a single point-like charged particle contains infinite field energy. The electric field strength is proportional to r^{-2}, and the energy density (8.30) of the field is hence proportional to r^{-4} which

integrates to infinity at small r. Note that relativity excludes the possibility of a rigid body: if one point of that body is set to motion, the other points can follow only with retardation limited by the light velocity c, thus causing deformation and hence change of internal energy of the body, i.e. of its rest mass. Hence, any classical elementary particle with a single definite rest mass and with kinetic energy, solely characterized by one velocity, must be point-like. On the other hand, the field energy of the electric field $E = -e/(4\pi\epsilon_0 r^2)$ of an electron, obtained by integration over (8.30) from radius r to infinity, is $e^2/(4\pi\epsilon_0 r)$ and exceeds the rest energy mc^2 for r-values smaller than the classical electron radius $r_c = e^2/(4\pi\epsilon_0 mc^2)$. For lengths close to and below r_c classical electrodynamics loses its physical meaning.

8.6 The Principles of Relativistic Quantum Theory

The superposition principle of quantum physics requires the representation of quantum states by Hilbert vectors $|\Psi\rangle$, so that linear superpositions of states are again states. Besides energy conservation, in relativistic quantum physics there is no separate conservation of rest mass. An eigenstate of an observable not commuting with the Hamiltonian contains components referring to arbitrarily large energies. From experiment we know that this opens the possibility of particle-antiparticle pair creation out of the vacuum. Hence, consequent relativistic quantum theory cannot be formulated for a fixed particle number, it definitely needs field quantization—e.g. Fock space—representation.

For the same reason, the particle coordinate cannot be an observable in relativistic quantum theory: because of the uncertainty principle, a position uncertainty of a particle, smaller than Δx causes a momentum uncertainty larger than $\hbar/\Delta x$ and thus an energy uncertainty larger than $\hbar c/\Delta x$. If this energy uncertainty exceeds twice the rest mass mc^2, then pair creation sets in, and the individuality of that particle cannot be maintained. Hence, the position uncertainty of a particle is bounded below by $\lambda\!\!\!^{-} = \hbar/mc$, the Compton wavelength of that particle. (Note that the classical electron radius mentioned in the last section is smaller than its Compton wavelength by a factor $\alpha = e^2/(4\pi\epsilon_0\hbar c) = 1/137$, the fine structure constant. Hence the defect of the classical theory is completely masked by quantum theory.) Unlike the particle coordinates, momenta *are* observable quantities as there is no general physical principle setting a lower limit to momentum uncertainty.

Since the position eigenstate $|\boldsymbol{r}\rangle$ has no physical meaning (even not as an improper eigenstate of the continuous spectrum), there is no wavefunction

$\Psi(\boldsymbol{r}) = \langle \boldsymbol{r}|\Psi\rangle$ figuring in quantum field theory. Nevertheless, local field operators like $\hat{\psi}(\boldsymbol{r})$, acting in the representation space on Hilbert states, are used as dynamical variables, in the same manner as creation and annihilation operators are used in occupation number representation of non-relativistic theory: observables are expressed in terms of those dynamical variables. Particularly, local expectation values may be determined with their use.

In order to describe physical dynamics, quantum states must be considered at a given instant of time t. Stationary states are characterized (in Schrödinger picture, $\hbar$ is put to unity again from here on throughout) by

$$i\partial_0|\Psi\rangle = |\Psi\rangle\, E/c \quad \text{or} \quad |\Psi(t)\rangle = |\Psi(0)\rangle\, e^{-iEt} \tag{8.40}$$

For the physical vacuum state, denoted by $|\rangle$,

$$\partial_0|\rangle = 0 \tag{8.41}$$

holds, and, in contrast to non-relativistic physics, energies E are non-negative in relativistic physics: a negative energy E of a state would imply a negative rest mass of the system, for the existence of which nature provides no evidence.

If, at $t = 0$, $\hat{c}^\dagger$ creates stationary matter with $E > 0$ out of the vacuum, and $\hat{c}$ annihilates it,

$$|\Psi(0)\rangle = \hat{c}^\dagger|\rangle, \qquad |\rangle = \hat{c}|\Psi(0)\rangle, \tag{8.42}$$

then time evolution is either described in Schrödinger picture by time-dependent states and time-independent operators or alternatively in Heisenberg picture by time-independent states and time-dependent operators, both types of quantities coinciding at $t = 0$. From (8.40, 8.42) one has

$$\begin{aligned} e^{iEt} &= \langle\Psi(t)|\Psi(0)\rangle = (\langle\Psi(0)|e^{iEt})(\hat{c}^\dagger|\rangle) = \\ &= \langle\Psi(0)|(e^{iEt}\hat{c}^\dagger)|\rangle = \langle\Psi|\hat{c}^\dagger(t)|\rangle, \end{aligned} \tag{8.43}$$

and

$$\begin{aligned} e^{-iEt} &= \langle\Psi(0)|\Psi(t)\rangle = (\langle|\hat{c})(e^{-iEt}|\Psi(0)\rangle) = \\ &= \langle|(\hat{c}e^{-iEt})|\Psi(0)\rangle = \langle|\hat{c}(t)|\Psi\rangle. \end{aligned} \tag{8.44}$$

Hence, in Heisenberg picture creation and annihilation operators, respectively, of stationary matter have definite time-dependences

$$\hat{c}^\dagger(t) = \hat{c}^\dagger\, e^{iEt}, \quad \hat{c}(t) = \hat{c}\, e^{-iEt}, \quad E > 0. \tag{8.45}$$

Field operators are generally composed of linear combinations of both creation and annihilation operators. In Heisenberg picture they are functions of position and of time, which generally allows for a covariant way of writing.

As an example the field operator of the free photon field

$$\hat{A}^\mu(x^\rho) = \sum_{\boldsymbol{k}\nu} \sqrt{\frac{\mu_0}{2k_0}} \left[\hat{c}^\nu_{\boldsymbol{k}}\, u^\mu_\nu\, e^{-ikx} + \hat{c}^{\nu\dagger}_{\boldsymbol{k}}\, u^\mu_\nu\, e^{ikx} \right], \tag{8.46}$$

$$[\hat{c}^\mu_{\boldsymbol{k}}, \hat{c}^\nu_{\boldsymbol{k}'}] = 0 = [\hat{c}^{\mu\dagger}_{\boldsymbol{k}}, \hat{c}^{\nu\dagger}_{\boldsymbol{k}'}], \quad [\hat{c}^\mu_{\boldsymbol{k}}, \hat{c}^{\nu\dagger}_{\boldsymbol{k}'}] = -g^{\mu\nu}\,\delta_{\boldsymbol{k}\boldsymbol{k}'}, \quad u^\mu_\sigma g^{\sigma\tau} u^\nu_\tau = g^{\mu\nu} \tag{8.47}$$

is considered. Here, kx in the exponents abbreviates $k_\mu x^\mu$, and, with $k_0 = |\boldsymbol{k}|$, (8.46) obeys the field equation $\Box\hat{A}^\mu = 0$ for the free photon field (cf. (8.27)), as it should. The u^μ_ν, $\nu = 0, 1, 2, 3$ are constant polarization vectors; completeness demands that they form an isometric transformation matrix in Minkowski's metric (last relation (8.47)). However, in order to have *covariant* canonical bosonic commutation relations as (8.47)—a demand of causality, otherwise the field components $\hat{A}^\mu$ would not commute outside of the light cone (i.e., $\hat{A}^\mu(x^\rho)$ with $\hat{A}^\nu(x'^\rho)$ for $(x^\rho - x'^\rho)(x_\rho - x'_\rho) < 0$, cf. [Itzykson and Zuber, 1980])—, for $\mu = \nu = 0$ it had to be given the wrong sign leading to creation of Fock-space states with negative norm.[11] On the other hand, (8.46) does not yet correspond to the classical electrodynamics, because $\Box A^\mu = 0$ was the correct classical field equation only under the Lorentz gauge $\partial_\mu A^\mu = 0$. This gauge, as applied to the field operator (8.46), again conflicts with (8.47) and hence with causality. Therefore, instead of posing the Lorentz gauge as a condition on the field operator, one must consider it as a condition on physical states $|\Psi\rangle$: only states with $\partial_\mu \hat{A}^\mu|\Psi\rangle = 0$ are to be considered physical. Physical states then have a positive norm and are characterized by occupation of transverse photon modes only. (For details cf. again e.g. [Itzykson and Zuber, 1980].)

Finally, the prefactor in (8.46) was chosen in such a way, that the Hamiltonian in correspondence with (8.30) is

$$\begin{aligned} \hat{H} &= \int d^3r\, \frac{1}{2}\left(\epsilon_0 \hat{\boldsymbol{E}}^2 + \mu_0 \hat{\boldsymbol{H}}^2\right) = \\ &= \sum_{\boldsymbol{k}} -c|\boldsymbol{k}|\, \hat{c}^{\rho\dagger}_{\boldsymbol{k}} u^\mu_\rho g_{\mu\nu} u^\nu_\sigma \hat{c}^\sigma_{\boldsymbol{k}} + \text{ const.}, \end{aligned} \tag{8.48}$$

[11] With the natural definition of the vacuum, $\langle|\rangle = 1$, $\hat{c}^{0\dagger}_{\boldsymbol{k}}|\rangle = |\boldsymbol{k}0\rangle$, $\hat{c}^0_{\boldsymbol{k}}|\rangle = 0$, one finds from (8.47) for $\mu = \nu = 0$ that $\langle \boldsymbol{k}0|\boldsymbol{k}0\rangle = \langle|\hat{c}^0_{\boldsymbol{k}}\hat{c}^{0\dagger}_{\boldsymbol{k}}|\rangle = -\langle|\rangle + \langle|\hat{c}^{0\dagger}_{\boldsymbol{k}}\hat{c}^0_{\boldsymbol{k}}|\rangle = -1$, i.e., a time-polarized photon would have a negative norm.

where $\hat{\boldsymbol{E}}$ and $\hat{\boldsymbol{H}}$ follow from $\hat{A}^\mu$ like in the classical case, and integration is over a normalization volume, which we tacitly put to unity. The (infinite) additive constant is put to zero by a simple renormalization, consisting of rearranging products of field operators to normal order, so that creators always precede annihilators from left to right. Normal order is denoted by putting the corresponding expressions between colons:

$$\hat{H} = \int d^3r \, \frac{1}{2} \, : \left(\epsilon_0 \hat{\boldsymbol{E}}^2 + \mu_0 \hat{\boldsymbol{H}}^2\right) : \, = \sum_{\boldsymbol{k}} -c|\boldsymbol{k}| \, \hat{c}^{\rho\dagger}_{\boldsymbol{k}} u^\mu_\rho g_{\mu\nu} u^\nu_\sigma \hat{c}^\sigma_{\boldsymbol{k}}. \tag{8.49}$$

Due to the wrong sign of the commutation relations for time-polarized photons, this Hamiltonian is not positive definite in the total Hilbert space; it is, however, definite in the physical sector of states containing only transverse photons. If the $\hat{c}^{(t)\dagger}_{\boldsymbol{k}} = \hat{c}^{\rho\dagger}_{\boldsymbol{k}} u^t_\rho, \; t = 1, 2$ create transverse photons, the projection of the Hamiltonian on the physical sector of the Hilbert space is $\sum_{\boldsymbol{k}} \sum_t^{1,2} \, c|\boldsymbol{k}| \, \hat{c}^{(t)\dagger}_{\boldsymbol{k}} \hat{c}^{(t)}_{\boldsymbol{k}}$.

Since photons are bosons, modes may be occupied in macroscopic number, forming a classical (Bose condensed) field, which may be embossed into the 'quantum vacuum' by renormalization. Time-independent electromagnetic fields are always classical, Bose condensed. Therefore their nonzero longitudinal components do not pose problems with the Hilbert space norm of the quantized (transverse) part of the field.

