Lecture notes

Numerical Methods in Quantum Mechanics

Corso di Laurea Magistrale in Fisica

Interateneo Trieste – Udine

Anno accademico 2020/2021

Paolo Giannozzi
University of Udine

Contains software and material written by

Furio Ercolessi[1] **and Stefano de Gironcoli**[2]
[1]*Formerly at University of Udine*
[2]*SISSA - Trieste*

Last modified May 27, 2021

Contents

Introduction

The aim of these lecture notes is to provide an introduction to methods and techniques used in the numerical solution of simple (non-relativistic) quantum-mechanical problems, with special emphasis on atomic and condensed-matter physics. The practical sessions are meant to be a sort of "computational laboratory", introducing the basic ingredients used in the calculation of materials properties at a much larger scale. The latter is a very important field of today's computational physics, due to its technological interest and potential applications.

The codes provided during the course are little more than templates. Students are expected to analyze them, to run them under various conditions, to examine their behavior as a function of input data, and most important, to interpret their output from a physical point of view. The students will be asked to extend or modify those codes, by adding or modifying some functionalities.

For further insight on the theory of Quantum Mechanics, many excellent textbooks are available (e.g. Griffiths, Schiff, or the ever-green Dirac and Landau). For further insight on the properly computational aspects of this course, we refer to the specialized texts quoted in the Bibliography section, and in particular to the book of Thijssen.

0.1 About Software

This course assumes some basic knowledge of how to write and execute simple programs, and how to plot their results. All that is needed is a Fortran or C compiler and some visualization software. The target machine is a PC running Linux, but you can also use a Macintosh or a Windows PC, as long as the mentioned software is installed and working, and if you know how to use it in practice. For Windows 10, the path of least resistance is to enable the Windows Subsystem for Linux (WSL-2) and to install a Linux distribution and an X window client. This gives access to a very functional Linux shell.

0.1.1 Compilers

In order to run a code written in any programming language, we must first translate it into machine language, i.e. a language that the computer can understand. The translation is done by an *interpreter* or by a *compiler*: the former translates and immediately executes each instruction, the latter takes the file, produces the so-called *object code* that together with other object codes

and with libraries is finally assembled into an *executable* file. Python, Java, or at an higher level, Matlab, Mathematica, are examples of "interpreted" language. Fortran, C, C++ are "compiled" languages.

Our codes are written in Fortran 90 (or 95, or later). This is a sophisticated and complex language offering dynamical memory management, arrays operations (e.g. matrix-vector products), modular and object-based structure. Fortran 90 maintains a wide compatibility with existing Fortran 77 codes, while remaining as efficient as Fortran 77 was. It is worth mentioning that the first applications of computers to physics date back to well before the birth of modern computer languages like C++, python, or even C: there is a large number of codes and libraries written in Fortran 77 (or even Fortran 66!) and still widely used in physics. Even among physicists, however, Fortran is no longer as common and widespread as it used to be, but online resources are still easy to find[1]. The codes themselves are very simple and make little usage of advanced language features. In any case, there is no problem if a student prefers to use a more widespread language like C/C++. A version of all codes in C is also available, with no warranty about the quality of the C code in terms of elegance and good coding practice.

In all cases, you need a C or Fortran compiler. The C compiler `gcc` is free and can be installed on all operating systems (in Linux PCs it is always present). Less-then-archaic versions of `gcc` include a Fortran compiler, called `gfortran`.

0.1.2 Visualization Tools

Visualization of data produced by the codes (wave functions, charge densities, various other quantities) has a central role in the analysis and understanding of the results. Code `gnuplot` can be used to make two-dimensional or three-dimensional plots of data or of analytical expressions. `gnuplot` is open-source software, available for all operating systems and often found pre-installed on Linux PCs. An introduction to `gnuplot`, with many links to more resources, can be found in `http://www.gnuplot.info/help.html`.

Another software that can be used is `Grace`[2], formerly known as `xmgr`. This is also open-source and highly portable, has a graphical user interface and thus it is easier to use than `gnuplot`, whose syntax is not always easy to remember.

0.1.3 Mathematical Libraries

The usage of efficient mathematical libraries is crucial in "serious" calculations. Some of the codes use routines from the BLAS[3] (Basic Linear Algebra Subprograms) library and from LAPACK[4] (Linear Algebra PACKage). The latter is an important and well-known library for all kinds of linear algebra operations:

[1] see for instance `http://fortran.bcs.org/resources.php`

[2] http://plasma-gate.weizmann.ac.il/Grace

[3] http://www.netlib.org/blas

[4] http://www.netlib.org/lapack

solution of linear systems, eigenvalue problems, etc.. LAPACK calls BLAS routines for all CPU-intensive calculations. Highly optimized versions of the latter are available for many different operating systems and architectures.

The original BLAS and LAPACK routines are written in Fortran 77. They are often found precompiled on many machines and can be linked directly by the compiler by adding `-llapack -lblas`. If called by a C code, it may be needed to add an underscore (_) in the calling program, as in: `dsyev_`, `dgemm_`. This is due to different, and machine-dependent, C-Fortran conventions for the naming of "symbols" (i.e. compiled routines). Note that the C compiler may also need `-lm` to link general mathematical libraries (i.e. operations like the square root).

0.1.4 Pitfalls in C-Fortran interlanguage calls

In addition to the above-mentioned potential mismatches between C and Fortran naming conventions, there are a few more pitfalls one has to be aware of when Fortran routines are called by C (or vice versa).

- Fortran passes *pointers* to subroutines and functions; C passes *values*. In order to call a Fortran routine from C, or vice versa, all C variables appearing in the call must be either pointers or arrays.
- In C, indices of vectors and arrays start from 0; in Fortran, they start from 1, unless differently specified in array declaration or allocation.
- Matrices in C are stored in memory row-wise, that is: `a[i][j+1]` follows `a[i][j]` in memory. In Fortran, they are stored column-wise, that is, the other way round: `a(i+1,j)` follows `a(i,j)` in memory.

An additional problem is that C does not provide run-time allocatable matrices like Fortran does, but only fixed-dimension matrices and arrays of pointers. The former are impractical, the latter are not usable as arguments to pass to Fortran. It would be possible, using either non-standard C syntax, or using C++ and the `new` command, to define dynamically allocated matrices similar to those used in Fortran. We have preferred for our simple C codes to "simulate" Fortran-style matrices (i.e. stored in memory column-wise) by mapping them onto one-dimensional C vectors.

We remark that Fortran 90 has a more advanced way of passing arrays to subroutines using "array descriptors". The codes used in this course however do not make use of this possibility but use the old-style Fortran 77 way of passing arrays via pointers.

0.2 Bibliography

J. M. Thijssen, *Computational Physics*, 2nd edition, Cambridge University Press, Cambridge, 2007:
`http://www.cambridge.org/gb/knowledge/isbn/item1171410`.

F. J. Vesely, *Computational Physics - An Introduction: Second Edition*, Kluwer, 2001. Also see the author's web page:
`http://www.ap.univie.ac.at/users/Franz.Vesely/cp0102/serious.html`, containing parts of the accompanying material.

S. E. Koonin e D. C. Meredith, *Computational physics - Fortran Version*, Addison-Wesley, 1990. See also Dawn Meredith's web page:
`https://mypages.unh.edu/dawnm/home`.

Chapter 1

One-dimensional Schrödinger equation

In this chapter we start from the harmonic oscillator to introduce a general numerical methodology to solve the one-dimensional, time-independent Schrödinger equation. The analytical solution of the harmonic oscillator will be first derived and described. A specific integration algorithm (Numerov) will be used. The extension of the numerical methodology to other, more general types of potentials does not present any special difficulty.

For a particle of mass m under a potential $V(x)$, the one-dimensional, time-independent Schrödinger equation is given by:

$$-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} + V(x)\psi(x) = E\psi(x), \tag{1.1}$$

where $\psi(x)$ is the wave function, that can be chosen to be real, and $\hbar$ is the Planck constant h divided by 2π. In the following we are focusing on the *discrete spectrum*: the set of isolated energy values for which Eq.(1.1) has normalizable solutions, localized in space.

1.1 The harmonic oscillator

The harmonic oscillator is a fundamental problem in classical dynamics as well as in quantum mechanics. It represents the simplest model system in which attractive forces are present and is an important paradigm for all kinds of vibrational phenomena. For instance, the vibrations around equilibrium positions of a system of interacting particles may be described, via an appropriate coordinate transformation, as a set of independent harmonic oscillators known as *normal vibrational modes*. The same holds in quantum mechanics. The study of the quantum oscillator allows a deeper understanding of quantization and of its effects and of wave functions of bound states.

In this chapter we first remind the main results of the theory of the harmonic oscillator, then we show how to set up a computer code that allows to numerically solve the Schrödinger equation for the harmonic oscillator. The resulting code can be easily modified and adapted to a different, not simply

quadratic, interaction potential. This allows to study problems that, unlike the harmonic oscillator, do not have a simple analytical solution.

1.1.1 Units

The Schrödinger equation for a one-dimensional harmonic oscillator is, in usual notations:

$$\frac{d^2\psi}{dx^2} = -\frac{2m}{\hbar^2}\left(E - \frac{1}{2}Kx^2\right)\psi(x) \tag{1.2}$$

where K is the force constant. The force on the mass is $F = -Kx$, proportional to the displacement x and directed towards the origin. Classically such an oscillator has a frequency (angular frequency)

$$\omega = \sqrt{\frac{K}{m}}. \tag{1.3}$$

It is convenient to work with adimensional units, the same that are used by the codes presented at the end of this chapter. Let us introduce an adimensional variable ξ and a length λ such that $x = \lambda\xi$. By substituting into Eq.(1.2), one finds

$$\frac{d^2\psi}{d\xi^2} = \left(-\frac{2mE\lambda^2}{\hbar^2} + \frac{mK\lambda^4}{\hbar^2}\xi^2\right)\psi. \tag{1.4}$$

The natural choice is to set $mK\lambda^4/\hbar^2 = 1$, leading to $\lambda = (\hbar^2/mK)^{1/4}$ and, using Eq.(1.3), to

$$\xi = \left(\frac{m\omega}{\hbar}\right)^{1/2} x \tag{1.5}$$

By further introducing an adimensional "energy" ε such that

$$\varepsilon = \frac{E}{\hbar\omega}, \tag{1.6}$$

one finally rewrites Eq.(1.2) in adimensional units:

$$\frac{d^2\psi}{d\xi^2} = -2\left(\varepsilon - \frac{\xi^2}{2}\right)\psi(\xi). \tag{1.7}$$

1.1.2 Exact solution

One can easily verify that for large ξ (such that ε can be neglected) the solutions of Eq.(1.7) must have an asymptotic behavior like

$$\psi(\xi) \sim \xi^n \mathrm{e}^{\pm\xi^2/2} \tag{1.8}$$

where n is any finite value. The + sign in the exponent must however be discarded: it would give raise to diverging, non-physical solutions (in which the particle would tend to leave the $\xi = 0$ point, instead of being attracted towards it by the elastic force). It is thus convenient to extract the asymptotic behavior and assume

$$\psi(\xi) = H(\xi)\mathrm{e}^{-\xi^2/2} \tag{1.9}$$

where $H(\xi)$ is a well-behaved function for large ξ (i.e. the asymptotic behavior is determined by the second factor $e^{-\xi^2/2}$). In particular, $H(\xi)$ must not grow like e^{ξ^2}, or else we fall back into a undesirable non-physical solution.

Under the assumption of Eq.(1.9), Eq.(1.7) becomes an equation for $H(\xi)$:

$$\frac{d^2H}{d\xi^2}(\xi) - 2\xi\frac{dH}{d\xi}(\xi) + (2\varepsilon - 1)H(\xi) = 0. \tag{1.10}$$

It is immediate to notice that $\varepsilon_0 = 1/2$, $H_0(\xi) = 1$ is the simplest solution. This is the *ground state*, i.e. the lowest-energy solution, as will soon be clear.

In order to find all solutions, we expand $H(\xi)$ into a series, in principle with infinite terms:

$$H(\xi) = \sum_{n=0}^{\infty} A_n \xi^n, \tag{1.11}$$

we derive the series to find $dH/d\xi$ and $d^2H/d\xi^2$, plug the results into Eq.(1.10) and regroup terms with the same power of ξ. We find an equation

$$\sum_{n=0}^{\infty} \left[(n+2)(n+1)A_{n+2} + (2\varepsilon - 2n - 1)A_n\right] \xi^n = 0 \tag{1.12}$$

that can be satisfied for any value of ξ only if the coefficients of all the orders are zero:

$$(n+2)(n+1)A_{n+2} + (2\varepsilon - 2n - 1)A_n = 0. \tag{1.13}$$

Thus, once A_0 and A_1 are given, Eq.(1.13) allows to determine by recursion the solution under the form of a power series.

For large n, the coefficient of the series behave like $A_{n+2} \sim 2A_n/n$, that is:

$$A_{n+2} \sim \frac{1}{(n/2)!}. \tag{1.14}$$

This asymptotic behavior is the same as for the series expansion $\exp(\xi^2) = \sum_n \xi^{2n}/n!$, indicating that the recursion relation Eq.(1.13) produces a function $H(\xi)$ that grows like $\exp(\xi^2)$ and yields a diverging nonphysical solution.

The only way to prevent this from happening is to have in Eq.(1.13) all coefficients beyond a given n vanish, so that the infinite series reduces to a finite-degree polynomial. This happens if and only if

$$\varepsilon = n + \frac{1}{2} \tag{1.15}$$

where n is a non-negative integer.

Allowed energies for the harmonic oscillator are thus *quantized*:

$$E_n = \left(n + \frac{1}{2}\right)\hbar\omega \quad n = 0, 1, 2, \ldots \tag{1.16}$$

The corresponding polynomials $H_n(\xi)$ are known as *Hermite polynomials*. $H_n(\xi)$ is of degree n in ξ, has n nodes, is even $[H_n(-\xi) = H_n(\xi)]$ for even n, odd

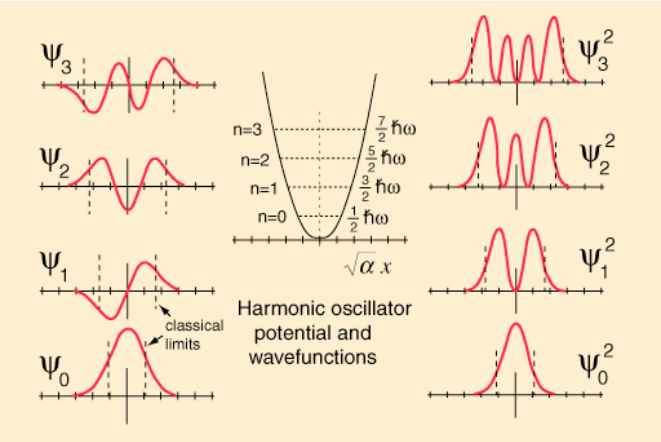


Figure 1.1: Wave functions and probability density for the harmonic oscillator.

$[H_n(-\xi) = -H_n(\xi)]$ for odd n. Since $\mathrm{e}^{-\xi^2/2}$ is node-less and even, the complete wave function corresponding to the energy E_n:

$$\psi_n(\xi) = H_n(\xi)\mathrm{e}^{-\xi^2/2} \tag{1.17}$$

has n nodes and the same parity as n. The fact that all solutions of the Schrödinger equation are either odd or even functions is a consequence of the symmetry of the potential: $V(-x) = V(x)$.

The lowest-order Hermite polynomials are

$$H_0(\xi) = 1, \quad H_1(\xi) = 2\xi, \quad H_2(\xi) = 4\xi^2 - 2, \quad H_3(\xi) = 8\xi^3 - 12\xi. \tag{1.18}$$

1.1.3 Comparison with classical probability density

The probability density for wave function $\psi_n(x)$ of the harmonic oscillator has in general $n+1$ peaks, whose height increases while approaching the corresponding classical inversion points (i.e. points where $V(x) = E$).

These probability density can be compared to that of the classical harmonic oscillator, in which the mass moves according to $x(t) = x_0 \sin(\omega t)$. The probability $\rho(x)dx$ to find the mass between x and $x + dx$ is proportional to the time needed to cross such a region, i.e. it is inversely proportional to the speed as a function of x:

$$\rho(x)dx \propto \frac{dx}{v(x)}. \tag{1.19}$$

Since $v(t) = x_0\omega\cos(\omega t) = \omega\sqrt{x_0^2 - x_0^2\sin^2(\omega t)}$, we have

$$\rho(x) \propto \frac{1}{\sqrt{x_0^2 - x^2}}. \tag{1.20}$$

This probability density has a minimum for $x = 0$, diverges at inversion points, is zero beyond inversion points.

The quantum probability density for the ground state is completely different: has a maximum for $x = 0$, decreases for increasing x. At the classical inversion point its value is still $\sim 60\%$ of the maximum value: the particle has a high probability to be in the classically forbidden region (for which $V(x) > E$).

In the limit of large quantum numbers, i.e., large n values, the quantum density tends however to look similar to the quantum one, but it still displays the oscillatory behavior in the allowed region, typical for quantum systems.

1.2 Quantum mechanics and numerical codes

1.2.1 Quantization

A first aspect to be considered in the numerical solution of quantum problems is the presence of *quantization* of energy levels for bound states, such as for instance Eq.(1.16) for the harmonic oscillator. The acceptable energy values E_n are not known *a priori*. Thus in the Schrödinger equation (1.1) the unknown is not just $\psi(x)$ but also E. For each allowed energy level, or *eigenvalue*, E_n, there will be a corresponding wave function, or *eigenfunction*, $\psi_n(x)$.

What happens if we try to solve the Schrödinger equation for an energy E that does not correspond to an eigenvalue? In fact, a "solution" exists for *any* value of E, but not a physical one. The quantization of the energy originates from boundary conditions, requiring no nonphysical divergence of the wave function in the forbidden regions. Thus, if E is not an eigenvalue, we will observe a divergence of $\psi(x)$. Numerical codes searching for allowed energies must be able to recognize when the energy is not correct and search for a better energy, until it coincides – within numerical or predetermined accuracy – with an eigenvalue. The first code presented at the end of this chapter implements such a strategy.

1.2.2 A pitfall: pathological asymptotic behavior

An important aspect of quantum mechanics is the existence of "negative" kinetic energies: i.e., the wave function can be non zero (and thus the probability to find a particle can be finite) in regions for which $V(x) > E$, forbidden according to classical mechanics. Based on (1.1) and assuming the simple case in which V is (or can be considered) constant, this means

$$\frac{d^2\psi}{dx^2} = k^2\psi(x) \tag{1.21}$$

where k^2 is a positive quantity. This in turns implies an exponential behavior, with both $\psi(x) \simeq \exp(kx)$ and $\psi(x) \simeq \exp(-kx)$ satisfying (1.21). As a rule only one of these two possibilities has a physical meaning: the one that gives raise to a wave function that *decreases* exponentially at large $|x|$.

It is very easy to distinguish between the "good" and the "bad" solution for a human, much less so for a numerical code that produces whatever comes out from the equations. If even a tiny amount of the "bad" solution, for instance due to numerical noise, is present, the integration algorithm will inexorably

make it grow in the classically forbidden region. As the integration goes on, the "bad" solution will dominate the "good" one and eventually produce crazy numbers (or crazy NaN's: Not a Number). Thus a nice-looking wave function in the classically allowed region, smoothly decaying in the classically forbidden region, may suddenly start to diverge at some point, unless a wise strategy is employed to prevent it. The second code presented at the end of this chapter implements such a strategy.

1.3 Numerov's method

In order to solve numerically the time-independent Schrödinger equation in one dimension, one has to *discretize* it on a suitable finite grid of points and to *integrate* (solve) it, the solution being also given on the grid of points.

There are many big thick books on this subject, describing old and new methods, from the very simple to the very sophisticated, for all kinds of differential equations and all kinds of discretization and integration algorithms. In the following, we will consider *Numerov's method*, named after Russian astronomer Boris Vasilyevich Numerov, as an example of a simple yet powerful and accurate algorithm.

Numerov's method is useful to integrate second-order differential equations of the general form

$$\frac{d^2y}{dx^2} = -g(x)y(x) + s(x) \tag{1.22}$$

where $g(x)$ and $s(x)$ are known functions. The Schrödinger equation, Eq.(1.1), has this form, with $g(x) = (2m/\hbar^2)[E - V(x)]$ and $s(x) = 0$. We will see in the next chapter that also the radial Schrödinger equations for three-dimensional systems having spherical symmetry belongs to such class. Another important equation falling into this category is Poisson's equation of electromagnetism,

$$\frac{d^2\phi}{dx^2} = -4\pi\rho(x) \tag{1.23}$$

where $\rho(x)$ is the charge density. In this case $g(x) = 0$ and $s(x) = -4\pi\rho(x)$.

Let us consider a finite box containing the system: for instance, $-x_{\max} \leq x \leq x_{\max}$, with $x_{\max}$ large enough for our solutions to decay to negligibly small values. Let us divide our finite box into N small intervals of equal size, Δx wide. We call x_i the points of the grid so obtained, $y_i = y(x_i)$ the values of the unknown function $y(x)$ on grid points. In the same way we indicate by g_i and s_i the values of the (known) functions $g(x)$ and $s(x)$ in the same grid points. In order to obtain a discretized version of the differential equation (i.e. to obtain an equation involving finite differences), we expand $y(x)$ into a Taylor series around a point x_n, up to fifth order:

$$\begin{aligned} y_{n-1} = \;& y_n - y'_n\Delta x + \tfrac{1}{2}y''_n(\Delta x)^2 - \tfrac{1}{6}y'''_n(\Delta x)^3 + \tfrac{1}{24}y''''_n(\Delta x)^4 - \tfrac{1}{120}y'''''_n(\Delta x)^5 \\ & + O[(\Delta x)^6] \\ y_{n+1} = \;& y_n + y'_n\Delta x + \tfrac{1}{2}y''_n(\Delta x)^2 + \tfrac{1}{6}y'''_n(\Delta x)^3 + \tfrac{1}{24}y''''_n(\Delta x)^4 + \tfrac{1}{120}y'''''_n(\Delta x)^5 \\ & + O[(\Delta x)^6]. \end{aligned} \tag{1.24}$$

We have switched to notation $y'(x) = dy/dx$ etc. for compactness. If we sum the two equations, we obtain:

$$y_{n+1} + y_{n-1} = 2y_n + y''_n(\Delta x)^2 + \frac{1}{12}y''''_n(\Delta x)^4 + O[(\Delta x)^6]. \tag{1.25}$$

Eq.(1.22) tells us that

$$y''_n = -g_n y_n + s_n \equiv z_n. \tag{1.26}$$

The quantity z_n above is introduced to simplify the notations. The following relation holds:

$$z_{n+1} + z_{n-1} = 2z_n + z''_n(\Delta x)^2 + O[(\Delta x)^4]. \tag{1.27}$$

This is the simple formula for the discretized second derivative, that can be obtained in a straightforward way by Taylor expansion up to third order. Thus:

$$y''''_n \equiv z''_n = \frac{z_{n+1} + z_{n-1} - 2z_n}{(\Delta x)^2} + O[(\Delta x)^2]. \tag{1.28}$$

By inserting back these results into Eq.(1.25) one finds

$$\begin{array}{ll} y_{n+1} & = 2y_n - y_{n-1} + (-g_n y_n + s_n)(\Delta x)^2 \\ & + \frac{1}{12}(-g_{n+1}y_{n+1} + s_{n+1} - g_{n-1}y_{n-1} + s_{n-1} + 2g_n y_n - 2s_n)(\Delta x)^2 \\ & + O[(\Delta x)^6] \end{array} \tag{1.29}$$

and finally the *Numerov's formula*

$$\begin{array}{ll} y_{n+1}\left[1 + g_{n+1}\frac{(\Delta x)^2}{12}\right] & = 2y_n\left[1 - 5g_n\frac{(\Delta x)^2}{12}\right] - y_{n-1}\left[1 + g_{n-1}\frac{(\Delta x)^2}{12}\right] \\ & + (s_{n+1} + 10s_n + s_{n-1})\frac{(\Delta x)^2}{12} + O[(\Delta x)^6] \end{array} \tag{1.30}$$

that allows to obtain y_{n+1} starting from y_n and y_{n-1}, and recursively the function in the entire box.

The value of the function in the first two points are needed in order to start the recurrence. We remark that such initial conditions differ from the more traditional ones for second-order differential equations:

$$y(x_0) = y_0, \quad y'(x_0) = y'_0, \tag{1.31}$$

in which the value at one point and the derivative in the same point are specified.

It is of course possible to integrate both in the direction of positive x and in the direction of negative x. In the presence of inversion symmetry, it will be sufficient to integrate in just one direction.

In our case—Schrödinger equation—the s_n terms are absent. It is convenient to introduce an auxiliary array f_n, defined as

$$f_n \equiv 1 + g_n\frac{(\Delta x)^2}{12}, \qquad \text{where} \qquad g_n = \frac{2m}{\hbar^2}[E - V(x_n)], \tag{1.32}$$

and to rewrite Numerov's formula as

$$y_{n+1} = \frac{(12 - 10f_n)y_n - f_{n-1}y_{n-1}}{f_{n+1}}. \tag{1.33}$$

The value of the energy is now hidden into g_n and f_n.

1.3.1 Code: harmonic0

Code `harmonic0.f90`[1] (or `harmonic0.c`[2]) solves the Schrödinger equation for the quantum harmonic oscillator, Eq.(1.7) in adimensional units. using the Numerov's algorithm above described for integration, and searching eigenvalues with a pre-determined number n of nodes using the "shooting method".

The shooting method is quite similar to the bisection procedure for the search of the zero of a function. We define an initial energy range $[E_{\min}, E_{\max}]$ that must contain the eigenvalue E_n. We start with an energy E equal to the mid-point of the energy range, $E = (E_{\max} + E_{\min})/2$. The wave function is integrated starting from $x = 0$ in the direction of positive x; at the same time, the number of nodes (i.e. of changes of sign of the function) is counted. If the number of nodes is larger than n, E is too high; if the number of nodes is smaller than n, E is too low. We then choose the lower half-interval $[E_{\min}, E_{\max} = E]$, or the upper half-interval $[E_{\min} = E, E_{\max}]$, respectively, select a new trial eigenvalue E in the mid-point of the new interval, iterate the procedure. When the energy interval is smaller than a pre-determined threshold, we assume that convergence has been reached.

For negative x the function is constructed using symmetry, since $\psi_n(-x) = (-1)^n \psi_n(x)$. This is of course possible only because $V(-x) = V(x)$, otherwise integration would have to be performed on the entire interval. The parity of the wave function determines the choice of the starting points for the recursion. For n odd, the two first points can be chosen as $y_0 = 0$ and an arbitrary finite value for y_1. For n even, y_0 is arbitrary and finite, y_1 is determined by Numerov's formula, Eq.(1.33), with $f_1 = f_{-1}$ and $y_1 = y_{-1}$:

$$y_1 = \frac{(12 - 10 f_0) y_0}{2 f_1}. \tag{1.34}$$

The code prompts for some input data:

- the limit $x_{\max}$ for integration (typical values: $5 \div 10$);
- the number N of grid points (typical values range from hundreds to a few thousand); note that the grid point index actually runs from 0 to N, so that $\Delta x = x_{\max}/N$;
- the name of the file where output data is written;
- the required number n of nodes (the code will stop if n is negative).

Finally the code prompts for a trial energy. You should answer 0 in order to search for an eigenvalue with n nodes. The code will start iterating on the energy, printing on standard output (i.e. at the terminal): iteration number, number of nodes found (on the positive x axis only), the current energy eigenvalue estimate. It is however possible to specify an energy, not necessarily an eigenvalue, to force the code to perform an integration at fixed energy and see

[1] http://www.fisica.uniud.it/%7Egiannozz/Didattica/MQ/Software/F90/harmonic0.f90
[2] http://www.fisica.uniud.it/%7Egiannozz/Didattica/MQ/Software/C/harmonic0.c

the resulting wave function. It is useful for testing purposes and to better understand how the eigenvalue search works (or doesn't work). Note that in this case the required number of nodes will not be honored; however the integration will be different for odd or even number of nodes, because the parity of n determines how the first two grid points are chosen.

The output file contains five columns: respectively, x, $\psi(x)$, $|\psi(x)|^2$, $\rho_{\rm cl}(x)$ and $V(x)$. $\rho_{\rm cl}(x)$ is the classical probability density (normalized to 1) of the harmonic oscillator, given in Eq.(1.20). All these quantities can be plotted as a function of x using any plotting program, such as `gnuplot`, shortly described in the introduction. Note that the code will prompt for a new value of the number of nodes after each calculation of the wave function: answer -1 to stop the code. If you perform more than one calculation, the output file will contain the result for all of them in sequence. Also note that the wave function are written for the entire box, from $-x_{\rm max}$ to $x_{\rm max}$.

It will become quickly evident that the code "sort of" works: the results look good in the region where the wave function is not vanishingly small, but invariably, the pathological behavior described in Sec.(1.2.2) sets up and wave functions diverge at large $|x|$. As a consequence, it is impossible to normalize the $\psi(x)$. The code definitely needs to be improved. The proper way to deal with such difficulty is to find an inherently stable algorithm.

1.3.2 Code: harmonic1

Code `harmonic1.f90`[3] (or `harmonic1.c`[4]) is the improved version of `harmonic0` that does not suffer from the problem of divergence at large x.