8.7 The Dirac Field

For the energy of a free electron, the dispersion relation

$$E^2 = \boldsymbol{p}^2 c^2 + m^2 c^4 \tag{8.50}$$

holds. Putting $i\partial^\mu = (E/c, \boldsymbol{p})$, one finds the Klein-Gordon equation

$$(\partial^\mu \partial_\mu + m^2 c^2)\psi(x^\nu) = 0. \tag{8.51}$$

Containing a second time-derivative and no spin operator, it does on the one hand not really look close to the Schrödinger equation, which should be its non-relativistic limit, and describes on the other hand a scalar (bosonic) field and not a spinor (fermionic) field. The Dirac equation instead is a four-component equation

$$i\partial_0 \psi = (-i\boldsymbol{\alpha} \cdot \nabla + \beta mc) \, \psi = \hat{H}_{\mathrm{D}} \, \psi, \quad \psi = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{pmatrix} \tag{8.52}$$

with 4×4 matrices α^i and β such that each component of a ψ obeying this Dirac equation is a solution of the Klein-Gordon equation. From (8.52) it follows that $0 = (i\partial_0 + \hat{H}_{\mathrm{D}})(i\partial_0 - \hat{H}_{\mathrm{D}})\psi = -(\partial_0^2 + \hat{H}_{\mathrm{D}}^2)\psi$. Hence we demand

$$-\partial_0\partial^0\psi = (-i\boldsymbol{\alpha}\cdot\nabla + \beta mc)^2\psi \stackrel{!}{=} (-\nabla^2 + m^2c^2)\,\psi. \tag{8.53}$$

The last relation requires the algebraic properties

$$[\alpha^i, \alpha^k]_+ = 2\delta_{ik}, \quad [\alpha^i, \beta]_+ = 0, \quad \beta^2 = 1, \tag{8.54}$$

where 1 means a 4×4 unit matrix, and δ_{ik} as well is for every index pair ik either a 4×4 unit matrix or a zero matrix. These algebraic relations may be cast into a covariant form

$$\gamma^0 = \beta, \quad \gamma^i = \beta\alpha^i, \qquad [\gamma^\mu, \gamma^\nu]_+ = 2g^{\mu\nu}. \tag{8.55}$$

having a simple representation as

$$\gamma^0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad \gamma^i = \begin{pmatrix} 0 & \sigma_i \\ -\sigma_i & 0 \end{pmatrix}, \tag{8.56}$$

or

$$\beta = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad \alpha^i = \begin{pmatrix} 0 & \sigma_i \\ \sigma_i & 0 \end{pmatrix}, \tag{8.57}$$

where 1 means a 2×2 unit matrix, and σ_i are the well known Pauli matrices (twice the matrices (1.10)).

In covariant writing the Dirac equation is now

$$(i\gamma^\mu\partial_\mu - mc)\psi = 0. \tag{8.58}$$

As charge conservation is implemented in quantum theory as a symmetry with respect to some inner conjugation, we want to consider the Hermitian conjugate of this equation. Observe in this respect, that α^i and $\beta = \gamma^0$ are Hermitian matrices, while $\gamma^{i\dagger} = \gamma^0\gamma^i\gamma^0$. Hermitian conjugation of the equation (8.58) thus leads to $\psi^\dagger\gamma^0(-i\overleftarrow{\partial}_\mu\,\gamma^\mu - mc)\gamma^0 = 0$ (since we had to reverse the order of matrix multiplications, differentiation must now operate to the left). The result is usually written as

$$\bar{\psi}(i\overleftarrow{\partial}_\mu\,\gamma^\mu + mc) = 0, \qquad \bar{\psi} \stackrel{\mathrm{def}}{=} \psi^\dagger\gamma^0, \qquad (\bar{\gamma}^\mu = \gamma^\mu). \tag{8.59}$$

Adding (8.58), multiplied by $\bar{\psi}$ from the left, and (8.59), multiplied by ψ from the right, yields

$$\bar{\psi}(\overleftarrow{\partial}_\mu\, \gamma^\mu + \gamma^\mu\, \overrightarrow{\partial}_\mu)\psi = \partial_\mu(\bar{\psi}\gamma^\mu\psi) = 0. \tag{8.60}$$

Hence, the conserved four-current with density

$$-ej^\mu = -ec\bar{\psi}\gamma^\mu\psi \tag{8.61}$$

appears as a proposed candidate for the electric four-current.

For physical interpretation of the field ψ itself, less compressed expressions are needed. If one composes the four-component field ψ as a bispinor by two two-component spinors ϕ and χ, the Dirac equation acquires the form

$$\psi = \begin{pmatrix} \phi \\ \chi \end{pmatrix}, \qquad \left\{ \begin{array}{rcl} i\dot{\phi} & = & mc^2\phi - ic\boldsymbol{\sigma}\cdot\nabla\chi \\ i\dot{\chi} & = & -mc^2\chi - ic\boldsymbol{\sigma}\cdot\nabla\phi \end{array} \right. \tag{8.62}$$

This shows, that the Dirac equation has two types of solutions: if ϕ is large, then the time-derivative corresponds to a positive energy $\sim mc^2$, and if χ is large, then the time-derivative corresponds to a negative energy $\sim -mc^2$. Hence, in the field operator of the free Dirac field

$$\begin{aligned} \hat{\psi}(x^\mu) &= \sum_{\boldsymbol{k}\sigma} \frac{1}{\sqrt{2k_0}} \left[\hat{a}_{\boldsymbol{k}\sigma}\, u_{\boldsymbol{k}\sigma}\, e^{-ikx} + \hat{b}^\dagger_{\boldsymbol{k}\sigma}\, v_{\boldsymbol{k}\sigma}\, e^{ikx} \right], \\ k_0 &= \sqrt{\boldsymbol{k}^2 + m^2c^2} \end{aligned} \tag{8.63}$$

the amplitude bispinor $u_{\boldsymbol{k}\sigma}$ must be taken from a solution of the Dirac equation with large ϕ, and $v_{\boldsymbol{k}\sigma}$ must be taken from a solution with large χ. For both cases there are two independent orthogonal solutions obeying

$$\sum_\sigma^{1,2} u_{\boldsymbol{k}\sigma}\bar{u}_{\boldsymbol{k}\sigma} = \gamma^\mu k_\mu + mc, \qquad \sum_\sigma^{1,2} v_{\boldsymbol{k}\sigma}\bar{v}_{\boldsymbol{k}\sigma} = -\gamma^\mu k_\mu + mc. \tag{8.64}$$

Since these conditions have a covariant form, no conflict between field quantization and causality appears.

To obtain (8.64), first observe that the Dirac equation (8.58) for $\hat{\psi}$ requires

$$(\gamma^\mu k_\mu - mc)\, u_{\boldsymbol{k}\sigma} = 0, \qquad (\gamma^\mu k_\mu + mc)\, v_{\boldsymbol{k}\sigma} = 0, \tag{8.65}$$

where k_0 was already chosen so that $(\gamma^\mu k_\mu)^2 - (mc)^2 = 0$. Hence we may put

$$u_{\boldsymbol{k}\sigma} \sim (\gamma^\mu k_\mu + mc) e_{\boldsymbol{k}\sigma}, \qquad v_{\boldsymbol{k}\sigma} \sim (\gamma^\mu k_\mu - mc) f_{\boldsymbol{k}\sigma}, \tag{8.66}$$

where the $e_{\boldsymbol{k}\sigma}$'s have zeros in the last two components and the $f_{\boldsymbol{k}\sigma}$'s have zeros in the first two components, whence

$$\sum_\sigma^{1,2} e_{\boldsymbol{k}\sigma}\bar{e}_{\boldsymbol{k}\sigma} \sim (1+\gamma^0)/2, \qquad \sum_\sigma^{1,2} f_{\boldsymbol{k}\sigma}\bar{f}_{\boldsymbol{k}\sigma} \sim (1-\gamma^0)/2. \tag{8.67}$$

Further observe, again because of $(\gamma^\mu k_\mu)^2 - (mc)^2 = 0$, that $(\pm\gamma^\mu k_\mu + mc)(1 \pm \gamma^0)(\pm\gamma^\mu k_\mu + mc) = (\pm\gamma^\mu k_\mu + mc)[(\pm\gamma^\mu k_\mu - mc)(1 \mp \gamma^0) + 2mc + 2k_0] = (\pm\gamma^\mu k_\mu + mc)2(mc + k_0)$. Hence, if we normalize $e_{\boldsymbol{k}\sigma}$, $f_{\boldsymbol{k}\sigma} \sim 1/\sqrt{mc + k_0}$, we arrive at (8.64). Along the same line scalar products $\bar{u}_{\boldsymbol{k}\sigma} u_{\boldsymbol{k}\sigma} = \bar{v}_{\boldsymbol{k}\sigma} v_{\boldsymbol{k}\sigma} = mc$, $\bar{u}_{\boldsymbol{k}\sigma} v_{\boldsymbol{k}'\sigma'} = 0, \bar{u}_{\boldsymbol{k}\sigma}\boldsymbol{\gamma} u_{\boldsymbol{k}\sigma} = 2\boldsymbol{k}$, and so on, are obtained.

In order to have a causal anticommutation rule for $\hat{\psi}$ and $\hat{\bar{\psi}}$, one must require

$$\begin{aligned} [\hat{a}_{\boldsymbol{k}\sigma}, \hat{a}_{\boldsymbol{k}'\sigma'}]_+ &= 0 = [\hat{b}_{\boldsymbol{k}\sigma}, \hat{b}_{\boldsymbol{k}'\sigma'}]_+, \quad [\hat{a}_{\boldsymbol{k}\sigma}, \hat{b}_{\boldsymbol{k}'\sigma'}]_+ = 0 = [\hat{a}_{\boldsymbol{k}\sigma}, \hat{b}^\dagger_{\boldsymbol{k}'\sigma'}]_+, \\ [\hat{a}_{\boldsymbol{k}\sigma}, \hat{a}^\dagger_{\boldsymbol{k}'\sigma'}]_+ &= \delta_{\sigma\sigma'}\delta_{\boldsymbol{k}\boldsymbol{k}'} = [\hat{b}_{\boldsymbol{k}\sigma}, \hat{b}^\dagger_{\boldsymbol{k}'\sigma'}]_+. \end{aligned} \tag{8.68}$$

(With commutation rules, one would not be able to obtain causal fields $\hat{\psi}$ and $\hat{\bar{\psi}}$, commuting outside of the light cone; this is how Pauli's spin-statistic theorem is obtained, which states that half-integer spin particles must be fermions, and integer spin particles must be bosons; e.g. [Itzykson and Zuber, 1980].) Since the electric charge density is, according to (8.61, 8.62), equal to $: -e\bar{\phi}\phi + e\bar{\chi}\chi :$, it is clear that $\hat{a}^\dagger_{\boldsymbol{k}\sigma}$ creates an electron and $\hat{b}^\dagger_{\boldsymbol{k}\sigma}$ creates a positron.

The Hamiltonian of the free Dirac field is obtained as

$$\begin{aligned} \hat{H} &= \int d^3r\, c : \hat{\bar{\psi}}(-i\boldsymbol{\gamma}\cdot\nabla + mc)\hat{\psi} : = \\ &= \sum_{\boldsymbol{k}\sigma} c\sqrt{\boldsymbol{k}^2 + m^2c^2}\, [\hat{a}^\dagger_{\boldsymbol{k}\sigma}\hat{a}_{\boldsymbol{k}\sigma} + \hat{b}^\dagger_{\boldsymbol{k}\sigma}\hat{b}_{\boldsymbol{k}\sigma}]. \end{aligned} \tag{8.69}$$

Here, in the normal ordering process fermionic sign changes are understood.

As in Section 8.5, electromagnetic coupling (with electron charge $-e$) is obtained by adding a term $ej^\mu A_\mu$ to the Lagrangian density. (In Section 8.5 j^μ was given by the particle position and velocity.) Since in the Dirac theory

$$\mathcal{L}_{\text{int}} = e : \hat{j}^\mu \hat{A}_\mu : = ec : \hat{\bar{\psi}}\gamma^\mu\hat{\psi}\hat{A}_\mu : \tag{8.70}$$

does not contain time-derivatives of the fields, it does not modify the canonical field momenta, hence the same interaction term with reversed sign enters the Hamiltonian density:

$$\mathcal{H}_{\text{int}} = -e : \hat{j}^{\mu} \hat{A}_{\mu} : . \tag{8.71}$$

This theory, taken literally, again has a defect: it can be shown to renormalize all observable electric charges to zero, hence it must be used with *phenomenological* charge renormalization to the observed value. The defect can be cured by replacing the electromagnetic theory by electro-weak theory with a not too large number of coupled fermions [Berestetskii, 1976], but this well confirmed theory again ...

9 Current Density Functional Theory

The basic concern of density functional theory is the ground state energy of an *inhomogeneous* interacting many-particle system as a function of the external potential acting on its density (and possibly as a function of the particle number, allowing for the consideration of ionization energies, affinities, and excitation gaps). This ground state energy function is replaced by its Legendre transform as a function of the particle density, and the Legendre back transformation leads via insertion of an interaction-free reference system to an effective one-particle equation whose solution finds the ground state energy and density to any given external potential. As was seen in Part I, this theory provides rather a reformulation than a solution of the many-body problem for the ground state, because it does not explicitly provide the Legendre transform $H[n]$ to $E[v, N]$ (cf. (6.22 and 6.25)). The original many-body problem is now contained in $H[n]$ which only can be guessed, although not without success.