Two integrations are performed: a forward recursion, starting from $x = 0$, and a backward one, starting from $x_{\rm max}$. The eigenvalue is fixed by the condition that the two parts of the function match with continuous first derivative (as required for a physical wave function, if the potential is finite). The matching point is chosen in correspondence of the classical inversion point, $x_{\rm cl}$, i.e. where $V(x_{\rm cl}) = E$. Such point depends upon the trial energy E. For a function defined on a finite grid, the matching point is defined with an accuracy that is limited by the interval between grid points. In practice, one finds the index `icl` of the first grid point $x_c =$ `icl`Δx such that $V(x_c) > E$; the classical inversion point will be located somewhere between $x_c - \Delta x$ and x_c.

The outward integration is performed until grid point `icl`, yielding a function $\psi_L(x)$ defined in $[0, x_c]$; the number n of changes of sign is counted in the same way as in `harmonic0`. If n is not correct the energy is adjusted (lowered if n too high, raised if n too low) as in `harmonic0`. We note that it is not needed to look for changes of sign beyond x_c: in fact we know *a priori* that in the classically forbidden region there cannot be any nodes (no oscillations, just decaying solutions).

If the number of nodes is the expected one, the code starts to integrate inward from the rightmost points. Note the statement `y(mesh) = dx`: its only goal is to force solutions to be positive, since the solution at the left of the

[3] http://www.fisica.uniud.it/%7Egiannozz/Didattica/MQ/Software/F90/harmonic1.f90
[4] http://www.fisica.uniud.it/%7Egiannozz/Didattica/MQ/Software/C/harmonic1.c

matching point is also positive. The value `dx` is arbitrary: the solution is anyway rescaled in order to be continuous at the matching point. The code stops the same index `icl` corresponding to x_c. We thus get a function $\psi_R(x)$ defined in $[x_c, x_{\max}]$.

In general, the two parts of the wave function have different values in x_c: $\psi_L(x_c)$ and $\psi_R(x_c)$. We first of all re-scale $\psi_R(x)$ by a factor $\psi_L(x_c)/\psi_R(x_c)$, so that the two functions match continuously in x_c. Then, the whole function $\psi(x)$ is renormalized in such a way that $\int |\psi(x)|^2 dx = 1$.

Now comes the crucial point: the two parts of the function will have in general a discontinuity at the matching point $\psi'_R(x_c) - \psi'_L(x_c)$. This difference should be zero for a good solution, but it will not in practice, unless we are really close to the good energy $E = E_n$. The sign of the difference allows us to understand whether E is too high or too low, and thus to apply again the bisection method to improve the estimate of the energy.

In order to calculate the discontinuity with good accuracy, we write the Taylor expansions:

$$\begin{array}{l} y^L_{i-1} = y^L_i - y'^L_i \Delta x + \frac{1}{2} y''^L_i (\Delta x)^2 + O[(\Delta x)^3] \\ y^R_{i+1} = y^R_i + y'^R_i \Delta x + \frac{1}{2} y''^R_i (\Delta x)^2 + O[(\Delta x)^3] \end{array} \tag{1.35}$$

For clarity, in the above equation i indicates the index `icl`. We sum the two Taylor expansions and obtain, noting that $y^L_i = y^R_i = y_i$, and that $y''^L_i = y''^R_i = y''_i = -g_i y_i$ as guaranteed by Numerov's method:

$$y^L_{i-1} + y^R_{i+1} = 2y_i + (y'^R_i - y'^L_i)\Delta x - g_i y_i (\Delta x)^2 + O[(\Delta x)^3] \tag{1.36}$$

that is

$$y'^R_i - y'^L_i = \frac{y^L_{i-1} + y^R_{i+1} - [2 - g_i(\Delta x)^2] y_i}{\Delta x} + O[(\Delta x)^2] \tag{1.37}$$

or else, by using the notations as in Eq.(1.32),

$$y'^R_i - y'^L_i = \frac{y^L_{i-1} + y^R_{i+1} - (14 - 12 f_i) y_i}{\Delta x} + O[(\Delta x)^2] \tag{1.38}$$

In this way the code calculated the discontinuity of the first derivative. If the sign of $y'^R_i - y'^L_i$ is positive, the energy is too high (can you give an argument for this?) and thus we move to the lower half-interval; if negative, the energy is too low and we move to the upper half interval. As usual, convergence is declared when the size of the energy range has been reduced, by successive bisection, to less than a pre-determined tolerance threshold.

During the procedure, the code prints on standard output a line for each iteration, containing: the iteration number; the number of nodes found (on the positive x axis only); if the number of nodes is the correct one, the discontinuity of the derivative $y'^R_i - y'^L_i$ (zero if number of nodes not yet correct); the current estimate for the energy eigenvalue. At the end, the code writes the final wave function (this time, correctly normalized to 1!) to the output file.

1.3.3 Laboratory

Here are a few hints for "numerical experiments" to be performed in the computer lab (or afterward), using both codes:

- Calculate and plot eigenfunctions for various values of n. It may be useful to plot, together with eigenfunctions or eigenfunctions squared, the classical probability density, contained in the fourth column of the output file. It will clearly show the classical inversion points. With `gnuplot`, e.g.:

  `plot "`*filename*`" u 1:3 w l, "`*filename*`" u 1:4 w l`

  (`u` = using, `1:3` = plot column3 vs column 1, `w l` = with lines; the second "*filename*" can be replaced by "").

- Look at the wave functions obtained by specifying an energy value not corresponding to an eigenvalue. Notice the difference between the results of `harmonic0` and `harmonic1` in this case.

- Look at what happens when the energy is close to but not exactly an eigenvalue. Again, compare the behavior of the two codes.

- Examine the effects of the parameters `xmax`, `mesh`. For a given Δx, how large can be the number of nodes?

- Verify how close you go to the exact results (remember that there is a convergence threshold on the energy hardwired in the code). What are the factors that affect the accuracy of the results?

Possible code modifications and extensions:

- Modify the potential, keeping inversion symmetry. This will require very little changes to be done. You might for instance consider a "double-well" potential described by the form:

$$V(x) = \epsilon \left[\left(\frac{x}{\delta} \right)^4 - 2 \left(\frac{x}{\delta} \right)^2 + 1 \right], \qquad \epsilon, \delta > 0. \tag{1.39}$$

- Modify the potential, breaking inversion symmetry. You might consider for instance the *Morse potential*:

$$V(x) = D \left[e^{-2ax} - 2e^{-ax} + 1 \right], \tag{1.40}$$

  widely used to model the potential energy of a diatomic molecule. Which changes are needed in order to adapt the algorithm to cover this case?

Chapter 2

Schrödinger equation for central potentials

In this chapter we extend the concepts and methods introduced in the previous chapter for a one-dimensional problem to a specific and very important class of three-dimensional problems: a particle of mass m under a central potential $V(r)$, i.e. depending only upon the distance r from a fixed center. The Schrödinger equation we are going to study in this chapter is thus

$$H\psi(\mathbf{r}) \equiv \left[-\frac{\hbar^2}{2m}\nabla^2 + V(r)\right]\psi(\mathbf{r}) = E\psi(\mathbf{r}). \tag{2.1}$$

The problem of two interacting particles via a potential depending only upon their distance, $V(|\mathbf{r}_1 - \mathbf{r}_2|)$, e.g. the Hydrogen atom, reduces to this case, with m equal to the reduced mass of the two particles.

The general solution proceeds via the separation of the Schrödinger equation into an angular and a radial part. In this chapter we consider the numerical solution of the radial Schrödinger equation. A non-uniform grid is introduced and the radial Schrödinger equation is transformed to an equation that can still be solved using Numerov's method introduced in the previous chapter.

2.1 Variable separation

Let us introduce a polar coordinate system (r, θ, ϕ), where θ is the polar angle, ϕ the azimuthal one, and the polar axis coincides with the z Cartesian axis. After some algebra, one finds the Laplacian operator in polar coordinates:

$$\nabla^2 = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial\phi^2} \tag{2.2}$$

It is convenient to introduce the operator $L^2 = L_x^2 + L_y^2 + L_z^2$, the square of the angular momentum vector operator, $\mathbf{L} = -i\hbar\mathbf{r}\times\nabla$. Both $\vec{L}$ and L^2 act only on angular variables. In polar coordinates, the explicit representation of L^2 is

$$L^2 = -\hbar^2\left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\right]. \tag{2.3}$$

The Hamiltonian can thus be written as

$$H = -\frac{\hbar^2}{2m}\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{L^2}{2mr^2} + V(r). \tag{2.4}$$

The term $L^2/2mr^2$ has a classical analogous: the radial motion of a mass having classical angular momentum L_{cl} can be described by an effective radial potential $\hat{V}(r) = V(r) + L_{cl}^2/2mr^2$, where the second term (the "centrifugal potential") takes into account the effects of rotational motion. For high L_{cl} the centrifugal potential "pushes" the equilibrium position outwards.

In the quantum case, both L^2 and one component of the angular momentum, for instance L_z:

$$L_z = -i\hbar\frac{\partial}{\partial\phi} \tag{2.5}$$

commute with the Hamiltonian, so L^2 and L_z are conserved and H, L^2, L_z have a (complete) set of common eigenfunctions. We can thus use the eigenvalues of L^2 and L_z to classify the states. Let us now proceed to the separation of radial and angular variables, as suggested by Eq.(2.4). Let us assume

$$\psi(r,\theta,\phi) = R(r)Y(\theta,\phi). \tag{2.6}$$

After some algebra we find that the Schrödinger equation can be split into an angular and a radial equation. The solution of the angular equations are the *spherical harmonics*, known functions that are eigenstates of both L^2 and L_z:

$$L_z Y_{\ell m}(\theta,\phi) = m\hbar Y_{\ell m}(\theta,\phi), \qquad L^2 Y_{\ell m}(\theta,\phi) = \ell(\ell+1)\hbar^2 Y_{\ell m}(\theta,\phi) \tag{2.7}$$

($\ell \geq 0$ and $m = -\ell, ..., \ell$ are integer numbers).

The radial equation is

$$-\frac{\hbar^2}{2m}\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial R_{n\ell}}{\partial r}\right) + \left[V(r) + \frac{\hbar^2\ell(\ell+1)}{2mr^2}\right]R_{n\ell}(r) = E_{n\ell}R_{n\ell}(r). \tag{2.8}$$

In general, the energy will depend upon ℓ because the effective potential does; moreover, for a given ℓ, we expect to find bound states with discrete energies and we have indicated with n the corresponding index.

Finally, the complete wave function will be

$$\psi_{n\ell m}(r,\theta,\phi) = R_{n\ell}(r)Y_{\ell m}(\theta,\phi) \tag{2.9}$$

The energy does not depend upon m. As already observed, m classifies the projection of the angular momentum on an arbitrarily chosen axis. Due to spherical symmetry of the problem, the energy cannot depend upon the orientation of the vector $\mathbf{L}$, but only upon his modulus. An energy level $E_{n\ell}$ will then have a *degeneracy* $2\ell + 1$ (or larger, if there are other observables that commute with the Hamiltonian and that we haven't considered).

2.1.1 Radial equation

The probability $p(r)dr$ to find the particle at a distance between r and $r+dr$ from the center is given by the integration over angular variables only of the wavefunction squared:

$$p(r)dr = \int_\Omega |\psi_{n\ell m}(r,\theta,\phi)|^2 r d\theta\, r\sin\theta\, d\phi dr = |R_{n\ell}|^2 r^2 dr = |\chi_{n\ell}|^2 dr \tag{2.10}$$

where we have introduced an auxiliary function $\chi(r)$, sometimes called *orbital* wavefunction,

$$\chi(r) = rR(r) \tag{2.11}$$

and exploited the normalization of the spherical harmonics:

$$\int_0^{2\pi} d\phi \int_0^{\pi} d\theta |Y_{\ell m}(\theta,\phi)|^2 \sin\theta = 1 \tag{2.12}$$

As a consequence the normalization condition for χ is

$$\int_0^{\infty} |\chi_{n\ell}(r)|^2 dr = 1 \tag{2.13}$$

The function $|\chi(r)|^2$ can thus be directly interpreted as the radial probability density. Let us re-write the radial equation for $\chi(r)$ instead of $R(r)$. Its is straightforward to find that Eq.(2.8) becomes

$$-\frac{\hbar^2}{2m}\frac{d^2\chi}{dr^2} + \left[V(r) + \frac{\hbar^2\ell(\ell+1)}{2mr^2} - E\right]\chi(r) = 0. \tag{2.14}$$

We note that this equation has the same form as the one-dimensional Schrödinger equation, Eq.(1.1), for a particle under an effective potential and $r \geq 0$:

$$\hat{V}(r) = V(r) + \frac{\hbar^2\ell(\ell+1)}{2mr^2}. \tag{2.15}$$

As already explained, the second term is the centrifugal potential. The same methods used to find the solution of Eq.(1.1), and in particular, Numerov's method, can be used to find the radial part of the eigenfunctions of the energy.

2.2 Coulomb potential

The most important and famous case is when $V(r)$ is the Coulomb potential:

$$V(r) = -\frac{Ze^2}{4\pi\epsilon_0 r}, \tag{2.16}$$

where $e = 1.6021 \times 10^{-19}$ C is the electron charge, Z is the atomic number (number of protons in the nucleus), $\epsilon_0 = 8.854187817 \times 10^{-12}$ in MKSA units. Physicists tend to prefer the CGS system, in which the Coulomb potential is written as:

$$V(r) = -Zq_e^2/r. \tag{2.17}$$

In the following we will use $q_e^2 = e^2/(4\pi\epsilon_0)$ so as to fall back into the simpler CGS form.

It is often practical to work with *atomic units* (a.u.): assuming that $\hbar = 1$, $m_e = 1/2$, $q_e^2 = 2$, the length is expressed in units of *Bohr radii* (or simply, "Bohr"), a_0:

$$a_0 = \frac{\hbar^2}{m_e q_e^2} = 0.529177\,\text{Å} = 0.529177 \times 10^{-10}\,\text{m}, \tag{2.18}$$

while energies are expressed in units of *Rydberg* (Ry):

$$1\,\text{Ry} = \frac{m_e q_e^4}{2\hbar^2} = 13.6058\,\text{eV}. \tag{2.19}$$

when $m_e = 9.11 \times 10^{-31}$ Kg is the electron mass, not the reduced mass of the electron and the nucleus.

By assuming instead $\hbar = 1, m_e = 1, q_e = 1$, we obtain another set of atomic units, in which the *Hartree* (Ha) is the unit of energy instead of the Ry:

$$1\,\text{Ha} = 2\,\text{Ry} = \frac{m_e q_e^4}{\hbar^2} = 27.212\,\text{eV}. \tag{2.20}$$

Beware! Never talk about "atomic units" without first specifying *which ones*. In the following, the first set ("Rydberg" units) will be occasionally used.