The most radical relativistic generalization of this theory, which is considered in the present chapter, does the same for the ground state energy of an *inhomogeneous* interacting quantum field of matter as a function of the external field acting on its particle current density (and possibly as a function of the conserved charges of the field of matter). It again provides an effective single-particle equation—the Kohn-Sham-Dirac equation—for obtaining the ground state energy and current density from the Legendre transform of the ground state energy functional. This is now a current density functional which potentially contains the whole information on ground states of interacting quantum fields (of a given type), and which, of course, is all the more unknown. One can but apply guesswork to account for part of relativistic effects of the problem.

In the first section, the ground state of an inhomogeneous quantum field and its current density are defined and the necessary longitudinal fields of particle interaction are introduced as condensed mean fields. The Kohn-Sham-Dirac equation in its most general form is derived in the subsequent section, and in the third section the four-current is decomposed into its orbital and spin parts in order to achieve closer resemblance to the non-

relativistic theory. With respect to the exchange and correlation energy functional, the shape of the relativistic theory also provides useful suggestions for non-relativistic functionals. Some related approximative functionals are presented in section four. The most important relativistic term is of course that of the kinetic energy, yielding the Dirac form of the Kohn-Sham equation. Any implementation of this theory must take care for a proper projection of the Dirac Hamiltonian onto the sector of electron states only. Otherwise the Hamiltonian in Schrödinger representation would not be bounded below (as the field-theoretic Hamiltonian is due to normal order). A very accurate numerical implementation together with a survay on relativistic density functional applications is presented in [Eschrig *et al.*, 2003b].

To be specific, the material of this chapter is presented for the case of the electron-positron-photon field, the only relevant known case in nature of a field subject to the action of an inhomogeneous *external* potential. In principle it applies, however, to any field theory, and even to cases which are inhomogeneous due to spontaneous symmetry breaking only, as e.g. atomic nuclei constituting droplets of nuclear matter in homogeneous surroundings. (The truly homogeneous ground state is of course also included in density functional theory, however, this theory does not provide any *result* for that case; the *independently* acquired knowledge on that case is on the contrary used to guess the density functional.)

9.1 QED Ground State in a Static External Field

We consider a quantum electrodynamical system in a static external classical (Bose condensed) field $F^{\mu\nu}$, given by the four-potential

$$A^\mu(x^\sigma): \qquad \partial_0 F^{\mu\nu} = \partial_0(\partial^\nu A^\mu - \partial^\mu A^\nu) = 0. \tag{9.1}$$

Of course, the static condition refers to a certain distinguished reference frame. We further assume spatial periodicity in a large periodic spatial volume $|\boldsymbol{T}^3|$ with respect to that reference frame, and refer all integrated quantities to that volume $|\boldsymbol{T}^3|$ (toroidal space; note that in QED this step has another completely independent advantage: it prevents an 'infrared catastrophe' of creation of long wavelength photons in unlimited number). Of course, a thorough treatment needs the same caution as described in Section 2.7.

The Hamiltonian in question is

$$\hat{H}_A = \int d^3r\,(\hat{\mathcal{H}} - e\hat{j}^\mu A_\mu), \tag{9.2}$$

where

$$\hat{j}^\mu = c : \hat{\bar{\psi}}\gamma^\mu\hat{\psi} :, \tag{9.3}$$

is the four-current density operator of the electron-positron field, and (cf. (8.69, 8.71, and 8.49)

$$\hat{\mathcal{H}} = c : \hat{\bar{\psi}}(-i\boldsymbol{\gamma}\cdot\nabla + mc)\hat{\psi} : - e\hat{j}^\mu\hat{A}_\mu + \frac{1}{2} : (\epsilon_0\hat{\boldsymbol{E}}^2 + \mu_0\hat{\boldsymbol{H}}^2) : . \tag{9.4}$$

(The colons for normal order of $\hat{j}^\mu$ were already included in the definition of the four-current density operator and need not be repeated in the second term of $\hat{\mathcal{H}}$; since $\hat{\psi}$ and $\hat{A}^\mu$ commute, no reordering of the product is needed.) The field operator $\hat{A}^\mu$ describes the field created in the quantum electrodynamical system and is to be distinguished from the external (c-number) field A^μ in (9.1, 9.2). Furthermore, (cf. (8.4, 8.16))

$$\hat{\boldsymbol{E}} = -c(\partial_0\hat{\boldsymbol{A}} + \nabla\hat{A}_0), \qquad \mu_0\hat{\boldsymbol{H}} = \nabla\times\hat{\boldsymbol{A}}. \tag{9.5}$$

Note that we did not include the field energy of the external field A^μ in the Hamiltonian. Therefore, the eigenvalues of that Hamiltonian are allowed to be negative (for negative $-eA^0$). For the following it is furthermore crucial that *normal order in* (9.4) *refers to the creation and annihilation operators of the (renormalized) asymptotic fields of the homogeneous system, i.e. with zero external four-potential A^μ, so that this intrinsic Hamiltonian density is independent of the external four-potential.*

In Heisenberg picture this Hamiltonian leads to operator field equations of motion

$$\left[i\gamma^\mu(\partial_\mu - ie\hat{A}_\mu - ieA_\mu) - mc\right]\hat{\psi} = 0, \tag{9.6}$$

$$\partial_\nu\hat{F}^{\mu\nu} = -\mu_0 e\hat{j}^\mu. \tag{9.7}$$

They determine the function dependence of the Heisenberg field operators on x^σ in accordance with boundary conditions to be given additionally. Particularly, equal-time canonical (anti-)commutation relations between canonical conjugate field variables must be required as initial conditions for the equations (9.6, 9.7).

In quantum field theory one mostly considers scattering situations (in an infinite spatial volume V). One considers given incoming (asymptotic) fields

at far past (initial conditions) and adds retarded (as required by causality) contributions appearing from interaction, or alternatively one considers given outgoing fields at far future and adds advanced contributions absorbed by interaction. The connection between given incoming and given outgoing fields is then determined by the scattering matrix S. The quantum state $|\Psi\rangle$ is fixed in this Heisenberg picture. The total electric charge $-eN = -(e/c)\int d^3r\,\langle\Psi|\hat{j}^0|\Psi\rangle$ of this state is also fixed due to charge conservation $\partial_\mu\hat{j}^\mu = 0$. N is the excess number of electrons against positrons. The number of electron-positron pairs as well as the number of photons, however, may change in time.

By way of contrast, here we are facing another situation, where the expectation values of all observables are stationary, i.e.,

$$\langle\Psi|\hat{j}^\mu|\Psi\rangle = J^\mu \tag{9.8}$$

and

$$\langle\Psi|\hat{F}^{\mu\nu}|\Psi\rangle = f^{\mu\nu} \tag{9.9}$$

are time-independent. We fix the total charge

$$Q = -\frac{e}{c}\int d^3r\,J^0(\boldsymbol{r}) \tag{9.10}$$

in the system and consider the ground state of the quantum field as that state minimizing

$$E[A,Q] \stackrel{\text{def}}{=} \min_{\Psi}\left\{\langle\Psi|\hat{H}_A|\Psi\rangle \;\middle|\; -\frac{e}{c}\int d^3r\,\langle\Psi|\hat{j}^0|\Psi\rangle = Q\right\}. \tag{9.11}$$

(In a spatial torus we can always assume that this ground state exists provided $\hat{H}_A$ is bounded below on the sector of states with fixed Q.) Because of (9.7) we have, however, $-\mu_0 eJ^\mu = \langle\Psi|\partial_\nu\hat{F}^{\mu\nu}|\Psi\rangle = \langle\Psi|\partial_\nu\partial^\nu\hat{A}^\mu - \partial_\nu\partial^\mu\hat{A}^\nu|\Psi\rangle$, and hence, whenever J^μ is non-zero, we face a situation with anomalous means $\langle\Psi|\hat{A}^\mu|\Psi\rangle \neq 0$, i.e. $\langle\Psi|\hat{c}^\mu_{\boldsymbol{k}}|\Psi\rangle \neq 0 \neq \langle\Psi|\hat{c}^{\mu\dagger}_{\boldsymbol{k}}|\Psi\rangle$, indicative of the presence of a condensate in the state $|\Psi\rangle$: a stationary current (e.g. a charge as its time component) produces a stationary (condensed) electromagnetic field.

The standard renormalization procedure [Itzykson and Zuber, 1980, §§ 11.4.2 and 12.5.3] in such a situation is to subtract $|\Psi\rangle$-dependent mean

values from the quantum fields $\hat{A}^\mu$ and $\hat{j}^\mu$, and to pass over to a description with reduced fields $\hat{a}^\mu$ and $\hat{J}^\mu$ defined by

$$\hat{A}^\mu = \hat{a}^\mu + a^\mu, \qquad \langle\Psi|\hat{a}^\mu|\Psi\rangle = \langle\hat{a}^\mu\rangle = 0, \tag{9.12}$$

$$\hat{j}^\mu = \hat{J}^\mu + J^\mu, \qquad \langle\Psi|\hat{J}^\mu|\Psi\rangle = \langle\hat{J}^\mu\rangle = 0. \tag{9.13}$$

Separating the mean value from the field equation (9.7) yields

$$\partial_\nu f^{\mu\nu} = -\mu_0 e J^\mu, \qquad \partial_\nu \hat{f}^{\mu\nu} = -\mu_0 e \hat{J}^\mu. \tag{9.14}$$

Recall that A^μ, a^μ, J^μ are time-independent, and hence $\nabla \cdot \boldsymbol{J} = 0$. The former state $|\Psi\rangle$ is now to be understood as belonging to the vacuum sector with respect to the fields $\hat{J}^\mu$, $\hat{a}^\mu$, i.e., to the Fock space above the vacuum of those fields, which does not allow for anomalous means of those fields.

The next step consists in transforming the Hamiltonian to the new variables. First observe that the static electric field $\boldsymbol{e} = -c\nabla a_0$ appears out of the time component of the original four-potential operator, end hence, because of the indefinite metric of the corresponding Hilbert space (cf. the footnote on page 190), $\langle: \hat{\boldsymbol{E}}^2 :\rangle = -\boldsymbol{e}^2 + \langle: \hat{\boldsymbol{e}}^2 :\rangle$, while $\langle: \hat{\boldsymbol{H}}^2 :\rangle = \boldsymbol{h}^2 + \langle: \hat{\boldsymbol{h}}^2 :\rangle$. Moreover,

$$-\epsilon_0 \boldsymbol{e}^2 = \epsilon_0 c \boldsymbol{e} \cdot \nabla a_0 \mathrel{\hat{=}} -\epsilon_0 c a_0 \nabla \cdot \boldsymbol{e} = e a_0 J^0, \tag{9.15}$$

$$\mu_0 \boldsymbol{h}^2 = (\nabla \times \boldsymbol{a}) \cdot \boldsymbol{h} \mathrel{\hat{=}} \boldsymbol{a} \cdot (\nabla \times \boldsymbol{h}) = -e \boldsymbol{a} \cdot \boldsymbol{J}. \tag{9.16}$$

The sign $\hat{=}$ means that the expressions are equivalent under the spatial integral (after integration by parts).

Terms linear in the new field operators are omitted in the transformed Hamiltonian, since they have zero expectation values in the vacuum sector of the reduced fields. The effective Hamiltonian of the inhomogeneous quantum field system is now obtained as

$$\hat{H}^{\text{eff}}_{A,Q} = \int d^3r \left(\hat{\mathcal{H}}^{\text{eff}}_J - e J^\mu A_\mu\right), \qquad -\frac{e}{c}\int d^3r\, J^0(\boldsymbol{r}) = Q, \tag{9.17}$$

$$\begin{aligned}\hat{\mathcal{H}}^{\text{eff}}_J \;&=\; c : \hat{\bar{\psi}}(-i\boldsymbol{\gamma}\cdot\nabla + mc)\hat{\psi} : \; + \; \frac{1}{2} : (\epsilon_0 \hat{\boldsymbol{e}}^2 + \mu_0 \hat{\boldsymbol{h}}^2) : \; - e\hat{J}^\mu \hat{a}_\mu \; - \\ &\quad - \; \frac{1}{2} e J^\mu a_\mu,\end{aligned} \tag{9.18}$$

where the last term collects the mean values of the two previous ones. It represents the mean-field interaction in the considered state. The effective

Hamiltonian is labeled J since its definition (the definition of the figuring field operators) depends on the expectation value J of the four-current, which also determines the expectation value a^μ of the intrinsic four-potential via the first equation (9.14) together with suitable boundary conditions. The normal order is the same as is in effect in (9.4).