We note first of all that for small r the centrifugal potential is the dominant term in the potential. The behavior of the solutions for $r \to 0$ will then be determined by

$$\frac{d^2\chi}{dr^2} \simeq \frac{\ell(\ell+1)}{r^2}\chi(r) \tag{2.21}$$

yielding $\chi(r) \sim r^{\ell+1}$, or $\chi(r) \sim r^{-\ell}$. The second possibility is not physical because $\chi(r)$ is not allowed to diverge.

For large r instead we remark that bound states may be present only if $E < 0$: there will be a classical inversion point beyond which the kinetic energy becomes negative, the wave function decays exponentially, only some energies can yield valid solutions. The case $E > 0$ corresponds instead to a problem of electron-nucleus scattering, with propagating solutions and a continuum energy spectrum. Here we deal with bound states; unbound states are the subject of the next chapter.

The asymptotic behavior of the solutions for large $r \to \infty$ will thus be determined by

$$\frac{d^2\chi}{dr^2} \simeq -\frac{2m_e}{\hbar^2}E\chi(r) \tag{2.22}$$

yielding $\chi(r) \sim \exp(\pm kr)$, where $k = \sqrt{-2m_e E}/\hbar$. The + sign must be discarded as nonphysical. It is thus sensible to assume for the solution a form like

$$\chi(r) = r^{\ell+1}\mathrm{e}^{-kr}\sum_{n=0}^{\infty} A_n r^n \tag{2.23}$$

which guarantees in both cases, small and large r, a correct behavior, as long as the series does not diverge exponentially.

2.2.1 Energy levels

The radial equation for the Coulomb potential can then be solved along the same lines as for the harmonic oscillator, Sec.1.1. The expansion of Eq.(2.23) is introduced into Eq.(2.14), a recursion formula for coefficients A_n is derived, one finds that the series in general diverges like $\exp(2kr)$ unless it is truncated to a finite number of terms, and this happens only for some particular values of E:

$$E_n = -\frac{Z^2}{n^2}\frac{m_e q_e^4}{2\hbar^2} = -\frac{Z^2}{n^2}\,\mathrm{Ry} \tag{2.24}$$

where $n \geq \ell + 1$ is an integer known as *main quantum number*. For a given ℓ we find solutions for $n = \ell + 1, \ell + 2, \ldots$; or, for a given n, possible values for ℓ are $\ell = 0, 1, \ldots, n-1$.

Although the effective potential appearing in Eq.(2.14) depends upon ℓ, and the angular part of the wave function also strongly depends upon ℓ, the energies, Eq.(2.24) depend only upon n. We have thus a degeneracy on the energy levels with the same n and different ℓ, in addition to the one due to the $2\ell + 1$ possible values of the quantum number m (as implied in Eq.(2.8) where m does not appear). The total degeneracy (not considering spin) for a given n is thus

$$\sum_{\ell=0}^{n-1}(2\ell + 1) = n^2. \tag{2.25}$$

2.2.2 Radial wave functions

Finally, the solution for the radial part of the wave functions is

$$\chi_{n\ell}(r) = \sqrt{\frac{(n-\ell-1)!Z}{n^2[(n+\ell)!]^3 a_0^3}}\, x^{\ell+1}\mathrm{e}^{-x/2}L_{n+1}^{2\ell+1}(x) \tag{2.26}$$

where

$$x \equiv \frac{2Z}{n}\frac{r}{a_0} = 2\sqrt{-\frac{2m_e E_n}{\hbar^2}}\, r \tag{2.27}$$

and $L_{n+1}^{2\ell+1}(x)$ are *Laguerre polynomials* of degree $n - \ell - 1$. The coefficient has been chosen in such a way that the following orthonormality relations hold:

$$\int_0^\infty \chi_{n\ell}(r)\chi_{n'\ell}(r)dr = \delta_{nn'}. \tag{2.28}$$

The ground state has $n = 1$ and $\ell = 0$: $1s$ in "spectroscopic" notation ($2p$ is $n = 2$, $\ell = 1$, $3d$ is $n = 3$, $\ell = 2$, $4f$ is $n = 4$, $\ell = 3$, and so on). For the Hydrogen atom ($Z = 1$) the ground state energy is -1Ry and the binding energy of the electron is 1 Ry (apart from the small correction due to the difference between electron mass and reduced mass). The wave function of the ground state is a simple exponential. With the correct normalization:

$$\psi_{100}(r, \theta, \phi) = \frac{Z^{3/2}}{a_0^{3/2}\sqrt{\pi}}\mathrm{e}^{-Zr/a_0}. \tag{2.29}$$

The dominating term close to the nucleus is the first term of the series, $\chi_{n\ell}(r) \sim r^{\ell+1}$. The larger ℓ, the quicker the wave function tends to zero when approaching the nucleus. This reflects the fact that the function is "pushed away" by the centrifugal potential. Thus radial wave functions with large ℓ do not appreciably penetrate close to the nucleus.

At large r the dominating term is $\chi(r) \sim r^n \exp(-Zr/na_0)$. This means that, neglecting the other terms, $|\chi_{n\ell}(r)|^2$ has a maximum about $r = n^2 a_0/Z$. This gives a rough estimate of the "size" of the wave function, which is mainly determined by n.

In Eq.(2.26) the polynomial has $n-\ell-1$ degree. This is also the number of nodes of the function. In particular, the eigenfunctions with $\ell = 0$ have $n-1$ nodes; those with $\ell = n-1$ are node-less. The form of the radial functions can be seen for instance on the Wolfram Research site[1] or explored via the Java applet at Davidson College[2]

2.3 Code: hydrogen_radial

The code `hydrogen_radial.f90`[3] or `hydrogen_radial.c`[4] solves the radial equation for a one-electron atom. It is based on `harmonic1`, but solves a slightly different equation on a logarithmically spaced grid. Moreover it uses a more sophisticated approach to locate eigenvalues, based on a perturbative estimate of the needed correction.

The code uses atomic (Rydberg) units, so lengths are in Bohr radii ($a_0 = 1$), energies in Ry, $\hbar^2/(2m_e) = 1$, $q_e^2 = 2$.

2.3.1 Logarithmic grid

The straightforward numerical solution of Eq.(2.14) runs into the problem of the singularity of the potential at $r = 0$. One way to circumvent this difficulty is to work with a variable-step grid instead of a constant-step one, as done in the previous chapter. Such grid becomes denser and denser as we approach the origin. Real-life solutions of the radial Schrödinger equation in atoms, especially heavy one, invariably involve such kind of grids, since wave functions close to the nucleus vary on a much smaller length scale than far from the nucleus. A detailed description of the scheme presented here can be found in chap.6 of *The Hartree-Fock method for atoms*, C. Froese Fischer, Wiley, 1977.

Let us introduce a new integration variable x and a constant-step grid in x, so as to be able to use Numerov's method without changes. We define a mapping between r and x via

$$x = x(r). \tag{2.30}$$

The relation between the constant-step grid spacing Δx and the variable-step grid spacing is given by

$$\Delta x = x'(r)\Delta r. \tag{2.31}$$

[1] http://library.wolfram.com/webMathematica/Physics/Hydrogen.jsp

[2] http://webphysics.davidson.edu/physlet_resources/cise_qm/html/hydrogenic.html

[3] http://www.fisica.uniud.it/%7Egiannozz/Didattica/MQ/Software/F90/hydrogen_radial.f90

[4] http://www.fisica.uniud.it/%7Egiannozz/Didattica/MQ/Software/C/hydrogen_radial.c

We make the specific choice

$$x(r) \equiv \log \frac{Zr}{a_0} \tag{2.32}$$

(note that with this choice x is adimensional) yielding

$$\Delta x = \frac{\Delta r}{r}. \tag{2.33}$$

The $\Delta r/r$ ratio remains thus constant on the grid of r so defined. This kind of grid is usually called *logarithmic grid*, but some people see the thing the other way round and call it *exponential*!

There is however a problem: by transforming Eq.(2.14) in the new variable x, a term with first derivative appears, preventing the usage of Numerov's method (and of other integration methods as well). The problem can be circumvented by transforming the unknown function as follows:

$$y(x) = \frac{1}{\sqrt{r}} \chi\left(r(x)\right). \tag{2.34}$$

It is easy to verify that by transforming Eq.(2.14) so as to express it as a function of x and y, the terms containing first-order derivatives disappear, and by multiplying both sides of the equation by $r^{3/2}$ one finds

$$\frac{d^2y}{dx^2} + \left[\frac{2m_e}{\hbar^2} r^2 \left(E - V(r)\right) - \left(\ell + \frac{1}{2}\right)^2\right] y(x) = 0 \tag{2.35}$$

where $V(r) = -Zq_e^2/r$ for the Coulomb potential. This equation no longer presents any singularity for $r = 0$, is in the form of Eq.(1.22), with

$$g(x) = \frac{2m_e}{\hbar^2} r(x)^2 \left(E - V(r(x))\right) - \left(\ell + \frac{1}{2}\right)^2 \tag{2.36}$$

and can be directly solved using the numerical integration formulae Eqs.(1.32) and (1.33) and an algorithm very similar to the one of Sec.1.3.2.

Subroutine `do_mesh` computes and stores the values of r, $\sqrt{r}$, r^2 for each grid point. The potential is also calculated and stored in `init_pot`. The grid is calculated starting from a minimum value, set in the code to $x = -8$, corresponding to $Zr_{\min} \simeq 3.4 \times 10^{-3}$ Bohr radii. Note that the grid in r does not include $r = 0$: this would correspond to $x = -\infty$. The known analytical behavior for $r \to 0$ and $r \to \infty$ are used to start the outward and inward recurrences, respectively.

2.3.2 Improving convergence with perturbation theory

A few words are needed to explain this section of the code:

```
i = icl
ycusp = (y(i-1)*f(i-1)+f(i+1)*y(i+1)+10.d0*f(i)*y(i)) / 12.d0
dfcusp = f(i)*(y(i)/ycusp - 1.d0)
! eigenvalue update using perturbation theory
de = dfcusp/ddx12 * ycusp*ycusp * dx
```

whose goal is to give an estimate, to first order in perturbation theory, of the difference δe between the current estimate of the eigenvalue and its final value.

Reminder: `icl` is the index corresponding to the classical inversion point. Integration is made with forward recursion up to this index, with backward recursion down to this index. `icl` is thus the index of the matching point between the two functions. The function at the right is rescaled so that the total function is continuous, but the first derivative dy/dx will be in general discontinuous, unless we have reached a good eigenvalue.

In the section of the code shown above, `y(icl)` is the value given by Numerov's method using either `icl-1` or `icl+1` as central point; `ycusp` is the value predicted by the Numerov's method using `icl` as central point. The problem is that `ycusp`$\neq$`y(icl)`.

What about if our function is the *exact solution*, but for a *different problem*? It is easy to find what the different problem could be: one in which a delta function, $v_0\delta(x-x_c)$, is superimposed at $x_c \equiv$`x(icl)` to the potential. The presence of a delta function causes a discontinuity (a "cusp") in the first derivative, as can be demonstrated by a limit procedure, and the size of the discontinuity is related to the coefficient of the delta function. Once the latter is known, we can give an estimate, based on perturbation theory, of the difference between the current eigenvalue (for the "different" potential) and the eigenvalue for the "true" potential.

One may wonder how to deal with a delta function in numerical integration. In practice, we assume the delta to have a value only in the interval Δx centered on `y(icl)`. The algorithm used to estimate its value is quite sophisticated. Let us look again at Numerov's formula, Eq.(1.33): note that the formula actually provides only the product `y(icl)f(icl)`. From this we usually extract `y(icl)` since `f(icl)` is assumed to be known. Now we suppose that `f(icl)` has a different and unknown value `fcusp`, such that our function satisfies Numerov's formula also in point `icl`. The following must hold:

```
fcusp*ycusp = f(icl)*y(icl)
```

since this product is provided by Numerov's method (by integrating from `icl-1` to `icl+1`), and `ycusp` is that value that the function `y` must have in order to satisfy Numerov's formula also in `icl`. As a consequence, the value of `dfcusp` calculated by the program is just `fcusp-f(icl)`, or δf.

The next step is to calculate the variation δV of the potential $V(r)$ appearing in Eq.(2.35) corresponding to δf. By differentiating Eq.(2.36) one finds $\delta g(x) = -(2m_e/\hbar^2)r^2\delta V$. Since $f(x) = 1 + g(x)(\Delta x)^2/12$, we have $\delta g = 12/(\Delta x)^2\delta f$, and thus

$$\delta V = -\frac{\hbar^2}{2m_e}\frac{1}{r^2}\frac{12}{(\Delta x)^2}\delta f. \tag{2.37}$$

First-order perturbation theory gives then the corresponding variation of the eigenvalue:

$$\delta e = \langle\psi|\delta V|\psi\rangle = \int |y(x)|^2 r(x)^2 \delta V dx. \tag{2.38}$$

Note that the change of integration variable from dr to dx adds a factor r to the integral:

$$\int_0^\infty f(r)dr = \int_{-\infty}^{\infty} f(r(x))\frac{dr(x)}{dx}dx = \int_{-\infty}^{\infty} f(r(x))r(x)dx. \tag{2.39}$$

We write the above integral as a finite sum over grid points, with a single non-zero contribution coming from the region of width Δx centered at point $x_c =$`x(icl)`. Finally:

$$\delta e = |y(x_c)|^2 r(x_c)^2 \delta V \Delta x = -\frac{\hbar^2}{2m_e}\frac{12}{(\Delta x)^2}|y(x_c)|^2 \Delta x \delta f \tag{2.40}$$

is the difference between the eigenvalue of the current potential (i.e. with a superimposed Delta function) and that of the true potential. This expression is used by the code to calculate the correction `de` to the eigenvalue. Since in the first step this estimate may have large errors, the line

```
e = max(min(e+de,eup),elw)
```

prevents the usage of a new energy estimate outside the bounds `[elw,eip]`. As the code proceeds towards convergence, the estimate becomes better and better and convergence is very fast in the final steps.

2.3.3 Laboratory

- Examine solutions as a function of n and ℓ; verify the presence of accidental degeneracy.
- Examine solutions as a function of the nuclear charge Z.
- Compare the numerical solution with the exact solution, Eq.(2.29), for the $1s$ case (or other cases if you know the analytic solution).
- Slightly modify the potential as defined in subroutine `init_pot`, verify that the accidental degeneracy disappears. Some suggestions: $V(r) = -Zq_e^2/r^{1+\delta}$ where δ is a small, positive or negative, number; or add an exponential Yukawa damping, $V(r) = -Zq_e^2 \exp(-Qr)/r$, where Q is a number of the order of 0.05 a.u..
- Calculate the expectation values of r and of $1/r$, compare them with the known analytical results.

Possible code modifications and extensions:

- Consider a different mapping: $r(x) = r_0(\exp(x) - 1)$, that unlike the one we have considered, includes $r = 0$. Which changes must be done in order to adapt the code to this mapping?