Although many questions regarding the save content of a quantum field theory with realistic interacting quantum fields have not yet got a final answer, one may expect that for every J^μ and every state $|\Psi\rangle$ in the vacuum sector of the reduced fields there exists a state $|\Psi'\rangle$ figuring in (9.11) and vice versa. The expectation values in corresponding to each other states of both Hamiltonians coincide. Hence, the ground state energy is now obtained as

$$E[A,Q] = \min_{\Psi} \left\{ \langle\Psi|\hat{H}^{\rm eff}_{A,Q}|\Psi\rangle \;\middle|\; \langle\Psi|\hat{J}^\mu|\Psi\rangle = 0 \right\}. \tag{9.19}$$

In the operator part of the Hamiltonian $\hat{H}^{\rm eff}_{A,Q}$, correlation and all vacuum polarization processes are retained. The corresponding states may again be restricted to those obeying $\partial_\mu \hat{a}^\mu|\Psi\rangle = 0$ and containing transversal photons of the field $\hat{a}^\mu$ only. Note that for reasons which soon become clear, in the first term of (9.18) the full current $\hat{j}^\mu = \hat{J}^\mu + J^\mu$ has been retained.

The Hamiltonian (9.17) is manifestly gauge invariant with respect to transformations

$$\boldsymbol{a}/2 + \boldsymbol{A} \longrightarrow \boldsymbol{a}/2 + \boldsymbol{A} + \nabla\chi, \quad \hat{\psi} \longrightarrow \hat{\psi}\exp(-ie\chi) \tag{9.20}$$

with an arbitrary c-number function $\chi(\boldsymbol{r})$. Furthermore, an additive potential constant to A^0 will not affect the ground state $|\Psi_{AQ}\rangle$ of (9.11) and only add a constant to its energy:

$$A^0 \longrightarrow A^0 + w \;: \qquad E[A,Q] \longrightarrow E[A,Q] + cQw \tag{9.21}$$

cf. (4.6). This has the same consequences as in the non-relativistic case. Note that the projection of the gauge of $\hat{\boldsymbol{a}}$ onto the physical sector of photon states (cf. the text after (8.47)) is fixed by the second relation (9.12) to be transversal.

The content of this section may appear rather esoteric to some reader, however, accepting the existence and objectivity of the functional $E[A,Q]$ of (9.19) is enough for all what follows. The preceding text tried to connect the *formal* theory of the following sections to QED.

9.2 Current Density Functionals and Kohn-Sham-Dirac Equation

The non-relativistic limit of $E[A,Q]$ from (9.11) or equivalently from (9.19) is $E[v,N]$, where $v = -ecA^0$ and $N = -Q/e$. One therefore may proceed in analogy to Section 6.1.

First, fix A and take two values of charge Q_i, $i = 1,2$ to be integer multiples of the charge quantum e. Let $|\Psi_{AQ_i}\rangle$ be the corresponding ground states according to (9.11) and let $0 < \alpha < 1$, $|\Psi_\alpha\rangle = \sqrt{\alpha}|\Psi_{AQ_1}\rangle + \sqrt{1-\alpha}|\Psi_{AQ_2}\rangle$. Note that the $|\Psi_{AQ_i}\rangle$ are orthogonal to each other (because of an integer difference of charge quanta) eigenstates of $\hat{H}_A$ and $\hat{Q}$, and that $\langle\Psi_\alpha|\hat{Q}|\Psi_\alpha\rangle = \alpha Q_1 + (1-\alpha)Q_2$. Hence, $E[A,\alpha Q_1 + (1-\alpha)Q_2] \leq \langle\Psi_\alpha|\hat{H}_A|\Psi_\alpha\rangle = \alpha\langle\Psi_{AQ_1}|\hat{H}_A|\Psi_{AQ_2}\rangle + (1-\alpha)\langle\Psi_{AQ_2}|\hat{H}_A|\Psi_{AQ_2}\rangle = \alpha E[A,Q_1] + (1-\alpha)E[A,Q_2]$. As in the non-relativistic case, states of an expectation value Q of charge which is not an integer multiple of the charge quantum have energies interpolating linearly between eigenenergies corresponding to integer numbers of charge quanta, because $\hat{Q}$ commutes with $\hat{H}_A$ and hence eigenstates of the one operator can always be chosen to be also eigenstates of the other. Altogether we find, that, *for fixed A, $E[A,Q]$ is always a convex function of Q* as, for fixed v, $E[v,N]$ was always a convex function of N (as defined with respect to the Fock space and comprising thermodynamic mixtures of phases in situations of phase separation).

Next, since $\hat{H}_A$ of (9.2) has an affine-linear dependence on the external four-potential A^μ, *$E[A,Q]$ for fixed Q is a concave functional of A^μ:* Fix Q, pick A_1, A_2 for which the $\hat{H}_{A_i}$ are bounded below and let $0 < \alpha < 1$. Then, obviously $\hat{H}_{\alpha A_1 + (1-\alpha)A_2}$ is also bounded below, and

$$\begin{aligned}
&E[\alpha A_1 + (1-\alpha)A_2, Q] = \\
&= \min_\Psi \Big\{ \alpha\langle\Psi|\hat{H}_{A_1}|\Psi\rangle + (1-\alpha)\langle\Psi|\hat{H}_{A_2}|\Psi\rangle \,\Big|\, \langle\Psi|\hat{Q}|\Psi\rangle = Q \Big\} \geq \\
&\geq \alpha \min_{\Psi'} \Big\{ \langle\Psi'|\hat{H}_{A_1}|\Psi'\rangle \,\Big|\, \langle\Psi'|\hat{Q}|\Psi'\rangle = Q \Big\} + \\
&\quad + (1-\alpha) \min_{\Psi''} \Big\{ \langle\Psi''|\hat{H}_{A_2}|\Psi''\rangle \,\Big|\, \langle\Psi''|\hat{Q}|\Psi''\rangle = Q \Big\} = \\
&= \alpha E[A_1,Q] + (1-\alpha)E[A_2,Q].
\end{aligned}$$

This result comes again from the simple fact that the minimum of a sum cannot be lower than the sum of the independent minima.

Considering further the gauge property (9.21), we may finally define the current density functional (in the sector of electron excess $Q \leq 0$) along the

same lines as $H[n]$ was introduced in Section 6.2: The convexity of $E[A,Q]$ with respect to Q gives rise to a pair of mutual Legendre transforms

$$\tilde{G}[A,\zeta] = \sup_Q\{\zeta Q - E[A,Q]\}, \quad E[A,Q] = \sup_\zeta\{Q\zeta - \tilde{G}[A,\zeta]\}.$$

where ζ has obviously the meaning of an electrochemical potential. The gauge property (9.21) implies $E[A,Q] - \zeta Q = E[A^\mu - \delta_0^\mu(\zeta/c), Q]$ and hence allows for a writing $\tilde{G}[A,\zeta] = \tilde{G}[A^\mu - \delta_0^\mu(\zeta/c), 0] = G[A^\mu - \delta_0^\mu(\zeta/c)]$ with

$$G[A] = -\inf_Q E[A,Q], \quad E[A,Q] = \sup_\zeta\{Q\zeta - G[A^\mu - \delta_0^\mu(\zeta/c)]\}. \tag{9.22}$$

Like $-E$, G is convex in A, hence another pair of mutual Legendre transforms is

$$H[J] = \sup_A\{(J|A) - G[A]\}, \quad G[A] = \sup_J\{(A|J) - H[J]\}, \tag{9.23}$$

where the notation

$$(J|A) = e\int d^3r\, J^\mu(\boldsymbol{r})A_\mu(\boldsymbol{r}) \tag{9.24}$$

is introduced. Substituting the second equation (9.23) into the second equation (9.22) yields

$$\begin{aligned} E[A,Q] &= \sup_\zeta\Big\{Q\zeta - \sup_J\{(A^\mu - \delta_0^\mu(\zeta/c)|J) - H[J]\}\Big\} = \\ &= \sup_\zeta\Big\{Q\zeta + \inf_J\{H[J] - (A^\mu - \delta_0^\mu(\zeta/c)|J)\}\Big\} \le \\ &\le \inf_J\Big\{H[J] - (A|J) + \sup_\zeta\big(Q + (\delta_0^\mu(1/c)|J)\big)\zeta\Big\}. \end{aligned}$$

Substitution of the first relation (9.22) into the first relation (9.23) yields

$$H[J] \stackrel{\text{def}}{=} \inf_{Q\le 0}\sup_A\left\{E[A,Q] + e\int d^3r\, J^\mu A_\mu\right\} = \inf_{Q\le 0} F[J,Q] \tag{9.25}$$

where the inversion of sup inf into inf sup may be justified in a manner analogous to the non-relativistic case. Also the definition of F and the sharpening of the inequality in the above expression for $E[A,Q]$ into an

equality is obtained analogously. The Hohenberg-Kohn variational principle now reads

$$E[A,Q] = \inf \left\{ H[J] - e \int d^3r\, J^\mu A_\mu \;\middle|\; J \in X,\;\; -\frac{e}{c} \int d^3r\, J^0 = Q \right\}. \tag{9.26}$$

Of course, $H[J]$ can be known only to that degree of approximation to which the problem (9.11) can be treated. The functional space X for the variation of J and its dual X^* which A has to belong to must still be defined.

In the non-relativistic theory, the kinetic energy is $T_{\rm nr} = \langle\Phi| - \sum \boldsymbol{\nabla}_i^2/2|\Phi\rangle$, and $T < +\infty$ implies $n \in \boldsymbol{L}^3(\boldsymbol{T}^3)$ for the density n via Sobolev's inequality (6.33). The relativistic kinetic energy is (in the Schrödinger representation for Φ) $T_{\rm r} = \langle\Phi| \sum ic\boldsymbol{\alpha} \cdot \boldsymbol{\nabla}_i|\Phi\rangle$, and one is tempted to suppose $J^\mu \in \boldsymbol{L}^{3/2}(\boldsymbol{T}^3)$. If for instance a Dirac bispinor orbital behaves like $\phi \sim r^{-s}$, then $J^\mu \in \boldsymbol{L}^{3/2}(\boldsymbol{T}^3)$ implies $s < 1$ for which indeed $T_{\rm r} < +\infty$ (see for instance [Landau and Lifshitz, 1982, Chapter IV]). Hence $A^\mu \in \boldsymbol{L}^3(\boldsymbol{T}^3)$ must be demanded as $\boldsymbol{L}^3(\boldsymbol{T}^3)$ is the dual of $\boldsymbol{L}^{3/2}(\boldsymbol{T}^3)$. For a potential $A^0 \sim r^{-s}$ this implies again $s < 1$ and hence excludes Coulomb potentials although it permits to treat them as a limiting case.

In a relativistic theory Coulomb potentials are a touchy case anyway. First of all, $\hat{H}_A$ would not be bounded below for $A^0 \sim r^{-s}$ with $s > 1$: an unlimited number of electrons (or positrons, depending on the sign of A_0) would plunge into the potential center as their mutual Coulomb repulsion could not compensate the attraction for small enough r. For $s = 1$, in the case $cA_0 = Z/r$, $Z > c$ a.u., a single electron would again slip down the potential thereby reducing Z effectively by one charge unit, and this process would continue until $Z_{\rm eff} < c$ a.u. The formal workaround with $A^\mu \in \boldsymbol{L}^3(\boldsymbol{T}^3)$ is to cut off Coulomb potentials at a nuclear radius $r_0 > 0$, in the course of which for $Z \leq 137$ the limes $r_0 \to 0$ may be taken. In truth QED is inherently inconsistent anyway unless the Coulomb interaction is cut off at very small radii by the asymptotic freedom of electroweak theory [Berestetskii, 1976]. (For the stability range of a slight caricature of the Hamiltonian $\hat{H}_A$ see Refs. [Lieb and Yau, 1988, Lieb *et al.*, 1996].)

A practically tractable case is an interaction-free electron-positron system in the external field, described by

$$\hat{\mathcal{H}}^0 = \; c : \hat{\bar{\psi}}(-i\boldsymbol{\gamma} \cdot \nabla + m_0 c)\hat{\psi} : . \tag{9.27}$$

(The difference of masses $m - m_0$ amounts to mass renormalization caused by interaction, which reduces in the non-relativistic case to the exclusion of

self-interaction.) We again denote the pendant of the functional $F[J,Q]$ for this case by $T[J,Q]$, and find for $Q=-eN$ the corresponding $E^0[A,Q]$ as the sum of the lowest N eigenvalues of the Dirac equation

$$(-ic\boldsymbol{\alpha}\cdot\nabla+\beta m_0c^2-ec\beta\gamma^\mu A_\mu)\psi_k=\psi_k\varepsilon_k \tag{9.28}$$

corresponding to solutions with large components in the electron sector (first two components of the bispinor in the usual representation (8.56) of the γ-matrices). The corresponding four-current density is

$$J^\mu=c\sum_k n_k\bar{\psi}_k\gamma^\mu\psi_k,\quad \langle\psi_k|\psi_{k'}\rangle=\delta_{kk'},\quad Q=-e\sum_k n_k. \tag{9.29}$$

Eventually, partial occupation of the HOMO is understood, if N is not integer.