Chapter 3

Scattering from a potential

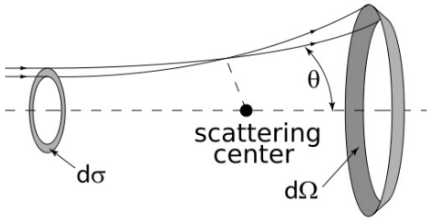
Until now we have considered the discrete energy levels of simple, one-electron Hamiltonians, corresponding to bound, localized states. Unbound, delocalized states exist as well for any physical potential (with the exception of idealized models like the harmonic potential) at sufficiently high energies. These states are relevant in the description of *scattering* from a potential, i.e. processes of diffusion of an incoming particle. Scattering is a really important subject in physics: what many experiments measure is how a particle is deflected by another. The comparison of measurements with calculated results makes it possible to understand the form of the interaction potential between the particles. In the following a short reminder of scattering theory is provided; then an application to a real problem (scattering of H atoms by rare gas atoms) is presented. This chapter is inspired to Ch.2 of the book of Thijssen.

3.1 Short reminder of the theory of scattering

The problem of scattering of a particle by another is first mapped onto the equivalent problem of scattering from a fixed center, with the usual coordinate change to relative and center-of-mass coordinates. In the following we consider *elastic* scattering. In the typical geometry, a free particle, described as a plane wave with wave vector along the z axis, is incident on the center and is scattered as a spherical wave at large values of r (distance from the center). An interesting quantity characterizing the scattering is the *differential cross section* $d\sigma(\Omega)/d\Omega$, i.e. the probability that in the unit of time a particle crosses the surface in the surface element $dS = r^2 d\Omega$ (where Ω is the solid angle, $d\Omega = \sin\theta d\theta d\phi$, where θ is the polar angle and ϕ the azimuthal angle). Another useful quantity is the *total cross section* σ_{tot}: the integral of the differential cross section over all angles,

$$\sigma_{tot} = \int \left(\frac{d\sigma(\Omega)}{d\Omega} \right) d\Omega. \tag{3.1}$$

For a central potential, the system is symmetric around the z axis and thus the differential cross section does not depend upon ϕ. Both the differential and the total cross section depend upon the energy of the incident particle.



Let us consider a wave function having the form:

$$\psi(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} + \frac{f(\theta)}{r}e^{ikr}, \qquad k = \sqrt{\frac{2mE}{\hbar^2}} \tag{3.2}$$

with $\mathbf{k} = (0, 0, k)$, parallel to the z axis. This represents the incident particle of energy E, described by a plane wave, plus a diffused spherical wave with *scattering amplitude* $f(\theta)$. The cross section can be obtained from the quantum-mechanical probability current $\mathbf{j}$:

$$\mathbf{j} = \frac{i\hbar}{2m}\left(\psi\nabla\psi^* - \psi^*\nabla\psi\right). \tag{3.3}$$

For our scattering wave function we find

$$\mathbf{j} = \mathbf{j}_i + \mathbf{j}_s + \mathcal{O}(1/r^3), \qquad \mathbf{j}_i = \frac{\hbar\mathbf{k}}{m}, \quad \mathbf{j}_s = \frac{\hbar k}{m}\frac{\hat{\mathbf{r}}}{r^2}|f(\theta)|^2 \tag{3.4}$$

where $\mathbf{j}_i$ is the flux of the incident particle, $\mathbf{j}_s$ the flux of scattered particles, rapidly oscillating terms (that average to zero) are neglected. We can write the cross section as a function of $f(\theta)$ as:

$$\frac{d\sigma(\Omega)}{d\Omega}d\Omega = \frac{\mathbf{j}_s\cdot\hat{\mathbf{r}}}{j_i}r^2 d\Omega = |f(\theta)|^2\sin\theta d\theta d\phi. \tag{3.5}$$

Let us look for solutions of the form (3.2). The wave function is in general given by the following expression:

$$\psi(\mathbf{r}) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} A_{lm}\frac{\chi_l(r)}{r}Y_{lm}(\theta,\phi), \tag{3.6}$$

which in our case, given the symmetry of the problem, can be simplified as

$$\psi(\mathbf{r}) = \sum_{l=0}^{\infty} A_l\frac{\chi_l(r)}{r}P_l(\cos\theta). \tag{3.7}$$

The functions $\chi_l(r)$ are solutions for (positive) energy $E = \hbar^2k^2/2m$ with angular momentum l of the radial Schrödinger equation:

$$\frac{\hbar^2}{2m}\frac{d^2\chi_l(r)}{dr^2} + \left[E - V(r) - \frac{\hbar^2 l(l+1)}{2mr^2}\right]\chi_l(r) = 0. \tag{3.8}$$

The asymptotic behavior at large r of $\chi_l(r)$ can be written as

$$\chi_l(r) \simeq kr\left[j_l(kr)\cos\delta_l - n_l(kr)\sin\delta_l\right] \tag{3.9}$$

where the j_l and n_l functions are the well-known *spherical Bessel functions*, respectively regular and divergent at the origin. These are the $R_l(r) = \chi_l(r)/r$ solutions of the radial equation for the free particle ($V(r) = 0$). The quantities δ_l are known as *phase shifts* and depend upon the potential and the energy.

The cross section can then be expressed in terms of the phase shifts. One looks for coefficients of Eq.(3.7) that yield the desired asymptotic behavior (3.2), using the following series expansion of a plane wave:

$$e^{i\mathbf{k}\cdot\mathbf{r}} = \sum_{l=0}^{\infty} i^l (2l+1) j_l(kr) P_l(\cos\theta), \tag{3.10}$$

where the P_l functions are Legendre polynomials and θ, angle between $\mathbf{k}$ and $\mathbf{r}$, is the same we have introduced earlier. One finds $A_l = i^l(2l+1)/k$. With some further algebra, and using the large-x behavior of spherical Bessel functions:

$$j_l(x) \simeq \frac{1}{x}\sin\left(x - \frac{\pi l}{2}\right), \qquad n_l(x) \simeq \frac{1}{x}\cos\left(x - \frac{\pi l}{2}\right), \tag{3.11}$$

one finally demonstrates that the differential cross section can be written as

$$\frac{d\sigma}{d\Omega} = \frac{1}{k^2}\left|\sum_{l=0}^{\infty}(2l+1)e^{i\delta_l}\sin\delta_l P_l(\cos\theta)\right|^2, \tag{3.12}$$

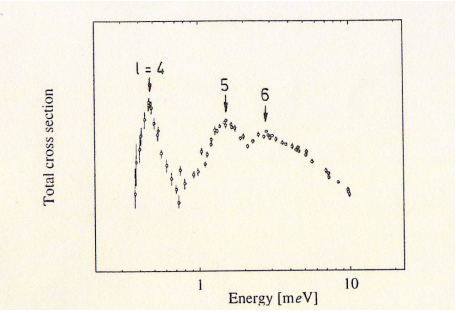
while the total cross section is given by

$$\sigma_{tot} = \frac{4\pi}{k^2}\sum_{l=0}^{\infty}(2l+1)\sin^2\delta_l \tag{3.13}$$

The energy- and angular momentum-dependent phase shifts thus contain *all* the information on the scattering properties of a potential.

3.2 Scattering of H atoms from rare gases

The total cross section $\sigma_{tot}(E)$ for the scattering of H atoms by rare gas atoms was measured by Toennies *et al.*, J. Chem. Phys. **71**, 614 (1979). At the right, the cross section for the H-Kr system as a function of the energy of the center of mass. One can notice "spikes" in the cross section, known as "resonances". One can connect the different resonances to specific values of the angular momentum l.



The H-Kr interaction potential can be modelled quite accurately as a Lennard-Jones (LJ) potential:

$$V(r) = \epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - 2\left(\frac{\sigma}{r}\right)^{6}\right] \tag{3.14}$$

where $\epsilon = 5.9$meV, $\sigma = 3.57$Å. The LJ potential is much used in molecular and solid-state physics to model interatomic interaction forces. The attractive $-1/r^6$ term describes weak van der Waals (or "dispersive", in chemical parlance) forces due to (dipole-induced dipole) interactions (see also Appendix C). The repulsive $1/r^{12}$ term models the repulsion between closed shells. While usually dominated by direct electrostatic interactions, the ubiquitous van der Waals forces are the dominant term for the interactions between closed-shell atoms and molecules. These play an important role in molecular crystals and in macromolecules. The LJ potential is the first realistic interatomic potential for which a molecular-dynamics simulation was performed (Rahman, 1965, liquid Ar).

It is straightforward to find that the LJ potential as written in Eq.(3.14) has a minimum $V_{min} = -\epsilon$ for $r = \sigma$, is zero for $r = \sigma/2^{1/6} = 0.89\sigma$ and becomes strongly positive (i.e. repulsive) at small r.

3.3 Code: crossection

Code `crossection.f90`[1], or `crossection.c`[2], calculates the total cross section $\sigma_{tot}(E)$ and its decomposition into contributions of the various values of the angular momentum for a scattering problem like the one described before.

The code is composed of different parts. It is always a good habit to verify separately the correct behavior of each piece before assembling them into the final code (this is how professional software is tested, by the way). The various parts are:

1. Solution of the radial Schrödinger equation, Eq.(3.8), with the Lennard-Jones potential of Eq.(3.14) for scattering states (i.e. withpositive energy). One can use Numerov's method with outwards integration only: there is no danger of numerical instability, since the solution is oscillating. The solution does not exhibit the wide variations in the spatial behavior of the Coulomb potential case, so simple uniform grid can be used. One has however to be careful with the $\sim r^{-12}$ divergence for $r \to 0$ and in general with the small-r region. The wave function decays very strongly towards zero for $r \to 0$, which means that it grows very quickly when integrated outwards. This may easily lead to overflow errors (numbers exceeding the maximum allowed size, typically 10^{308} for double precision). A simple way to avoid trouble is to start the integration from $r_{min} \sim 0.5\sigma$, where the wave function is very small but not too close to zero. The first two points can be calculated in the following way, [3] by assuming the asymptotic (vanishing) form for $r \to 0$:

$$\chi''(r) \simeq \frac{2m\epsilon}{\hbar^2}\frac{\sigma^{12}}{r^{12}}\chi(r) \quad \Longrightarrow \quad \chi(r) \simeq \exp\left(-\sqrt{\frac{2m\epsilon\sigma^{12}}{25\hbar^2}}r^{-5}\right). \quad (3.15)$$

[1] http://www.fisica.uniud.it/%7Egiannozz/Didattica/MQ/Software/F90/crossection.f90

[2] http://www.fisica.uniud.it/%7Egiannozz/Didattica/MQ/Software/C/crossection.c

[3] Ronald Cohen (Carnegie) noticed that this procedure introduces an error of lower order, i.e. worse, than that of the Numerov's algorithm. In fact by assuming such form for the first two steps of the recursion, we use a solution that is neither analytically exact, nor consistent with Numerov's algorithm.

The choice of the units in this code is (once again) different from that of the previous codes. It is convenient to choose units in which $\hbar^2/2m$ is a number of the order of 1. Two possible choices are meV·Å^2, or meV·σ^2. This code uses the former choice. Note that m here is not the electron mass! it is the reduced mass of the H-Kr system. As a first approximation, m is here the mass of H.

2. Calculation of the phase shifts $\delta_l(E)$. Phase shifts can be calculated by comparing the computed wave functions with the expected asymptotic solution at two different values r_1 and r_2, both larger than the distance r_{max} beyond which the potential can be considered to be negligible. Let us write

$$\chi_l(\mathbf{r}_1) = Akr_1 \left[j_l(kr_1) \cos\delta_l - n_l(kr_1) \sin\delta_l \right] \tag{3.16}$$
$$\chi_l(\mathbf{r}_2) = Akr_2 \left[j_l(kr_2) \cos\delta_l - n_l(kr_2) \sin\delta_l \right], \tag{3.17}$$

from which, by dividing the two relations, we obtain an auxiliary quantity K

$$K \equiv \frac{r_2 \chi_l(\mathbf{r}_1)}{r_1 \chi_l(\mathbf{r}_2)} = \frac{j_l(kr_1) - n_l(kr_1)\tan\delta_l}{j_l(kr_2) - n_l(kr_2)\tan\delta_l} \tag{3.18}$$

from which we can deduce the phase shift:

$$\tan\delta_l = \frac{K j_l(kr_2) - j_l(kr_1)}{K n_l(kr_2) - n_l(kr_1)}. \tag{3.19}$$

The choice of r_1 and r_2 is not very critical but requires some care. r_1 can be chosen at r_{max}, for which, given the fast decay of the LJ potential for large r, a good choice is $r_{max} = 5\sigma$. r_2 should not be too close to r_1. A good choice seems to be $r_2 = r_1 + \lambda/4$, that is, at 1/4 of the wave length $\lambda = 2\pi/k$ of the scattered wave function from r_1. [4]

3. Calculation of the spherical Bessel functions j_l and n_l. The analytical forms of these functions are known, but they become quickly unwieldy for high l. One can use recurrence relation. In the code the following simple recurrence is used:

$$z_{l+1}(x) = \frac{2l+1}{x} z_l(x) - z_{l-1}(x), \qquad z = j, n \tag{3.20}$$

with the initial conditions

$$j_{-1}(x) = \frac{\cos x}{x}, \quad j_0(x) = \frac{\sin x}{x}; \qquad n_{-1}(x) = \frac{\sin x}{x}, \quad n_0(x) = -\frac{\cos x}{x}. \tag{3.21}$$

Note that this recurrence is unstable for large values of l: in addition to the good solution, it has a "bad" divergent solution. This is not a serious problem in practice: the above recurrence should be sufficiently stable up to at least $l = 20$ or so, but you may want to verify this.

[4] A previous choice, half the wavelength, was occasionally giving numerical problems: the resulting δ was fitting very well in the end points, much less so in the middle.

4. Sum the phase shifts as in Eq.(3.13) to obtain the total cross section and a graph of it as a function of the incident energy. The relevant range of energies is of the order of a few meV, something like $0.1 \leq E \leq 3 \div 4$ meV. If you go too close to $E = 0$, you will run into numerical difficulties (the wave length of the wave function diverges). The angular momentum ranges from 0 to a value of l_{max} to be determined.

The code writes a file containing the total cross section and each angular momentum contribution as a function of the energy (beware: up to $l_{max} = 13$; for larger l, lines will be wrapped).

3.3.1 Laboratory

- Verify the effect of all the parameters of the calculation: grid step for integration, r_{min}, $r_1 = r_{max}$, r_2, l_{max}. In particular the latter is important: you should start to find good results when $l_{max} \geq 6$.

- Print, for some selected values of the energy, the wave function and its asymptotic behavior. You should verify that they match beyond r_{max}.