For a non-interacting particle field in a static external potential no pair-processes take place, and therefore wavefunctions may be used. Therefore, the variational result (9.28, 9.29) is obtained in complete analogy to the non-relativistic case as described in Section 4.2 with the relation to the general T-functional as discussed in Section 6.4. Moreover, in this case no coherent, Bose condensed internal fields appear and hence no separate treatment of the coherent and fluctuating parts of the fields as introduced in the renormalization procedure of last section is needed. In order to directly compare with the interacting case, the full current $\hat{j}^\mu$ via the full field operator $\hat{\psi}$ had been retained in the first term of (9.18).

In the general case of interacting electrons, (9.29) is again used as a parameterization of the variational four-current J^μ in terms of Kohn-Sham-Dirac spinor orbitals ψ_k and orbital occupation numbers n_k.

In essentially all practical cases the normal order of (9.3, 9.4) demands the Dirac spinor orbitals ψ_k to be electron orbitals which according to the remark after (9.5) are orbitals which develop continuously from the free-electron continuum when the external potential A^μ is continuously switched on. It is this restriction which makes the Kohn-Sham-Dirac Hamiltonian bounded below.

The four-current density functional $H[J]$ is again split into an orbital variation part K and a density integral L:

$$\begin{aligned}
H[J]&=K[J]+L[J],\\
K[J]&=\min_{\psi_k,n_k}\Big\{k[\psi_k,n_k]\;\Big|\;J^\mu=c\sum_k n_k\bar{\psi}_k\gamma^\mu\psi_k,\;(\psi_k|\psi_{k'})=\delta_{kk'}\Big\},\\
L[J]&=\int d^3r\;J^0(\boldsymbol{r})\;l(J^\mu(\boldsymbol{r}),\nabla J^\mu(\boldsymbol{r}),\ldots).
\end{aligned} \tag{9.30}$$

Substitution into (9.26) yields the Kohn-Sham-Dirac variational problem

$$E[A,Q] = \min_{\psi_k, n_k} \Bigg\{ k[\psi_k, n_k] + L\Big[c \sum_k n_k \bar{\psi}_k \gamma^\mu \psi_k\Big] - \\ - ec \sum_k n_k (\psi_k | \gamma^0 \gamma^\mu A_\mu | \psi_k) \;\Bigg|\; (\psi_k|\psi_{k'}) = \delta_{kk'},\ -e \sum_k n_k = Q \Bigg\} \quad (9.31)$$

which were equivalent to (9.26) would not the local ansatz for $L[J]$ already introduce an approximation.

An alternative, still potentially exact splitting of $H[J]$ is

$$H[J] = K[J] - \frac{e}{2} \int d^3r\, J^\mu a_\mu + E_{\mathrm{XC}}[J], \quad (9.32)$$

where

$$k[\psi_k, n_k] = \sum_k n_k \langle \psi_k | \beta(-ic\boldsymbol{\gamma} \cdot \boldsymbol{\nabla} + c^2) | \psi_k \rangle \quad (9.33)$$

and a^μ is linearly related to J^μ by the first equation (9.14), so that all terms of (9.30) except of E_{XC} have been defined previously and (9.30) thus defines E_{XC}, the variation of (9.26) with the general *ansatz* (9.29) for the four-current density leads to the most general Kohn-Sham-Dirac equation [Eschrig *et al.*, 1985], cf. also [Dreizler and Gross, 1990]

$$\left[-ic\boldsymbol{\alpha} \cdot \nabla + \beta m_0 c^2 - ec\beta\gamma^\mu (A_\mu + a_\mu + a_\mu^{\mathrm{XC}})\right] \psi_k = \psi_k \varepsilon_k, \quad (9.34)$$

with the Kohn-Sham exchange and correlation four-potential

$$-e a_\mu^{\mathrm{XC}} \stackrel{\mathrm{def}}{=} \frac{\delta E_{\mathrm{XC}}[J]}{\delta J^\mu}. \quad (9.35)$$

Given a functional $E_{\mathrm{XC}}[J]$, this Kohn-Sham-Dirac equation in connection with the ansatz (9.29) is to be solved self-consistently to find the ground state energy and the ground state four-current density.

Recall that, if quantum electrodynamics would be taken literally, the amount of mass renormalization $m - m_0$ would be infinite due to an infinite self-interaction of an electron via its own electromagnetic field. This basic field-theoretic defect prevents an *ab-initio* treatment of mass renormalization. In lack of anything better one substitutes m_0 by the phenomenological electron mass m and simultaneously excludes the static self-interaction contributions from $E_{\mathrm{XC}}[J]$ as in the non-relativistic theory.

The gauge invariance properties of this Kohn-Sham-Dirac equation are

$$\boldsymbol{A}+\boldsymbol{a}+\boldsymbol{a}^{\mathrm{XC}} \longrightarrow \boldsymbol{A}+\boldsymbol{a}+\boldsymbol{a}^{\mathrm{XC}}+\nabla\chi, \quad \psi_k \longrightarrow \psi_k \exp(-ie\chi) \tag{9.36}$$

and

$$A_0+a_0+a_0^{\mathrm{XC}} \longrightarrow A_0+a_0+a_0^{\mathrm{XC}}+w, \quad \varepsilon_k \longrightarrow \varepsilon_k - ecw \tag{9.37}$$

with an arbitrary function $\chi(\boldsymbol{r})$ and an arbitrary constant w. (The latter gauge transformation fits into the former frame by writing $\psi_k(t) \longrightarrow \psi_k(t)\exp(iecwt)$.)

9.3 The Gordon Decomposition and Spin Density

The four-current density (9.29) may be decomposed in a manner allowing for a physical interpretation [Gordon, 1928]. This decomposition was introduced in the first relativistic formulation of density functional theory by [Rajagopal and Callaway, 1973]. It starts with use of certain algebraic properties of the γ-matrices:

$$[\gamma^\mu,\gamma^\nu]_+ = 2g^{\mu\nu}, \qquad [\gamma^\mu,\gamma^\nu]_- = -2i\sigma^{\mu\nu}, \tag{9.38}$$

where the anticommutator is proportional to the unit matrix (as previously, for every pair $\mu\nu$ of tensor indices, the tensor elements $g^{\mu\nu}$ and $\sigma^{\mu\nu}$ are 4×4-matrices acting in the bispinor space), and the commutator defines an antisymmetric form, the elements of which are

$$\sigma^{0k} = i\alpha^k, \qquad \sigma^{jk} = \epsilon^{jkl}\Sigma_l, \qquad \Sigma_l = \begin{pmatrix} \sigma_l & 0 \\ 0 & \sigma_l \end{pmatrix}. \tag{9.39}$$

σ_l are Pauli's 2×2 spin matrices as before.

If we now write the Dirac equation (for the sake of simplicity we take (9.28), because we need only its general structure) in its covalent form, replacing the energy term by the time-derivative,

$$[\gamma^\nu(i\partial_\nu + eA_\nu) - mc]\psi = 0 = \bar{\psi}[(-i\overleftarrow{\partial}_\nu + eA_\nu)\gamma^\nu - mc], \tag{9.40}$$

we find

$$\gamma^\mu\psi = \frac{1}{mc}\gamma^\mu\gamma^\nu(i\partial_\nu + eA_\nu)\psi = \frac{1}{mc}(g^{\mu\nu} - i\sigma^{\mu\nu})(i\partial_\nu + eA_\nu)\psi, \tag{9.41}$$

$$\bar{\psi}\gamma^\mu = \frac{1}{mc}\bar{\psi}(-i\overleftarrow{\partial}_\nu + eA_\nu)(g^{\nu\mu} - i\sigma^{\nu\mu}). \tag{9.42}$$

Adding both expressions, multiplied by $\bar{\psi}$ and ψ, respectively, yields

$$\bar{\psi}\gamma^\mu\psi = \frac{1}{2mc}\bar{\psi}(-i\overleftarrow{\partial}^\mu + i\partial^\mu + 2eA^\mu)\psi + \frac{1}{2mc}\partial_\nu\bar{\psi}\sigma^{\mu\nu}\psi. \tag{9.43}$$

For the three-current density, this result may be written as

$$\boldsymbol{J} = \boldsymbol{I} + \frac{1}{m}\nabla\times\boldsymbol{S} + \frac{\partial\boldsymbol{G}}{\partial t}, \tag{9.44}$$

with the orbital current density

$$\boldsymbol{I} = \frac{1}{2m}\bar{\psi}(-i\nabla + i\overleftarrow{\nabla} + 2e\boldsymbol{A})\psi, \tag{9.45}$$

the spin current density, derived from the spin density

$$\boldsymbol{S} = \frac{1}{2}\bar{\psi}\boldsymbol{\Sigma}\psi, \tag{9.46}$$

and a relativistic correction term, derived from

$$\boldsymbol{G} = -\frac{i}{2mc}\bar{\psi}\boldsymbol{\alpha}\psi. \tag{9.47}$$

In the stationary situation considered here this last term vanishes and we are left with the sum of orbital and spin current densities.

The total stationary current density must have zero divergence due to charge conservation. Since the divergence of the spin current density vanishes by its very structure as a curl, the orbital current density must also be divergence free:

$$\nabla\boldsymbol{I} = 0. \tag{9.48}$$

Since further $\boldsymbol{I}$ either vanishes at infinity or is periodic, it may likewise be expressed as a curl of some vector field $\boldsymbol{L}$, to be visualized as an 'angular momentum density':

$$\boldsymbol{I} = \frac{1}{2m}\nabla\times\boldsymbol{L}. \tag{9.49}$$

(Recall that an orbital angular momentum *density* cannot really figure in quantum mechanics because position and momentum cannot have sharp values at the same time. Accordingly, (9.49) defines $\boldsymbol{L}$ only up to an arbitrary additive gradient term.)

The total electric current density may then be expressed as

$$-e\boldsymbol{J} = \nabla \times \boldsymbol{M} = -\frac{e}{2m}\nabla \times (\boldsymbol{L} + 2\boldsymbol{S}). \tag{9.50}$$

$\boldsymbol{M}$ has the dimension of a 'magnetization density', related in a non-renormalized way to the angular momenta by the Bohr magneton. (Note that except for the non-relativistic case, the decomposition of the current given above is formal, which is also indicated by the appearance of ideal gyromagnetic factors. Moreover, what was said above on the 'angular momentum density' refers likewise to the orbital part of the 'magnetization density'.)

The four-current density is now given by $J^\mu = (nc, -\nabla \times \boldsymbol{M}/e)$, and hence the functional $E_{\mathrm{XC}}[J]$ may be rewritten as a functional $E_{\mathrm{XC}}[n, \boldsymbol{M}]$ with the mechanical exchange and correlation potential acting on an electron

$$v^{\mathrm{XC}} = -eu^{\mathrm{XC}} = -eca_0^{\mathrm{XC}} = \frac{\delta E_{\mathrm{XC}}}{\delta n} \tag{9.51}$$

and, expressing formally the curl with the help of (9.50) as a functional derivative of $\boldsymbol{J}$ with respect to $\boldsymbol{M}$, with a magnetic exchange and correlation field[12]

$$\mu_0 \boldsymbol{h}^{\mathrm{XC}} = \nabla \times \boldsymbol{a}^{\mathrm{XC}} = \int d^3r' \, \frac{\delta \boldsymbol{J}(\boldsymbol{r}')}{\delta \boldsymbol{M}} \frac{\delta E_{\mathrm{XC}}}{\delta \boldsymbol{J}(\boldsymbol{r}')} = \frac{\delta E_{\mathrm{XC}}}{\delta \boldsymbol{M}}. \tag{9.52}$$

The interaction terms $-eJ^\mu A_\mu$ are cast into

$$-eJ^0 A_0 = nV \tag{9.53}$$

and

$$-eJ^k A_k = e\boldsymbol{J}\cdot\boldsymbol{A} = -\boldsymbol{A}\cdot\nabla\times\boldsymbol{M} \mathrel{\hat{=}} -\boldsymbol{M}\cdot\nabla\times\boldsymbol{A} = -\mu_0 \boldsymbol{H}\cdot\boldsymbol{M}. \tag{9.54}$$

The sign $\hat{=}$ again means equivalence under the spatial integral after integration by parts. V and $\boldsymbol{H}$ are the mechanical potential and the magnetic field corresponding to A^μ.