- Observe the contributions of the various l and the peaks for $l = 4, 5, 6$ (resonances). Make a plot of the effective potential as a function of l: can you see why there are resonances only for a few values of l?

Possible code modifications and extensions:

- Modify the code to calculate the cross section from a different potential, for instance, the following one:

$$V(r) = -A \exp\left[-(r - r_0)^2\right], \quad r < r_{max}; \qquad V(r > r_{max}) = 0$$

  What changes do you need to apply to the algorithm, in addition to changing the form of the potential?

- In the limit of a weak potential (such as e.g. the potential introduced above), the phase shifts δ_l are well approximated by the *Born approximation*:

$$\delta_l \simeq -\frac{2m}{\hbar^2} k \int_0^\infty r^2 j_l^2(kr) V(r) dr,$$

  ($k = \sqrt{2mE/\hbar}$). Write a simple routine to calculate this integral, compare with numerical results.

Chapter 4

The Variational Method

The exact analytical solution of the Schrödinger equation is possible only in a few cases. Even the direct numerical solution by integration is often not feasible in practice, especially in systems with more than one particle. There are however extremely useful approximated methods that can in many cases reduce the complete problem to a much simpler one. In the following we will consider the *variational principle* and its consequences. This constitutes, together with suitable approximations for the electron-electron interactions, the basis for most practical approaches to the solution of the Schrödinger equation in condensed-matter physics.

4.1 The variational principle

Let us consider a Hamiltonian H and a function ψ, that can be varied at will with the sole condition that it stays normalized. In general, ψ is not an eigenfunction of H, but we can calculate the expectation value of the energy for such function

$$\langle H \rangle = \int \psi^* H \psi \, dv \tag{4.1}$$

where v represents all the integration coordinates.

The *variational principle* states that functions ψ for which $\langle H \rangle$ is *stationary*—i.e. does not vary to first order in small variations of ψ—are the eigenfunctions of the energy. In other words, the Schrödinger equation is equivalent to a stationarity condition.

4.1.1 Demonstration of the variational principle

Let us consider the wave function ψ and an arbitrary variation of it, $\delta\psi$. In general, $\psi + \delta\psi$ is no longer normalized even is ψ is. It is thus convenient to use a more general definition of expectation value, valid also for non-normalized functions:

$$\langle H \rangle = \frac{\int \psi^* H \psi \, dv}{\int \psi^* \psi \, dv} \tag{4.2}$$

The expectation value of H over wavefunction $\psi + \delta\psi$ becomes

$$\begin{aligned}
\langle H\rangle + \delta\langle H\rangle &= \frac{\int(\psi^* + \delta\psi^*)H(\psi+\delta\psi)\,dv}{\int(\psi^*+\delta\psi^*)(\psi+\delta\psi)\,dv} \\
&\simeq \frac{\int \psi^* H\psi\,dv + \int \delta\psi^* H\psi\,dv + \int \psi^* H\delta\psi\,dv}{\int \psi^*\psi\,dv + \int \delta\psi^*\psi\,dv + \int \psi^*\delta\psi\,dv} \\
&= \left(\int \psi^* H\psi\,dv + \int \delta\psi^* H\psi\,dv + \int \psi^* H\delta\psi\,dv\right)\times \\
&\quad \frac{1}{\int \psi^*\psi\,dv}\left(1 - \frac{\int \delta\psi^*\psi\,dv}{\int \psi^*\psi\,dv} - \frac{\int \psi^*\delta\psi\,dv}{\int \psi^*\psi\,dv}\right)
\end{aligned} \tag{4.3}$$

where second-order terms in $\delta\psi$ have been omitted and we have used the approximation $1/(1+x) \simeq 1-x$, valid for $x << 1$. By omitting again higher-order terms, one finds for the variation $\delta\langle H\rangle$:

$$\delta\langle H\rangle = \frac{\int \delta\psi^* H\psi\,dv}{\int \psi^*\psi\,dv} + \frac{\int \psi^* H\delta\psi\,dv}{\int \psi^*\psi\,dv} - \langle H\rangle\left(\frac{\int \delta\psi^*\psi\,dv}{\int \psi^*\psi\,dv} + \frac{\int \psi^*\delta\psi\,dv}{\int \psi^*\psi\,dv}\right). \tag{4.4}$$

One of the two terms in parentheses is the complex conjugate of the other; the same holds for the two remaining terms, because H is a hermitian operator, satisfying

$$\int a^* H b\,dv = \left(\int b^* H a\,dv\right)^* \tag{4.5}$$

for any pair of functions a and b. We can thus simplify the above expression as

$$\delta\langle H\rangle = \left(\frac{\int \delta\psi^* H\psi\,dv}{\int \psi^*\psi\,dv} + \text{c.c.}\right) - \langle H\rangle\left(\frac{\int \delta\psi^*\psi\,dv}{\int \psi^*\psi\,dv} + \text{c.c.}\right) \tag{4.6}$$

and finally to

$$\delta\langle H\rangle = \frac{\int \delta\psi^*\left[H - \langle H\rangle\right]\psi\,dv + \text{c.c.}}{\int \psi^*\psi\,dv}. \tag{4.7}$$

This immediately shows that if ψ is a solution of the Schrödinger equation, $\delta\langle H\rangle = 0$ and $\langle H\rangle$ is stationary. Let us now consider the opposite implication: if $\langle H\rangle$ is stationary with respect to any variation of ψ, then

$$\int \delta\psi^*\left[H - \langle H\rangle\right]\psi\,dv = 0 \tag{4.8}$$

is true for any arbitrary $\delta\psi$ if and only if $\left[H - \langle H\rangle\right]\psi = 0$, that is, if ψ is a solution of the Schrödinger equation: $H\psi = E\psi$.

4.1.2 Alternative demonstration of the variational principle

A different and more general way to demonstrate the variational principle, which will be useful later, is based upon Lagrange multipliers method. This method deals with the problem of finding stationarity conditions for an integral I_0 while keeping at the same time constant other integrals $I_1 \ldots I_k$. One can solve the equivalent problem

$$\delta\left(I_0 + \sum_k \lambda_k I_k\right) = 0 \tag{4.9}$$

where λ_k are additional variables called *Lagrange multipliers*. In our case we have

$$I_0 = \int \psi^* H \psi \, dv \tag{4.10}$$

$$I_1 = \int \psi^* \psi \, dv \tag{4.11}$$

and thus we assume

$$\delta(I_0 + \lambda I_1) = 0 \tag{4.12}$$

where λ must be determined. By proceeding like in the previous section, we find

$$\delta I_0 = \int \delta\psi^* H \psi \, dv + \text{c.c.} \tag{4.13}$$

$$\delta I_1 = \int \delta\psi^* \psi \, dv + \text{c.c.} \tag{4.14}$$

and thus the condition to be satisfied is

$$\delta(I_0 + \lambda I_1) = \int \delta\psi^* [H + \lambda] \psi \, dv + \text{c.c.} = 0 \tag{4.15}$$

that is

$$H\psi = -\lambda\psi \tag{4.16}$$

i.e. the Lagrange multiplier is equal, apart from the sign, to the energy eigenvalue. Again we see that states whose expectation energy is stationary with respect to any variation in the wave function are the solutions of the Schrödinger equation.

4.1.3 Variational principle for the ground state energy

Let us consider the eigenfunctions ψ_n of a Hamiltonian H, with associated eigenvalues (energies) E_n:

$$H\psi_n = E_n \psi_n. \tag{4.17}$$

We label the ground state with $n = 0$ and the ground-state energy as E_0. Let us demonstrate that for any different function ψ, we necessarily have

$$\langle H \rangle = \frac{\int \psi^* H \psi \, dv}{\int \psi^* \psi \, dv} \geq E_0. \tag{4.18}$$

In order to demonstrate it, let us expand ψ into energy eigenfunctions (this is always possible because energy eigenfunctions are a complete orthonormal set):

$$\psi = \sum_n c_n \psi_n. \tag{4.19}$$

Then one finds

$$\langle H \rangle = \frac{\sum_n |c_n|^2 E_n}{\sum_n |c_n|^2} = E_0 + \frac{\sum_n |c_n|^2 (E_n - E_0)}{\sum_n |c_n|^2} \tag{4.20}$$

This demonstrates Eq.(4.18), since the second term is either positive or zero, as $E_n \geq E_0$ by definition of ground state. Note that if the ground state is non-degenerate, the inequality is strict: $\langle H \rangle = E_0$ only for $\psi = \psi_0$.

This simple result is extremely important: it tells us that any function ψ yields for the expectation energy an upper estimate of the energy of the ground state. The best approximation to the ground state is can be found by varying ψ inside a given set of functions and looking for the function that minimizes $\langle H \rangle$. This is the essence of the variational method.

A simple way to use such result is to introduce a set of normalized *trial* wave functions $\psi(v; \alpha_1, \ldots, \alpha_r)$, where v are the variables of the problem (coordinates etc), $\alpha_i, i = 1, \ldots, r$ are parameters. The ground-state energy will be a function of the parameters:

$$E(\alpha_1, \ldots, \alpha_r) = \int \psi^*(v; \alpha_1, \ldots, \alpha_r) H \psi(v; \alpha_1, \ldots, \alpha_r)\, dv \tag{4.21}$$

We now look for the minimum of E with respect to a variation of the parameters, that is, we impose

$$\frac{\partial E}{\partial \alpha_1} = \ldots = \frac{\partial E}{\partial \alpha_r} = 0. \tag{4.22}$$

The function ψ satisfying these conditions with the lowest E is the function that better approximates the ground state, among the considered set of trial wave functions. It is clear that the quality of the results depends in a crucial way upon a suitable choice of the set of trial wave functions.

4.2 The variational method in practice

The variational method leads to the solution of an algebraic problem if the wave function is expanded into a finite *basis set* of functions. By applying the variational principle, one finds the optimal coefficients of the expansion. Based on Eq. (4.9), this means calculating the *functional* (i.e. a "function" of a function):

$$\begin{aligned} G[\psi] &= \langle \psi | H | \psi \rangle - E \langle \psi | \psi \rangle \\ &= \int \psi^* H \psi \, dv - E \int \psi^* \psi \, dv \end{aligned} \tag{4.23}$$

and imposing the stationary condition on $G[\psi]$. Such procedure produces an equation for the expansion coefficients that we are going to determine.

It is important to notice that our basis is formed by a *finite* number N of functions, and thus cannot be a complete system: in general, it is not possible to write any function ψ (including exact solutions of the Schrödinger equation) as a linear combination of the functions in this basis set. What we are going to do is to find the ψ function that better approaches the true ground state, or in general, a true eigenstate, among all functions that can be expressed as linear combinations of the N chosen basis functions. A smart choice of the basis functions is crucial to obtain a computationally efficient algorithm.

4.2.1 Expansion into a basis set of orthonormal functions

Let us assume to have a basis of N functions b_i, between which orthonormality relations hold:

$$\langle b_i|b_j\rangle \equiv \int b_i^* b_j \, dv = \delta_{ij} \tag{4.24}$$

Let us expand the generic ψ in such basis:

$$\psi = \sum_{i=1}^{N} c_i b_i \tag{4.25}$$

By replacing Eq.(4.25) into Eq.(4.23) one can immediately notice that the latter takes the form

$$\begin{aligned} G(c_1, \ldots, c_N) &= \sum_{ij} c_i^* c_j H_{ij} - E \sum_{ij} c_i^* c_j \delta_{ij} \\ &= \sum_{ij} c_i^* c_j (H_{ij} - E\delta_{ij}) \end{aligned} \tag{4.26}$$

where we have introduced the *matrix elements* H_{ij}:

$$H_{ij} = \langle b_i|H|b_j\rangle = \int b_i^* H b_j \, dv \tag{4.27}$$

Since both H and the basis functions are given, H_{ij} is a known square matrix of numbers. The hermiticity of the Hamiltonian operator implies that such matrix is hermitian:

$$H_{ji} = H_{ij}^* \tag{4.28}$$

(i.e. symmetric if all elements are real). According to the variational method, let us minimize Eq. (4.26) with respect to the coefficients:

$$\frac{\partial G}{\partial c_i^*} = 0 \tag{4.29}$$

This produces the condition

$$\sum_j (H_{ij} - E\delta_{ij}) c_j = 0 \tag{4.30}$$

If the derivative with respect to complex quantities bothers you: write the complex coefficients as sums of a real and an imaginary part $c_k = x_k + iy_k$, require that derivatives with respect to both x_k and y_k are zero. By exploiting hermiticity you will find a system

$$\begin{aligned} W_k + W_k^* &= 0 \\ -iW_k + iW_k^* &= 0 \end{aligned}$$

where $W_k = \sum_j (H_{kj} - E\delta_{kj}) c_j$, that allows as only solution $W_k = 0$.

We note that, if the basis were a complete (and thus infinite) system, this would be the Schrödinger equation in the Heisenberg representation. We have finally demonstrated that the same equations, for a finite basis set, yield the best approximation to the true solution according to the variational principle.

4.2.2 Secular equation

Eq.(4.30) is a system of N algebraic linear equations, homogeneous (there are no constant term) in the N unknown c_j. In general, this system has only the trivial, and obviously nonphysical, solution $c_j = 0$ for all coefficients. A non-zero solution exists if and only if the following condition on the determinant is fulfilled:

$$\det |H_{ij} - E\delta_{ij}| = 0 \tag{4.31}$$

Such condition implies that one of the equations is a linear combination of the others and the system has in reality $N - 1$ equations and N unknowns, thus admitting non-zero solutions.

Eq.(4.31) is known as *secular equation*. It is an algebraic equation of degree N in E (as it is evident from the definition of the determinant, with the main diagonal generating a term E^N, all other diagonals generating lower-order terms), that admits N roots, or *eigenvalues*. Eq.(4.30) can also be written in matrix form

$$H\mathbf{c} = E\mathbf{c} \tag{4.32}$$

where H is here the $N \times N$ matrix whose matrix elements are H_{ij}, $\mathbf{c}$ is the vector formed with c_i components. The solutions $\mathbf{c}$ are also called *eigenvectors*. For each eigenvalue there will be a corresponding eigenvector (known within a multiplicative constant, fixed by the normalization). We have thus N eigenvectors and we can write that there are N solutions:

$$\psi_k = \sum_i C_{ik} b_i, \qquad k = 1, \ldots, N \tag{4.33}$$

where C_{ik} is a matrix formed by the N vectors $\mathbf{c}^{(k)}$, written as columns and disposed side by side:

$$C \equiv \begin{pmatrix} \mathbf{c}^{(1)} & \mathbf{c}^{(1)} & \ldots & \mathbf{c}^{(N)} \end{pmatrix}. \tag{4.34}$$

Eq.(4.30) can be rewritten as:

$$\sum_j H_{ij} C_{jk} = E_k C_{ik}. \tag{4.35}$$

Eq.(4.32) is a common equation in linear algebra and there are standard methods to solve it. Given a matrix H, it is possible to obtain, using standard library routines, the C matrix and the set E_k of eigenvalues.