The Kohn-Sham-Dirac equation (9.34) was obtained by varying (9.26) with respect to $\bar{\psi}_k$. From (9.50) and (9.46) we find

$$\frac{\delta \boldsymbol{M}(\boldsymbol{r})}{\delta \bar{\psi}_k(\boldsymbol{r}')} = -\frac{e}{2m}\left(\frac{\delta \boldsymbol{L}(\boldsymbol{r})}{\delta \bar{\psi}_k(\boldsymbol{r}')} + \delta(\boldsymbol{r}' - \boldsymbol{r})\boldsymbol{\Sigma}\psi_k(\boldsymbol{r})\right). \tag{9.55}$$

[12] Writing (9.50) formally as $-e\boldsymbol{J}(\boldsymbol{r}') = \int d^3r \, \delta(\boldsymbol{r}'-\boldsymbol{r})\boldsymbol{\nabla}\times\boldsymbol{M}(\boldsymbol{r})$, one immediately finds $-e\delta\boldsymbol{J}(\boldsymbol{r}')/\delta\boldsymbol{M}(\boldsymbol{r}) = \delta(\boldsymbol{r}' - \boldsymbol{r})\boldsymbol{\nabla}\times$.

This allows for an alternative way of writing down the general Kohn-Sham-Dirac equation:

$$\begin{aligned}&\left[-ic\boldsymbol{\alpha}\cdot\nabla+\beta mc^2+V(\boldsymbol{r})+v(\boldsymbol{r})+v^{\mathrm{XC}}(\boldsymbol{r})\right]\,\psi_k(\boldsymbol{r})-\\&-\mu_0\beta\int d^3r'\,\left(\boldsymbol{H}(\boldsymbol{r}')+\boldsymbol{h}(\boldsymbol{r}')+\boldsymbol{h}^{\mathrm{XC}}(\boldsymbol{r}')\right)\,\cdot\,\frac{\delta\boldsymbol{M}(\boldsymbol{r}')}{\delta\bar{\psi}_k(\boldsymbol{r})}\;=\;\psi_k(\boldsymbol{r})\varepsilon_k.\end{aligned}\tag{9.56}$$

At the price of replacing the four-current, local in terms of the Kohn-Sham orbitals ψ_k, by a 'magnetization density' $\boldsymbol{M}$, non-locally depending on the ψ_k and whose orbital part $\boldsymbol{L}$ is subject to another gauge (undetermined gradient term), the vector potential with its unpleasant far-ranging character has been eliminated. Note that for a homogeneous electron liquid in a homogeneous magnetic field the orbital current density is zero except close to the boundary of the volume while the vector potential $\boldsymbol{A}(\boldsymbol{r})$ increases linearly with increasing $\boldsymbol{r}$.

The crucial problem remaining to be solved is to find a suitable expression for $\delta\boldsymbol{L}(\boldsymbol{r})/\delta\bar{\psi}_k(\boldsymbol{r}')$. Whether a quasi-local functional expression $\boldsymbol{L}[\psi_k]$ can be found or not, obviously depends on the choice of the undetermined gradient term contained in $\boldsymbol{L}$, since in the just mentioned example of a homogeneous situation this is the only term of $\boldsymbol{L}$ present at finite $\boldsymbol{r}$.

9.4 Approximative Variants

If we completely neglect the orbital current $\boldsymbol{I}$, then we find from (9.55)

$$\frac{\delta\boldsymbol{M}(\boldsymbol{r})}{\delta\bar{\psi}_k(\boldsymbol{r}')}=-\frac{e}{2m}\boldsymbol{\Sigma}\psi_k\,\delta(\boldsymbol{r}-\boldsymbol{r}')\quad\text{for}\quad\boldsymbol{I}=0.\tag{9.57}$$

In this case, the Kohn-Sham-Dirac equation acquires the simple form

$$\begin{aligned}&\Big[-ic\boldsymbol{\alpha}\cdot\nabla+\beta mc^2+V+v+v^{\mathrm{XC}}+\\&\qquad+\frac{e\mu_0}{2m}\beta\boldsymbol{\Sigma}\cdot(\boldsymbol{H}+\boldsymbol{h}+\boldsymbol{h}^{\mathrm{XC}})\Big]\,\psi_k\;=\;\psi_k\varepsilon_k,\end{aligned}\tag{9.58}$$

where the magnetic field couples to the spin only. As in many applications (however by far not always) the influence of the orbital current is small, this form is widely used in computations.

Without the spin-dependent $\boldsymbol{\Sigma}$-term, the Kohn-Sham-Dirac equation was investigated by [Rajagopal, 1978] and by [MacDonald and Vosko, 1979].

The $\boldsymbol{\Sigma}$-term with the exchange and correlation field $\boldsymbol{h}^{\mathrm{XC}}$, first introduced by [Rajagopal and Callaway, 1973], has been essential already in the non-relativistic theory where the exchange and correlation field appears as the difference of the Kohn-Sham potential for spin up and down electrons, respectively (cf. Section 4.7 as well as Sections 7.1 and 7.2).

In the non-relativistic LDA for collinear spin structures, the exchange and correlation field is given by

$$\frac{e\mu_0}{m}\boldsymbol{h}^{\mathrm{XC,LDA}}(n(\boldsymbol{r}),\zeta(\boldsymbol{r})) = v^{\mathrm{LDA}}_{\mathrm{XC}+}(n(\boldsymbol{r}),\zeta(\boldsymbol{r})) - v^{\mathrm{LDA}}_{\mathrm{XC}-}(n(\boldsymbol{r}),\zeta(\boldsymbol{r})) \quad (9.59)$$

with $v^{\mathrm{LDA}}_{\mathrm{XC}\pm}$ obtained from $n(\boldsymbol{r})$ and $\Sigma(\boldsymbol{r}) = n(\boldsymbol{r})\zeta(\boldsymbol{r})$ via (7.23). Note that, by the very nature it enters the theory via (9.52), the exchange and correlation field has to be divergence-free, which is not provided by the LDA. The consequences of this defect of the spin-dependent LDA have not yet been investigated.

In the non-relativistic limit, spin and orbital currents decouple, so that the exchange and correlation energy functional may separately depend on both. This situation has been analyzed in [Vignale and Rasolt, 1987, Vignale and Rasolt, 1988]. Those authors introduce the 'paramagnetic current density' (of nonrelativistic spin-orbitals $\psi(x)$)

$$\boldsymbol{I}_p(x) \stackrel{\mathrm{def}}{=} \frac{1}{2m}\psi^*(x)(-i\nabla + i\overleftarrow{\nabla})\psi(x) \quad (9.60)$$

instead of the physical orbital current density (9.45). The current density $\boldsymbol{I}_p$ is not gauge invariant. Using gauge arguments, it has been shown in the above cited papers, that the exchange and correlation energy functional can depend on $\boldsymbol{I}_p$ only via the 'vorticity'

$$\boldsymbol{\nu}(x) \stackrel{\mathrm{def}}{=} \nabla \times \left(\frac{\boldsymbol{I}_p(x)}{n(x)}\right). \quad (9.61)$$

Up to second order in $\boldsymbol{I}_p$, a local current density approximation

$$E_{\mathrm{XC}}[n,\boldsymbol{\nu}] = E_{\mathrm{XC}}[n,0] + \int dx \left(\frac{9\pi}{4}\right)^{1/3} \frac{1}{24\pi^2 r_s}\left(\frac{\chi_L}{\chi_L^0} - 1\right)|\boldsymbol{\nu}(x)|^2 \quad (9.62)$$

has been found, where $r_s = r_s(x)$ and

$$\left(\frac{\chi_L}{\chi_L^0} - 1\right) = 0.027\,64 r_s \ln r_s + 0.014\,07 r_s + O(r_s^2 \ln r_s) \quad (9.63)$$

is the ratio of the diamagnetic susceptibilities for the interacting and non-interacting homogeneous electron systems minus one (in the high-density limit). The use of the paramagnetic current density also slightly modifies the Kohn-Sham equations. (Cf. also [Trickey, 1990, p. 235–254].)

Later on [Skudlarski and Vignale, 1993], the exchange and correlation energy $E_{\mathrm{XC}}(n, \zeta, \boldsymbol{H})$ of a homogeneous electron liquid in an arbitrarily strong homogeneous magnetic field $\boldsymbol{H}$ has been found numerically. Thereby, the degree of spin polarization ζ was varied independently via a varying Zeeman coupling constant g between spin and field. A more general local current density approximation

$$\begin{aligned} E_{\mathrm{XC}}[n, \zeta, \boldsymbol{\nu}] &= \\ &= \int d^3r\, n(\boldsymbol{r}) E_{\mathrm{XC}}(n(\boldsymbol{r}), \zeta(\boldsymbol{r}), \boldsymbol{H}(\boldsymbol{r}) = -mc|\boldsymbol{\nu}(\boldsymbol{r})|/e\mu_0) \end{aligned} \tag{9.64}$$

has been suggested. Experience still must be collected with those approximations.

Alternatively, a non-relativistic current density functional approach using the physical orbital current density instead of the paramagnetic one has been proposed in [Diener, 1991]. However, explicit expressions have not yet been tried.

A largely unsolved problem regarding magnetically polarized ground states is the orbital polarization contribution of tightly bound d- and f-electrons to the exchange and correlation energy functional. (For nearly free electrons in a solid at moderate field strength orbital polarization is quenched by kinetic energy.) Take some complete set ϕ_{nlm} of atomic-like orbitals and use the ansatz

$$\begin{aligned} \boldsymbol{L}(\boldsymbol{r}) \;\approx\; & \sum_k^{\text{occ.}} \int d^3r'\, \bar{\psi}_k(\boldsymbol{r}') \;* \\ & * \sum_{nlm} \left\{ \delta(\boldsymbol{r} - \boldsymbol{r}') \hat{\boldsymbol{L}}' - \nabla \lambda_{nlm}(\boldsymbol{r}, \boldsymbol{r}') \right\} \phi_{nlm}(\boldsymbol{r}')\, (\phi_{nlm}|\psi_k), \end{aligned} \tag{9.65}$$

where $\hat{\boldsymbol{L}}'$ is the non-relativistic orbital angular momentum operator, so that the spatial integral over $\boldsymbol{L}(\boldsymbol{r})$ gives the total non-relativistic orbital angular momentum for any choice of the λ_{nlm}. With this ansatz one would find expressions of the type

$$\int d^3r\, \boldsymbol{h}(\boldsymbol{r}) \cdot \frac{\delta \boldsymbol{L}(\boldsymbol{r})}{\delta \bar{\psi}_k(\boldsymbol{r}')} = f_{nlm}(\boldsymbol{r}')(\phi_{nlm}|\psi_k), \tag{9.66}$$

where one could try to adjust the f_{nlm} phenomenologically to standard situations. This would loosely give a justification of the orbital polarization corrections first introduced in [Brooks, 1985]. Note, however, that (9.65) is an approximation because it does not locally fulfill $\boldsymbol{\nabla} \times \boldsymbol{L} = 2m\boldsymbol{I}$.

Bibliography

[Ando, 1963] T. Ando, "Properties of Fermion Density Matrices," Rev. Mod. Phys. **35**, 690–702 (1963).

[Anisimov *et al.*, 1997] V. I. Anisimov, F. Aryasetiawan, and A. I. Lichtenstein, "First-principles calculations of the electronic structure and spectra of strongly correlated systems: the LDA+U method," J. Phys.: Condens. Matter **9**, 767–808 (1997).

[Berestetskii, 1976] V. B. Berestetskii, "Zero Mass and Asymptotic Freedom," Uspekhi Fizicheskich Nauk (in russian) **120**, 439–454 (1976).

[Berezin, 1965] F. A. Berezin, *The Method of Second Quantization* (Academic Press, New York, 1965).

[Born and Huang, 1968] M. Born and K. Huang, *Dynamical Theory of Crystal Lattices* (Clarendon Press, Oxford, 1968).

[Brooks, 1985] M. S. S. Brooks, "Calculated Ground State Properties of Light Actinide Metals and their Compounds," Physica **130B**, 6–12 (1985).

[Carr and Maradudin, 1964] W. J. Carr, Jr. and A. A. Maradudin, "Ground-state Energy of a High-Density Electron Gas," Phys. Rev. **133**, A371–A374 (1964).

[Ceperly and Alder, 1980] D. M. Ceperly and B. J. Alder, "Ground State of the Electron Gas by a Stochastic Method," Phys. Rev. Lett. **45**, 566–569 (1980).

[Coleman, 1963] A. J. Coleman, "Structure of Fermion Density Matrices," Rev. Mod. Phys. **35**, 668–689 (1963).

[Cook, 1953] J. M. Cook, "The Mathematics of Second Quantisation," Trans. Am. Math. Soc. **74**, 224–245 (1953).

[Czyżyk and Sawatzky, 1994] M. T. Czyżyk and G. A. Sawatzky, "Local-density functional and on-site correlations: The electronic structure of La_2CuO_4 and $LaCuO_3$," Phys. Rev. **B49**, 14211–14228 (1994).