The solution process is usually known as *diagonalization*. This name comes from the following important property of C. Eq.(4.33) can be seen as a *transformation* of the N starting functions into another set of N functions, via a transformation matrix. It is possible to show that if the b_i functions are orthonormal, the ψ_k functions are orthonormal as well. Then the transformation is *unitary*, that is,

$$\sum_i C^*_{ij} C_{ik} = \delta_{jk} \tag{4.36}$$

holds. In matrix notations,

$$(C^{-1})_{ij} = C^*_{ji} \equiv C^\dagger_{ij} \tag{4.37}$$

that is, the inverse matrix is equal to the conjugate of the transpose matrix, or *hermitian conjugate* matrix. A matrix C having such property is called a *unitary* matrix and typically represents a rotation.

Let us consider now the matrix product $C^{-1}HC$ and let us calculate its elements:

$$(C^{-1}HC)_{kn} = \sum_{ij} (C^{-1})_{ki} H_{ij} C_{jn} = \sum_i C^*_{ik} \sum_j H_{ij} C_{jn} \tag{4.38}$$

from Eq.(4.37). Using Eqs.(4.35) and (4.36), one finally finds

$$(C^{-1}HC)_{kn} = \sum_i C^*_{ik} E_n C_{in} = E_n \sum_i C^*_{ik} C_{in} = E_n \delta_{kn} \tag{4.39}$$

The transformation matrix C reduces H to a diagonal matrix, whose non-zero N elements are the eigenvalues. We can thus see our eigenvalue problem as the search for a transformation that brings from the original basis to a new basis in which the H operator has a diagonal form, that is, it acts on the elements of the basis by simply multiplying them by a constant (as in the Schrödinger equation).

4.3 Plane-wave basis set

A good example of orthonormal basis set, and one commonly employed in physics, is the *plane-wave* basis set. This basis set is closely related to Fourier transforms and it can be easily understood if concepts from Fourier analysis are known.

A function $f(x)$ defined on the entire real axis can be always expanded into Fourier components, $\widetilde{f}(k)$:

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \widetilde{f}(k) e^{ikx} dk \tag{4.40}$$

$$\widetilde{f}(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{-ikx} dx. \tag{4.41}$$

For a function defined on a finite interval $[-a/2, a/2]$, we can instead write

$$f(x) = \frac{1}{\sqrt{a}} \sum_n \widetilde{f}(k_n) e^{ik_n x} \tag{4.42}$$

$$\widetilde{f}(k_n) = \frac{1}{\sqrt{a}} \int_{-a/2}^{a/2} f(x) e^{-ik_n x} dx \tag{4.43}$$

where $k_n = 2\pi n/a$, $n = 0, \pm 1, \pm 2, ...$ Note that the $f(x)$ function of Eq.(4.42) is by construction a periodic function, with period equal to a: $f(x + a) = f(x)$, as can be immediately verified. This implies that $f(-a/2) = f(+a/2)$

must hold (also known under the name of *periodic boundary conditions*). The expressions reported here are easily generalized to three or more dimensions. In the following only a simple one-dimensional case will be shown.

Let us define our plane-wave basis set $b_i(x)$ as:

$$b_i(x) = \frac{1}{\sqrt{a}} e^{ik_i x}, \qquad k_i = \frac{2\pi}{a} i, \quad i = 0, \pm 1, \pm 2, ..., \pm N. \tag{4.44}$$

The corresponding coefficients c_i for the wave function $\psi(x)$ are

$$c_i = \int_{-a/2}^{a/2} b_i^*(x)\psi(x)dx = \langle b_i|\psi\rangle, \qquad \psi(x) = \sum_i c_i b_i(x). \tag{4.45}$$

This base, composed of $2N+1$ functions, becomes a complete basis set in the limit $N \to \infty$. This is a consequence of well-known properties of Fourier series. It is also straightforward to verify that the basis is orthonormal: $S_{ij} = \langle b_i|b_j\rangle = \delta_{ij}$. The solution of the problem of a particle under a potential requires thus the diagonalization of the Hamiltonian matrix, whose matrix elements:

$$H_{ij} = \langle b_i|H|b_j\rangle = \langle b_i|\frac{p^2}{2m} + V(x)|b_j\rangle \tag{4.46}$$

can be trivially calculated. The kinetic term is *diagonal* (i.e. it can be represented by a diagonal matrix):

$$\langle b_i|\frac{p^2}{2m}|b_j\rangle = -\frac{\hbar^2}{2m}\int_{-a/2}^{a/2} b_i^*(x)\frac{d^2 b_j}{dx^2}(x)dx = \delta_{ij}\frac{\hbar^2 k_i^2}{2m}. \tag{4.47}$$

The potential term is nothing but the Fourier transform of the potential (apart from a multiplicative factor):

$$\langle b_i|V(x)|b_j\rangle = \frac{1}{a}\int_{-a/2}^{a/2} V(x)e^{-i(k_i-k_j)x}dx = \frac{1}{\sqrt{a}}\widetilde{V}(k_i - k_j). \tag{4.48}$$

A known property of Fourier transform ensures that the matrix elements of the potential tend to zero for large values of $k_i - k_j$. The decay rate will depend upon the spatial variation of the potential: faster for slowly varying potentials, and vice versa. Potentials and wave functions varying on a typical length scale λ have a significant Fourier transform up to $k_{max} \sim 2\pi/\lambda$. In this way we can estimate the number of plane waves needed to solve a problem.

4.4 Code: pwell

Let us consider the simple problem of a potential well with finite depth V_0:

$$V(x) = 0 \quad \text{per} \quad x < -\frac{b}{2}, x > \frac{b}{2} \tag{4.49}$$

$$V(x) = -V_0 \quad \text{per} \quad -\frac{b}{2} \le x \le \frac{b}{2} \tag{4.50}$$

with $V_0 > 0$, $b < a$. The matrix elements of the Hamiltonian are given by Eq.(4.47) for the kinetic part. by Eq.(4.48) for the potential. The latter can be explicitly calculated:

$$\langle b_i|V(x)|b_j\rangle = -\frac{1}{a}\int_{-b/2}^{b/2} V_0 e^{-i(k_i-k_j)x}dx \tag{4.51}$$

$$= -\frac{V_0}{a}\left.\frac{e^{-i(k_i-k_j)x}}{-i(k_i-k_j)}\right|_{-b/2}^{b/2} V_0 \tag{4.52}$$

$$= -\frac{V_0}{a}\frac{\sin\left(b(k_i-k_j)/2\right)}{(k_i-k_j)/2}, \qquad k_i \neq k_j. \tag{4.53}$$

The case $k_i = k_j$ must be separately treated, yielding

$$\widetilde{V}(0) = -\frac{V_0 b}{a}. \tag{4.54}$$

Code `pwell.f90`[1] (or `pwell.c`[2]) generates the k_i, fills the matrix H_{ij} and diagonalizes it. The code uses units in which $\hbar^2/2m = 1$ (e.g. atomic Rydberg units). Input data are: width (b) and depth (V_0) of the potential well, width of the box (a), number of plane waves ($2N+1$). On output, the code prints the three lowest energy levels; moreover it writes to file `gs-wfc.out` the wave function of the ground state.

4.4.1 Diagonalization routines

The practical solution of the secular equation, Eq.(4.31), is not done by naively calculating the determinant and finding its roots. Various well-established, robust and fast diagonalization algorithms are known. Typically they are based on the reduction of the original matrix to Hessenberg or tridiagonal form via successive transformations. All such algorithms require the entire matrix (or at least half, exploiting hermiticity) to be available in memory at the same time, plus some work arrays. The time spent in the diagonalization is practically independent on the content of the matrix and it is invariably of the order of $\mathcal{O}(N^3)$ floating-point operations for a $N \times N$ matrix, even if eigenvalues only and not eigenvectors are desired. Matrix diagonalization used to be a major bottleneck in computation, due to its memory and time requirements. With modern computers, diagonalization of 1000×1000 matrix is done in less than no time. Still, memory growing as N^2 and time as N^3 are a serious obstacle towards larger N. At the end of these lecture notes alternative approaches will be mentioned.

The computer implementation of diagonalization algorithms is also rather well-established. In our code we use subroutine `dsyev.f`[3] from the linear algebra library LAPACK[4]. Several subroutines from the basic linear algebra library

[1] http://www.fisica.uniud.it/%7Egiannozz/Didattica/MQ/Software/F90/pwell.f90
[2] http://www.fisica.uniud.it/%7Egiannozz/Didattica/MQ/Software/C/pwell.c
[3] http://www.fisica.uniud.it/%7Egiannozz/Didattica/MQ/Software/dsyev.f
[4] http://www.netlib.org/lapack/

BLAS[5] (collected here: `dgemm.f`[6]) are also required. `dsyev` implements reduction to tridiagonal form for a real symmetric matrix (`d`=double precision, `sy`=symmetric matrix, `ev`=calculate eigenvalues and eigenvectors). The usage of `dsyev` requires either linking to a pre-compiled LAPACK and BLAS libraries, or compilation of the Fortran version and subsequent linking. Instructions on the correct way to call `dsyev` are contained in the header of the subroutine. Beware: most diagonalization routines overwrite the input matrix!

For the C version of the code, it may be necessary to add an underscore (as in `dsyev_()`) in the calling program. Moreover, the tricks explained in Sec.0.1.4 are used to define matrices and to pass arguments to BLAS and LAPACK routines.

4.4.2 Laboratory

- Observe how the results converge with respect to the number of plane waves, verify the form of the wave function. Verify the energy you get for a known case. You may use for instance the following case: for $V_0 = 1$, $b = 2$, the exact result is $E = -0.4538$. You may (and should) also verify the limit $V_0 \to \infty$ (what are the energy levels?).

- Observe how the results converge with respect to a. Note that for values of a not so big with respect to b, the energy calculated with the variational method is *lower* than the exact value. Why is it so?

- Plot the ground-state wave function. You can also modify the code to write excited states. Do they look like what you expect?

Possible code modifications and extensions:

- Modify the code, adapting it to a potential well having a Gaussian form (whose Fourier transform can be analytically calculated: what is the Fourier transform of a Gaussian function?) For the same "width", which problem converges more quickly: the square well or the Gaussian well?

- We know that for a symmetric potential, i.e. $V(-x) = V(x)$, the solutions have a well-defined parity, alternating even and odd parity (ground state even, first excited state odd, and so on). Exploit this property to reduce the problem into two sub-problems, one for even states and one for odd states. Use sine and cosine functions, obtained by suitable combinations of plane waves as above. Beware the correct normalization and the $k_n = 0$ term! Why is this convenient? What is gained in computational terms?

[5] http://www.netlib.org/blas/

[6] http://www.fisica.uniud.it/%7Egiannozz/Didattica/MQ/Software/dgemm.f

Chapter 5

Non-orthonormal basis sets

In the variational method as seen in action in the previous chapter the wave function is expanded over a set of orthonormal basis functions. In many physically relevant cases, it is useful to adopt a non-orthonormal basis set instead. A paradigmatic case is the calculation of the electronic structure of molecules using atom-centered localized functions, in particular, Gaussian functions, as frequently done in Quantum Chemistry methods. In this chapter we consider the extension of the variational method to the case of non-orthonormal basis sets.

5.1 Variational method for non-orthonormal basis set

The variational method can be extended with no special difficulty to cover the case in which the basis is formed by functions that are not orthonormal, i.e. for which

$$S_{ij} = \langle b_i | b_j \rangle = \int b_i^* b_j \, dv \tag{5.1}$$

is not simply equal to δ_{ij}. The quantities S_{ij} are known as *overlap integrals*.

In principle, one can always derive an orthogonal basis set from a non-orthoginal one using an orthogonalization procedure such as the *Gram-Schmid* algorithm. Given a non-orthogonal basis set b_i, the corresponding orthogonal basis set $\tilde{b}_i$ is obtained as follows:

$$\tilde{b}_1 = b_1 \tag{5.2}$$

$$\tilde{b}_2 = b_2 - \tilde{b}_1 \langle \tilde{b}_1 | b_2 \rangle / \langle \tilde{b}_1 | \tilde{b}_1 \rangle \tag{5.3}$$

$$\tilde{b}_3 = b_3 - \tilde{b}_2 \langle \tilde{b}_2 | b_3 \rangle / \langle \tilde{b}_2 | \tilde{b}_2 \rangle - \tilde{b}_1 \langle \tilde{b}_1 | b_3 \rangle / \langle \tilde{b}_1 | \tilde{b}_1 \rangle \tag{5.4}$$

and so on. The $\tilde{b}_i$ are then normalized. In practice, such procedure is computationally expensive and numerically not very stable so usually one prefers to adapt the approach of Sec.4.2.1 to a non-orthonormal basis set.

For a non-orthonormal basis set, Eq.(4.26) becomes

$$G(c_1, \ldots, c_N) = \sum_{ij} c_i^* c_j (H_{ij} - \epsilon S_{ij}) \tag{5.5}$$

and the minimum condition, Eq.(4.30), becomes

$$\sum_j (H_{ij} - \epsilon S_{ij}) c_j = 0 \tag{5.6}$$

or, in matrix form,

$$H\mathbf{c} = \epsilon S\mathbf{c}. \tag{5.7}$$

This is called a *generalized eigenvalue problem*.

The solution of a generalized eigenvalue problem is in practice equivalent to the solution of two simple eigenvalue problems. Let us first solve the auxiliary problem:

$$S\mathbf{d} = \sigma\mathbf{d} \tag{5.8}$$

completely analogous to the problem (4.32). We can thus find a unitary matrix D (obtained by putting eigenvectors as columns side by side), such that $D^{-1}SD$ is diagonal ($D^{-1} = D^\dagger$), and whose non-zero elements are the eigenvalues σ. We find an equation similar to Eq.(4.39):

$$\sum_i D^*_{ik} \sum_j S_{ij} D_{jn} = \sigma_n \delta_{kn}. \tag{5.9}$$

Note that all $\sigma_n > 0$: an overlap matrix is *positive definite*. In fact,

$$\sigma_n = \langle \tilde{b}_n | \tilde{b}_n \rangle, \qquad |\tilde{b}_n\rangle = \sum_j D_{jn} |b_j\rangle \tag{5.10}$$

and $|\tilde{b}\rangle$ is the rotated basis set in which S is diagonal. Note that a zero eigenvalue σ means that the corresponding $|\tilde{b}\rangle$ has zero norm, i.e. one of the b functions is a linear combination of the other functions. In that case, the matrix is called *singular* and some matrix operations (e.g. inversion) are not well defined.