[Dawydow, 1987] A. S. Dawydow, *Quantenmechanik*, 7th ed. (VEB Deutscher Verlag der Wissenschaften, Berlin, 1987).

[Diener, 1991] G. Diener, "Current-Density-Functional Theory for a Non-Relativistic Electron Gas in a Strong Magnetic Field," J. Phys.: Condens. Matter **3**, 9417–9428 (1991).

[Dirac, 1930] P. A. M. Dirac, "Note on Exchange Phenomena in the Thomas Atom," Proc. Camb. Phil. Soc. **26**, 376–385 (1930).

[Dirac, 1958] P. A. M. Dirac, *Principles of Quantum Mechanics*, 4th ed. (Clarendon Press, Oxford, 1958).

[Dreizler and da Providencia, 1985] *Density Functional Methods in Physics, Proceedings of a NATO ASI Held in Alcabideche, Portugal, September 1983*, edited by R. M. Dreizler and J. da Providencia (Plenum Press, New York, 1985).

[Dreizler and Gross, 1990] R. M. Dreizler and E. K. U. Gross, *Density Functional Theory* (Springer-Verlag, Berlin, Heidelberg, New York, 1990).

[Dyson and Lenard, 1967] F. J. Dyson and A. Lenard, "Stability of Matter. I.," J. Math. Phys. **8**, 423–434 (1967).

[Englisch and Englisch, 1984a] H. Englisch and R. Englisch, "Exact Density Functionals for Ground-State Energies. I General Results," phys. stat. sol. (b) **123**, 711–721 (1984).

[Englisch and Englisch, 1984b] H. Englisch and R. Englisch, "Exact Density Functionals for Ground-State Energies. II Details and Remarks," phys. stat. sol. (b) **124**, 373–379 (1984).

[Erdahl and Smith, 1987] *Density Matrices and Density Functionals, Proceedings of the Symposium in Honor of John Coleman Held in Kingston, Ontario, August 1985*, edited by R. Erdahl and V. H. Smith, Jr. (Reidel Publishing Company, Dordrecht, 1987).

[Eschrig *et al.*, 2003a] H. Eschrig, K. Koepernik, and I. Chaplygin, "Density Functional Application to Strongly Correlated Electron Systems," J. Solid St. Chem. (Special Issue 2003).

[Eschrig and Pickett, 2001] H. Eschrig and W. Pickett, "Density functional theory of magnetic systems revisited," Solid St. Commun **118**, 123–127 (2001).

[Eschrig *et al.*, 2003b] H. Eschrig, M. Richter, and I. Opahle, in *Relativistic Electronic Structure Theory—Part II: Applications*, edited by P. Schwerdtfeger (Elsevier, Amsterdam, 2003), Chap. 12, Relativistic Solid State Calculations.

[Eschrig *et al.*, 1985] H. Eschrig, G. Seifert, and P. Ziesche, "Current Density Functional Theory of Quantum Electrodynamics," Solid St. Commun. **56**, 777–780 (1985).

[Ewald, 1921] P. P. Ewald, "Die Berechnung Optischer und Elektrostatischer Gitterpotentiale," Ann. Physik (Leipzig) **64**, 253–287 (1921).

[Fermi, 1927] E. Fermi, "Un metodo statistico per la determinazione di alcune priorieta dell'atome," Rend. Accad. Naz. Lincei **6**, 602–607 (1927).

[Feynman, 1939] R. P. Feynman, "Forces in Molecules," Phys. Rev. **56**, 340–343 (1939).

[Fuchs, 1935] K. Fuchs, "A Quantum Mechanical Investigation of the Cohesive Forces of Metallic Copper," Proc. Roy. Soc. **A151**, 585–602 (1935).

[Gáspár, 1954] R. Gáspár, "Über eine Approximation des Hartree-Fockschen Potentials durch eine universelle Potentialfunktion," Acta Phys. Acad. Sci. Hung. **3**, 263–286 (1954).

[Gell-Mann and Low, 1951] M. Gell-Mann and F. Low, "Bound States in Quantum Field Theory," Phys. Rev. **84**, 350–353 (1951).

[Gilbert, 1975] T. L. Gilbert, "Hohenberg-Kohn Theorem for Nonlocal External Potentials," Phys. Rev. **B12**, 2111–2120 (1975).

[Gordon, 1928] W. Gordon, "Der Strom der Diracschen Elektronentheorie," Z. Physik **50**, 630–632 (1928).

[Grabo and Gross, 1995] T. Grabo and E. K. U. Gross, "Density-Functional Theory Using an Optimized Exchange-Correlation Potential," Chem. Phys. Lett. **240**, 141–150 (1995).

[Gross and Dreizler, 1995] *Density Functional Theory, Proceedings of a NATO ASI Held in Il Ciocco, Italy, August 1993*, edited by E. K. U. Gross and R. M. Dreizler (Plenum Press, New York, 1995).

[Gunnarsson *et al.*, 1979] O. Gunnarsson, M. Jonson, and B. I. Lundqvist, "Description of Exchange and Correlation Effects in Inhomogeneous Electron Systems," Phys. Rev. **B20**, 3136–3164 (1979).

[Gunnarsson and Lundqvist, 1976] O. Gunnarsson and B. I. Lundqvist, "Exchange and Correlation in Atoms, Molecules and Solids by the Spin-Density Formalism," Phys. Rev. **B13**, 4274–4298 (1976).

[Harriman, 1980] J. E. Harriman, "Orthonormal Orbitals for the Representation of an Arbitrary Density," Phys. Rev. **A24**, 680–682 (1980).

[Hedin, 1965] L. Hedin, "New Method for Calculating the One-Particle Green's Function with Application to the Electron-Gas Problem," Phys. Rev. **139**, A796–A823 (1965).

[Hedin and Lundqvist, 1971] L. Hedin and B. I. Lundqvist, "Explicit Local Exchange and Correlation Potentials," J. Phys. C: Solid St. Phys. **4**, 2064–2083 (1971).

[Hellmann, 1937] H. Hellmann, *Einführung in die Quantenchemie* (Deuticke, Leipzig, 1937).

[Hoffmann-Ostenhof *et al.*, 1980] M. Hoffmann-Ostenhof, T. Hoffmann-Ostenhof, R. Ahlrichs, and J. Morgan, "On the Exponential Falloff of Wave Functions and Electron Densities," in *Mathematical Problems in Theoretical Physics, Proceedings of the International Conference on Mathematical Physics held in Lausanne, Switzerland, August 20–25, 1979*, Vol. 116 of *Springer Lecture Notes in Physics*, edited by K. Osterwalder (Springer-Verlag, Berlin, Heidelberg, New York, 1980), pp. 62–67.

[Hohenberg and Kohn, 1964] P. Hohenberg and W. Kohn, "Inhomogeneous Electron Gas," Phys. Rev. **136**, B864–B871 (1964).

[Itzykson and Zuber, 1980] C. Itzykson and J.-B. Zuber, *Quantum Field Theory* (McGraw-Hill Book Company, New York, 1980).

[Janak, 1978] J. F. Janak, "Proof that $\partial E/\partial n_i = \varepsilon_i$ in Density Functional Theory," Phys. Rev. **B18**, 7165–7168 (1978).

[Kirshnits, 1957] D. A. Kirshnits, "Quantum Corrections to the Thomas-Fermi Equation," JETP **5**, 64–71 (1957).

[Kohn, 1985] W. Kohn, "Density Functional Theory: Fundamentals and Applications," in *Highlights of Condensed Matter Theory*, edited by F. Bassani, F. Fumi, and M. P. Tosi (North-Holland, Amsterdam, 1985), pp. 1–15.

[Kohn and Sham, 1965] W. Kohn and L. J. Sham, "Self-Consistent Equations Including Exchange and Correlation Effects," Phys. Rev. **140**, A1133–A1138 (1965).

[Kolmogorov and Fomin, 1970] A. Kolmogorov and S. Fomin, *Introductory Real Analysis* (Prentice Hall, Englewood Cliffs, NJ, 1970).

[Koopmans, 1934] T. Koopmans, "Über die Zuordnung von Wellenfunktionen und Eigenwerten zu den einzelnen Elektronen eines Atoms," Physica **1**, 104–113 (1934).

[Landau and Lifshitz, 1977] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics (Non-Relativistic Theory)* (Pergamon Press, Oxford, 1977).

[Landau and Lifshitz, 1982] L. D. Landau and E. M. Lifshitz, *Quantum Electrodynamics* (Pergamon Press, Oxford, 1982).

[Lang, 1965] S. Lang, *Algebra* (Addison-Wesley, Reading, Mass., 1965).

[Lannoo *et al.*, 1985] M. Lannoo, M. Schlüter, and L. J. Sham, "Calculation of the Kohn Sham Potential and its Discontinuity for a Model Semiconductor," Phys. Rev. **B32**, 3890–3899 (1985).

[Lenard and Dyson, 1968] A. Lenard and F. J. Dyson, "Stability of Matter. II.," J. Math. Phys. **9**, 698–711 (1968).

[Levy, 1982] M. Levy, "Electron Densities in Search of Hamiltonians," Phys. Rev. **A26**, 1200–1208 (1982).

[Lieb, 1976] E. H. Lieb, "The Stability of Matter," Rev. Mod. Phys. **48**, 553–569 (1976).

[Lieb, 1981] E. H. Lieb, "Thomas-Fermi and Related Theories of Atoms and Molecules," Rev. Mod. Phys. **53**, 603–641 (1981), erratum, Rev. Mod. Phys. **54**, 311 (1982).

[Lieb, 1983] E. H. Lieb, "Density Functionals for Coulomb Systems," Int. J. Quant. Chem. **XXIV**, 243–277 (1983).

[Lieb *et al.*, 1996] E. H. Lieb, M. Loss, and H. Siedentop, "Stability of Relativistic Matter via Thomas-Fermi Theory," Helv. Phys. Acta **69**, 974–984 (1996).

[Lieb and Simon, 1977] E. H. Lieb and B. Simon, "The Thomas-Fermi Theory of Atoms, Molecules and Solids," Adv. in Math. **23**, 22–116 (1977).

[Lieb and Thirring, 1975] E. H. Lieb and W. E. Thirring, "Bound for the Kinetic Energy of Fermions Which Proves the Stability of matter," Phys. Rev. Lett. **35**, 687–689 (1975).

[Lieb and Yau, 1988] E. H. Lieb and H.-T. Yau, "The Stability and Instability of Relativistic Matter," Commun. Math. Phys. **118**, 177–213 (1988).

[MacDonald and Vosko, 1979] A. H. MacDonald and S. H. Vosko, "A Relativistic Density Functional Formalism," J. Phys. C **12**, 2977–2990 (1979).

[Macke, 1950] W. Macke, "Über die Wechselwirkungen im Fermi-Gas," Z. Naturforschung **5a**, 192–208 (1950).

[Marc and McMillan, 1985] G. Marc and W. G. McMillan, "The Virial Theorem," Adv. Chem. Phys. **58**, 209–361 (1985).

[McWeeny, 1960] R. McWeeny, "Some Recent Advances in Density Matrix Theory," Rev. Mod. Phys. **32**, 335–369 (1960).

[Mermin, 1965] N. D. Mermin, "Thermal Properties of the Inhomogeneous Electron Gas," Phys. Rev. **137**, A1441–A1443 (1965).

[Misawa, 1965] S. Misawa, "Ferromagnetism of an Electron Gas," Phys. Rev. **140**, A1645–A1648 (1965).

[Parr and Yang, 1989] R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules* (Oxford University Press, Oxford, 1989).

[Pauli, 1933] W. Pauli, "Die allgemeinen Prinzipien der Wellenmechanik," in *Handb. d. Phys.*, 2. ed., edited by H. Geiger and K. Scheel (Springer-Verlag, Berlin, 1933), Vol. 24, part 1, pp. 83–272.

[Perdew, 1985] J. Perdew, "What do the Kohn-Sham Orbital Energies Mean? How do Atoms Dissociate?," in *Density Functional Methods*

in Physics, Proceedings of a NATO ASI Held in Alcabideche, Portugal, September 1983, edited by R. M. Dreizler and J. da Providencia (Plenum Press, New York, 1985), pp. 265–308.

[Perdew and Levy, 1983] J. P. Perdew and M. Levy, "Physical Content of the Exact Kohn-Sham Orbital Energies: Band Gaps and Derivative Discontinuities," Phys. Rev. Lett. **51**, 1884–1887 (1983).

[Perdew and Zunger, 1981] J. P. Perdew and A. Zunger, "Self-Interaction Correction to Density-Functonal Approximations for Many-Electron Systems," Phys. Rev. **B23**, 5048–5079 (1981).

[Rajagopal, 1978] A. K. Rajagopal, "Inhomogeneous Relativistic Electron Gas," J. Phys. C **11**, L943–L948 (1978).

[Rajagopal and Callaway, 1973] A. K. Rajagopal and J. Callaway, "Inhomogeneous Electron Gas," Phys. Rev. **B7**, 1912–1919 (1973).

[Reed and Simon, 1973] M. Reed and B. Simon, *Methods of Modern Mathematical Physics. Vol. I: Functional Analysis* (Academic Press, New York, 1973).

[Sakurai, 1985] J. J. Sakurai, *Modern Quantum Mechanics* (Benjamin, Menlo Park, 1985).

[Schwarz, 1994] A. S. Schwarz, *Topology for Physicists* (Springer-Verlag, Berlin, 1994).

[Seidl *et al.*, 1996] A. Seidl *et al.*, "Generalized Kohn-Sham schemes and the band-gap problem," Phys. Rev. **B53**, 3764–3774 (1996).

[Sewell, 1986] G. L. Sewell, *Quantum Theory of Collective Phenomena*, *Monographs on the Physics and Chemistry of Materials* (Oxford University Press, Oxford, 1986).

[Sham and Schlüter, 1983] L. J. Sham and M. Schlüter, "Density-Functional Theory of the Energy Gap," Phys. Rev. Lett. **51**, 1888–1891 (1983).

[Skudlarski and Vignale, 1993] P. Skudlarski and G. Vignale, "Exchange-Correlation Energy of a Three-Dimensional Electron Gas in a Magnetic Field," Phys. Rev. **B48**, 8547–8559 (1993).

[Slater, 1951] J. C. Slater, "A Simplification of the Hartree-Fock Method," Phys. Rev. **81**, 385–390 (1951).

[Slater, 1960] J. C. Slater, *Quantum Theory of Atomic Structure* (McGraw-Hill Book Company, New York, 1960), Vol. 1.

[Svane, 1995] A. Svane, "Comment on Self-Interaction-Corrected Density-Functional Formalism," Phys. Rev. **B51**, 7924–7926 (1995).

[Teller, 1962] E. Teller, "On the Stability of Molecules in the Thomas-Fermi Theory," Rev. Mod. Phys. **34**, 627–631 (1962).

[Thomas, 1927] L. H. Thomas, "The Calculation of Atomic Fields," Proc. Camp. Philos. Soc. **23**, 542–548 (1927).

[Trickey, 1990] *Density Functional Theory of Many-Fermion Systems*, Vol. 21 of *Advances in Quantum Chemistry*, edited by S. B. Trickey (Academic Press, Inc., San Diego, 1990).

[Vignale and Rasolt, 1987] G. Vignale and M. Rasolt, "Density-Functional Theory in Strong Magnetic Fields," Phys. Rev. Lett. **59**, 2360–2363 (1987).

[Vignale and Rasolt, 1988] G. Vignale and M. Rasolt, "Current- and Spin-Density-Functional Theory for Inhomogeneous Electron Systems in Strong Magnetic Fields," Phys. Rev. **B37**, 10685–10696 (1988).

[von Barth and Hedin, 1972] U. von Barth and L. Hedin, "A Local Exchange-Correlation Potential for the Spin Polarized Case: I," J. Phys. C **5**, 1629–1642 (1972).

[von Neumann, 1955] J. von Neumann, *Mathematical Foundation of Quantum Mechanics* (Princeton University Press, Princeton NJ, 1955).

[von Weizsäcker, 1935] C. F. von Weizsäcker, "Zur Theorie der Kernmassen," Z. Physik **96**, 431–458 (1935).

[Wilson, 1962] E. B. Wilson, Jr., "Four-Dimensional Electron Density Function," J. Chem. Phys. **36**, 2232–2233 (1962).

[Young, 1969] L. C. Young, *Lectures on the Calculus of Variations and Optimal Control Theory* (W. B. Saunders Company, Philadelphia, London, Toronto, 1969).

[Zeidler, 1986] E. Zeidler, *Nonlinear Functional Analysis and its Applications, vol. I-III* (Springer-Verlag, New York, 1986).

[Zumbach and Maschke, 1985] G. Zumbach and K. Maschke, "Density-Matrix Functional Theory for the N-Particle Ground State," J. Chem. Phys. **82**, 5604–5607 (1985).

Index

absolutely summable, 124
action, 183, 187
adiabatic approximation, 57
adiabatic forces, 56
adiabaticity assumption, 53
affine-linear, 77, 110
almost everywhere (a.e.), 63, 123
angular momentum density, 208
anticommutator, 32
aufbau principle, 95

Banach space, 120
    reflexive, 129
Banach-Alaoglu theorem, 130
base of topology, 117
bispinor, 193
Bohr radius, 17
Bolzano-Weierstrass theorem, 118
Borel set, 121
Bose condensed field, 191
bound minimum, 135
bounded linear functional, 129
bounded linear operator, 128

canonical momentum density, 185
Cauchy net, 120
Cauchy sequence, 120
chain rule for functional derivatives, 134
charge conservation, 180
chemical potential, 70, 98
classical electron radius, 188
closed set, 117
closed-shell Hartree-Fock method, 25
commutator, 31
compact, 118
complete set, 29
    of N-particle states, 24
    of spin-orbitals, 15
complete space, 120
Compton wavelength, 188
conjugate function, 112
conjugate functional, 131
continuous, 120
continuous linear operator, 128
convex, 111, 131
convex hull, 115, 132
correlation insulator, 171
coupling constant, 17, 52
creation and annihilation operators, 32
    bosonic, 30
    fermionic, 31
current density functional $H[J]$, 202
current density, paramagnetic, 211

d'Alembert operator, 181
de Rham's cohomology, 183
density functional
    universal, 145
density functional by Hohenberg and Kohn, 79
    of kinetic energy, 81
density functional by Levy and Lieb, 88
density functional for N-particle density matrices, 93
density matrix, 90
    reduced, 35–47
density operator, 33, 41
density parameter r_s, 70
Dirac equation, 191–193
distance, 119
dual pair, 131, 132

electromagnetic field, 180
electromagnetic potentials, 180
electron affinity, 99
electron liquid, homogeneous, 69, 156–161
electronegativity, 100
ensemble state, 90
essential supremum (ess sup), 63
exchange and correlation energy, 48, 161, 212
exchange and correlation field, 209
exchange and correlation hole, 48–51
exchange energy, 26
    of the homogeneous electron liquid, 159
exchange hole, 40, 49

exchange potential operator, 27
exchange term, 20

F-derivative, 133
Fenchel's duality, 137
Fermi energy, 70
Fermi radius, 23
fermion gas, interaction-free, 38
  homogeneous, 23, 37, 40, 49
field operators, 33
fine structure constant, 179, 188
finite from below, 111
Fock operator, 27, 29
Fock space, 30
four-current density, 182, 193, 205
four-field tensor, 182
four-momentum, 186
four-potential, 182
four-velocity, 186
fractional orbital occupation numbers, 90
free Dirac field, 193
free minimum, 135
free photon field, 190
functional derivative, 85, 132–135
fundamental form, 181
fundamental tensor, 181

G-derivative, 132
gap correction, 164
gauge invariance, 78, 201, 207
generalized sequence, 118
gradient expansion, 75
ground state, 78
  degenerate, 80
  determinantal, 82
  ensemble, 90
  non-degenerate, 80
  of the quantum field, 199
  spin-polarized, 102
ground state energy, 77, 139, 201
  ensemble, 90, 141

Hahn-Banach theorem, 113, 132
Hamiltonian, 13
  in Heisenberg representation, 24
  effective, of the inhomogeneous quantum field system, 200
Hamiltonian,
  field quantized, 33
  for Coulomb systems, 55
    total, 56
  in momentum representation, 19
  in natural units, 17
  in occupation number representation, 32
  in Schrödinger representation, 16
  of QED, 197
  of the free Dirac field, 194
  of the free photon field, 190
  of the homogeneous electron liquid, 156
  on the torus, 58
Hamiltonian density, 185
  of electromagnetic coupling, 195
hard potential barrier, 79
Hartree energy, 26, 48
Hartree potential, 27
Hartree unit, 17
Hartree-Fock energy, 26, 28
Hartree-Fock equations, 27
Hartree-Fock orbitals, 28
Hausdorff property, 117
Heisenberg picture, 189
Heitler-London ansatz, 51
Hellmann-Feynman theorem, 53
Hilbert space, 13, 127
  N-particle, bosonic, 16
  N-particle, fermionic, 15
Hölder's inequality, 126
Hohenberg-Kohn variational principle, 80, 144, 204
hole state, 34
HOMO, 28, 100
hyperplane, 130
hyperplane of support, 131

impure spin states, 107
inf-sup problem, 136
inner point, 116
ionization potential, 100

Janak's theorem, 95, 99
jump of the Kohn-Sham potential, 153

kinetic energy, 43, 48

kinetic energy,
exchange and correlation contribution, 53
kinetic energy functional, 62
Klein-Gordon equation, 191
Kohn-Sham exchange and correlation energy, 53
Kohn-Sham exchange and correlation hole, 54, 162
Kohn-Sham exchange and correlation potential, 85, 161, 206
Kohn-Sham orbital, 86
Kohn-Sham orbital energy, 100, 105
Kohn-Sham-Dirac equation, 210
general, 206, 210
Kohn-Sham-Dirac spinor orbitals, 205
Kohn-Sham-Dirac variational problem, 206
Koopmans' theorem, 28

Lagrange density, 184
Lagrange function, 137
Lagrange multiplier, 135–137
abstract, 137
LDA+U density functional, 175
Lebesgue integral, 123
Lebesgue measure, 121
Lebesgue space $\boldsymbol{L}^p$, 125–128
Legendre transform, 112, 131
limes, 116
of a net, 118
local current density approximation, 211, 212
local density approximation (LDA), 87, 161–165
locally convex space, 119
Lorentz gauge, 181, 190
Lorentz transformation, 183
lower semicontinuous, 111, 131
L^p-spaces, 63
LUMO, 28, 99
$\mathcal{L}(X, Y)$, 128

magnetization density, 209
main theorem for extremal problems, 130
Maxwell's equations, 180, 182
mean field, 28
measurable function, 122
measurable set, 122
metric space, 119
minimum problem, 130
Minkowski's geometry, 181
Minkowski's inequality, 126
molecular field, 28, 29
Mott insulator, 171

n-representable potential, 79
N-representable density, 88
natural orbitals, 93
natural units, 17
neighborhood, 116
net, 118
norm, 119
normal order, 191, 198

occupation matrix, 174
open set, 116
open-shell Hartree-Fock method, 25
orbital current density, 208
orbital relaxation energy, 28

p-summable function, 125
pair correlation function, 40
pair density, 38
paramagnetic state, 92
particle density, 23, 35
particle number operator, 31
Pauli-Hellmann-Feynman theorem, 53
Perdew-Zunger interpolation of $\varepsilon_{\mathrm{XC}}$, 160
periodic boundary conditions, 18
phase separation, 98
Poisson's equation, 66
potential gauge constant, 78
pure spin states, 107
pure state, 35
pure-state v-representable density, 79

quantum electrodynamics (QED), 179

regular measure, 122
relativistic dispersion relation, 187
relativistic effects
magnitude, 179

renormalization, 34, 191, 199, 204
rest mass, 186
Riesz-Fisher theorem, 126

saddle point, 137
scaled nuclear charges, 55
Schrödinger picture, 189
Schwarz' inequality, 127
self-consistent field, 29
self-interaction, 27, 46, 51, 205, 206
seminorm, 119
side condition, 135
σ-additive, 122
simplex, 91
Slater determinant, 15
Sobolev's inequality, 149
spin current density, 208
spin density, 44
spin magnetization density, 44
spin operator, 14
spin polarization, 100–105
    degree of, 44, 158
spin-density matrix, 36, 44
spin-orbital, 15
spinor, 14
subdifferential, 113, 131
subgradient, 112, 131

tangent of support, 113
Taylor expansion of a functional, 134
Teller's no-binding theorem, 72
thermodynamic limit, 33, 57–60, 98
Thomas-Fermi energy, 64
    of the atom, 68
Thomas-Fermi equation, 65, 166
Thomas-Fermi functional, 62
Thomas-Fermi screening length, 70
Thomas-Fermi-Dirac theory, 73
Thomas-Fermi-λWeizsäcker theory, 74
topological space, 117
topology, 117
torus, 18, 57, 140, 150
total charge, 199
trace, 36
transition state, 100

U-potential, 176

vacuum permeability μ_0, 180
vacuum permittivity ϵ_0, 180
virial theorem, 56
    of Thomas-Fermi theory, 72
von Barth-Hedin interpolation function
        $f(\zeta)$, 159
vorticity, 211

wavenumber space, 18
weak lower semicontinuous, 131
weak topology, 130
weak* topology, 130

X_α-approach, 87

Young's inequality, 112