Let us define now a second transformation matrix

$$A_{ij} \equiv \frac{D_{ij}}{\sqrt{\sigma_j}}. \tag{5.11}$$

We can write

$$\sum_i A^*_{ik} \sum_j S_{ij} A_{jn} = \delta_{kn} \tag{5.12}$$

(note that unlike D, the matrix A is not unitary) or, in matrix form, $A^\dagger S A = I$. Let us now define

$$\mathbf{c} = A\mathbf{v} \tag{5.13}$$

With this definition, Eq.(5.7) becomes

$$HA\,\mathbf{v} = \epsilon S A\,\mathbf{v} \tag{5.14}$$

We multiply to the left by $A^\dagger$:

$$A^\dagger H A\,\mathbf{v} = \epsilon A^\dagger S A\,\mathbf{v} = \epsilon\,\mathbf{v} \tag{5.15}$$

Thus, by solving the secular problem for operator $A^\dagger H A$, we find the desired eigenvalues for the energy. In order to obtain the eigenvectors in the starting base, it is sufficient, following Eq.(5.13), to apply operator A to each eigenvector.

5.1.1 Gaussian basis set

Atom-centered Gaussian functions are frequently used as basis functions, especially for atomic and molecular calculations. They are known as GTO: Gaussian-Type Orbitals. An important feature of Gaussian functions is that the product of two Gaussian functions, even if centered at different centers, can be written as a single Gaussian. With rather straightworward algebra, one demonstrates that

$$e^{-\alpha(\mathbf{r}-\mathbf{r}_1)^2}e^{-\beta(\mathbf{r}-\mathbf{r}_2)^2} = e^{-(\alpha+\beta)(\mathbf{r}-\mathbf{r}_0)^2}e^{-\frac{\alpha\beta}{\alpha+\beta}(\mathbf{r}_1-\mathbf{r}_2)^2}, \tag{5.16}$$

where the new center $\mathbf{r}_0$ lies on the line connecting the two centers $\mathbf{r}_1$ and $\mathbf{r}_2$:

$$\mathbf{r}_0 = \frac{\alpha\mathbf{r}_1 + \beta\mathbf{r}_2}{\alpha+\beta}. \tag{5.17}$$

Some useful integrals involving Gaussian functions:

$$\int_0^\infty e^{-\alpha x^2}dx = \frac{1}{2}\left(\frac{\pi}{\alpha}\right)^{1/2}, \tag{5.18}$$

$$\int_0^\infty xe^{-\alpha x^2}dx = \left[-\frac{e^{-\alpha x^2}}{2\alpha}\right]_0^\infty = \frac{1}{2\alpha}, \tag{5.19}$$

from which one derives

$$\int_0^\infty e^{-\alpha x^2}x^{2n}dx = (-1)^n\frac{\partial^n}{\partial\alpha^n}\int_0^\infty e^{-\alpha x^2}dx = \frac{(2n-1)!!\pi^{1/2}}{2^{n+1}\alpha^{n+1/2}} \tag{5.20}$$

$$\int_0^\infty e^{-\alpha x^2}x^{2n+1}dx = (-1)^n\frac{\partial^n}{\partial\alpha^n}\int_0^\infty xe^{-\alpha x^2}dx = \frac{n!}{2\alpha^{n+1}}. \tag{5.21}$$

All the needed matrix elements and overlaps can be computed starting from these formulae.

5.1.2 Other kinds of localized basis functions

Basis functions composed of Hydrogen-like wave functions (i.e. exponentials) are also used in Quantum Chemistry as alternatives to Gaussian functions. They are known as STO: Slater-Type Orbitals. It is also possible to use linear combinations of atomic orbitals (LCAO) in molecular and solid-state calculations. STO's and LCAO's do not have the nice analytical properties of Gaussian functions but may have other advantages, such as a better description of the behavior of the wave function close to the nucleus and at large distances.

Some useful integrals involving STOs (also used in Appendix D.1 and D.2) are listed here:

$$\int\frac{e^{-2Zr}}{r}d^3r = 4\pi\int_0^\infty re^{-2Zr}dr = 4\pi\left[e^{-2Zr}\left(-\frac{r}{2Z}-\frac{1}{4Z^2}\right)\right]_0^\infty = \frac{\pi}{Z^2} \tag{5.22}$$

$$\int\frac{e^{-2Z(r_1+r_2)}}{|\mathbf{r}_1-\mathbf{r}_2|}d^3r_1d^3r_2 = \frac{5\pi^2}{8Z^5}. \tag{5.23}$$

5.2 Code: hydrogen_gauss

Code `hydrogen_gauss.f90`[1] (or `hydrogen_gauss.c`[2]) solves the secular problem for the hydrogen atom using two different non-orthonormal basis sets:

1. a Gaussian, "S-wave" basis set:

$$b_i(\mathbf{r}) = e^{-\alpha_i r^2}; \tag{5.24}$$

2. a Gaussian "P-wave" basis set, existing in three different choices, corresponding to the different values m of the projection of the angular momentum L_z:

$$b_i(\mathbf{r}) = xe^{-\alpha_i r^2}, \quad b_i(\mathbf{r}) = ye^{-\alpha_i r^2}, \quad b_i(\mathbf{r}) = ze^{-\alpha_i r^2} \tag{5.25}$$

   (actually only the third choice corresponds to a well-defined value $m = 0$).

The Hamiltonian operator for this problem is obviously

$$H = -\frac{\hbar^2\nabla^2}{2m_e} - \frac{Zq_e^2}{r} \tag{5.26}$$

For the hydrogen atom, $Z = 1$.

Calculations for S- and P-wave Gaussians are completely independent. In fact, the two sets of basis functions are mutually orthogonal: $S_{ij} = 0$ if i is a S-wave, j is a P-wave Gaussian, as evident from the different parity of the two sets of functions. Moreover the matrix elements H_{ij} of the Hamiltonian are also zero between states of different angular momentum, for obvious symmetry reasons. The S and H matrices are thus *block matrices* and the eigenvalue problem can be solved separately for each block. The P-wave basis is clearly unfit to describe the ground state, being orthogonal to it by construction, and it is included mainly as an example.

The code reads from file a list of exponents, α_i, and proceeds to evaluate all matrix elements H_{ij} and S_{ij}. The calculation is based upon analytical results for integrals of Gaussian functions (Sec.5.1.1). In particular, for S-wave Gaussians one has

$$S_{ij} = \int e^{-(\alpha_i+\alpha_j)r^2} d^3r = \left(\frac{\pi}{\alpha_i+\alpha_j}\right)^{3/2} \tag{5.27}$$

while the kinetic and Coulomb terms in H_{ij} are respectively

$$H_{ij}^K = \int e^{-\alpha_i r^2} \left[-\frac{\hbar^2\nabla^2}{2m_e}\right] e^{-\alpha_j r^2} d^3r = \frac{\hbar^2}{2m_e} \frac{6\alpha_i\alpha_j}{\alpha_i+\alpha_j} \left(\frac{\pi}{\alpha_i+\alpha_j}\right)^{3/2} \tag{5.28}$$

$$H_{ij}^V = \int e^{-\alpha_i r^2} \left[-\frac{Zq_e^2}{r}\right] e^{-\alpha_j r^2} d^3r = -\frac{2\pi Zq_e^2}{\alpha_i+\alpha_j} \tag{5.29}$$

[1] http://www.fisica.uniud.it/%7Egiannozz/Didattica/MQ/Software/F90/hydrogen_gauss.f90
[2] http://www.fisica.uniud.it/%7Egiannozz/Didattica/MQ/Software/C/hydrogen_gauss.c

For the P-wave basis the procedure is analogous, using the corresponding (and more complex) analytical expressions for integrals.

The code then calls subroutine `diag` that solves the generalized secular problem (i.e. it applies the variational principle). Subroutine `diag` returns a vector `e` containing eigenvalues (in order of increasing energy) and a matrix `v` containing the eigenvectors, i.e. the expansion coefficients of wave functions.

Internally, `diag` performs the calculation described in the preceding section in two stages. The solution of the simple eigenvalue problem is performed by the subroutine `dsyev` we have already seen in Sec.4.4.

In principle, one could use a single LAPACK routine, `dsygv`, that solves the generalized secular problem, $H\psi = \epsilon S\psi$, with a single call. In practice, one has to be careful to avoid numerical instabilities related to the problem of *linear dependencies* among basis functions (see Eq.(5.10) and the following discussion). Inside routine `diag`, all eigenvectors of matrix S corresponding to very small eigenvectors, i.e. smaller than a pre-fixed threshold, are discarded, before proceeding with the second diagonalization. The number of linearly independent eigenvectors is reprinted in the output.

The reason for such procedure is that it is not uncommon to discover that some basis functions can almost exactly be written as sums of some other basis functions. This does not happen if the basis set is well chosen, but it can happen if the basis set functions are too numerous or not well chosen (e.g. with too close exponents). A wise choice of the α_j coefficients is needed in order to have a reach accuracy without numerical instabilities.

The code then proceeds and writes to files `s-coeff.out` (or `p-coeff.out`) the coefficients of the expansion of the wave function into Gaussians. The ground state wave function is written into file `s-wfc.out` (or `p-wfc.out`).

Notice the usage of `dgemm` calls to perform matrix-matrix multiplication. The header of `dgemm.f` contains a detailed documentation on how to call it. Its usage may seem awkward at a first sight (and also at a second one). This is a consequence in part of the Fortran way to store matrices, requiring the knowledge of the first, or "leading", dimension of matrices; in part, of the old-style Fortran way to pass variables, including vectors and arrays, to subroutines under the form of pointers. One may wonder why bother with `dgemm` and its obscure syntax and why not use the Fortran-90 syntax and the `MATMUL` intrinsic function instead: they are so much easier to use! The reason is *efficiency*: very efficient implementations of `dgemm` exist for modern architectures. On the other hand, modern language features that look great on paper, and sometimes *are* great also in practice, may turn out to be inefficient.

For the C version of the code, and how matrices are introduced and passed to Fortran routines, see Sec.0.1.4.

5.2.1 Laboratory

- Verify the accuracy of the energy eigenvalues, starting with one Gaussian, then 2, then 3. Try to find the best values for the coefficients for the $1s$ state (i.e. the values that yield the lowest energy).

- Compare with the the solutions obtained using code `hydrogen_radial`. Plot the $1s$ numerical solution (calculated with high accuracy) and the "best" $1s$ solution for 1, 2, 3, Gaussians (you will need to multiply the latter by a factor $\sqrt{4\pi}$: why? where does it come from?). What do you observe? where is the most significant error concentrated?

- Compare with the results for the following optimized basis sets (a.u.):

  - three Gaussians: $\alpha_1 = 0.109818$, $\alpha_2 = 0.405771$, $\alpha_3 = 2.22776$ (known as "STO-3G" in Quantum-Chemistry jargon)
  - four Gaussians: $\alpha_1 = 0.121949$, $\alpha_2 = 0.444529$, $\alpha_3 = 1.962079$, $\alpha_4 = 13.00773$

- Observe and discuss the ground state obtained using the P-wave basis set

- Observe the effects related to the number of basis functions, and to the choice of the parameters α. Try for instance to choose the characteristic Gaussian widths, $\lambda = 1/\sqrt{\alpha}$, as uniformly distributed between suitably chosen λ_{min} and λ_{max}.

- For $Z > 1$, how would you re-scale the coefficients of the optimized Gaussians above?

Chapter 6

Self-consistent field

Let us move now to many-particle problems. A way to solve a system of many electrons is to consider each electron under the effective field generated by all other electrons. The many-body problem is thus reduced to the *self-consistent* solution of many single-electron Schrödinger equations. This idea is formalized in a rigorous way in the Hartree-Fock and density-functional theories. In the following we consider the simple case of the ground state of the He atom, where the Hartree-Fock method leads to an equation in which the self-consistent field is a function of the charge density only.

6.1 The Hartree-Fock method

The idea of the Hartree-Fock method is to approximate the wave function for a $N-$electron system with a suitably antisymmetrized product of N single-electron functions (*orbitals*). The best set of orbitals is then found by applying the variational principle, that is: by minimizing the expectation value of the energy $E = \langle\psi|H|\psi\rangle$ for state $|\psi\rangle$.

Let us consider the case of a $N-$electron atom. The Hamiltonian is

$$H = -\sum_i \frac{\hbar^2}{2m_e}\nabla_i^2 - \sum_i \frac{Zq_e^2}{r_i} + \sum_{\langle ij\rangle} \frac{q_e^2}{r_{ij}} \tag{6.1}$$

where Z is the charge of the nucleus and the sum is over *pairs* of electrons i and j, i.e. each pair appears only once. Alternatively:

$$\sum_{\langle ij\rangle} = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \tag{6.2}$$

It is convenient to introduce one-electron f_i and two-electrons g_{ij} operators:

$$f_i \equiv -\frac{\hbar^2}{2m_e}\nabla_i^2 - \frac{Zq_e^2}{r_i} \quad , \quad g_{ij} \equiv \frac{q_e^2}{r_{ij}} \tag{6.3}$$

With such notation, the Hamiltonian is written as

$$H = \sum_i f_i + \sum_{\langle ij\rangle} g_{ij}. \tag{6.4}$$

6.1.1 Slater determinants

A $N-$electron wave function can be built from a set of orbitals ϕ_i, $i = 1, \ldots, N$ as a *Slater determinant*:

$$\psi(1, \ldots, N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(1) & \ldots & \phi_1(N) \\ . & . & . \\ . & . & . \\ \phi_N(1) & \ldots & \phi_N(N) \end{vmatrix} \tag{6.5}$$

Here the notation $(N) \equiv (\mathbf{r}_N, \sigma_N)$ indicates position and spin variables for the $N-$th electron. A Slater determinant has by construction the required antisymmetry properties for a many-electron wave function: in fact, the exchange of two particles is equivalent to the exchange of two columns, which produces, due to a known property of determinants, a change of sign. Note that if two rows are equal, the determinant is zero: all ϕ_i's must be different. This demonstrates Pauli's exclusion principle: *two (or more) identical fermions cannot occupy the same state*.

Note that the single-electron orbitals ϕ_i are assumed to be orthonormal:

$$\int \phi_i^*(1)\phi_j(1) dv_1 = \delta_{ij} \tag{6.6}$$

where the "integration" on dv_1 means "integration on coordinates, sum over spin components".

Since a determinant for N electrons has $N!$ terms, we need a way to write matrix elements between determinants on a finite surface of paper. The following property, valid for any (symmetric) operator F and determinantal functions ψ and ψ', is very useful:

$$\begin{aligned} \langle\psi|F|\psi'\rangle &= \frac{1}{N!} \int \begin{vmatrix} \phi_1^*(1) & . & \phi_1^*(N) \\ . & . & . \\ \phi_N^*(1) & . & \phi_N^*(N) \end{vmatrix} F \begin{vmatrix} \phi'_1(1) & . & \phi'_1(N) \\ . & . & . \\ \phi'_N(1) & . & \phi'_N(N) \end{vmatrix} dv_1 \ldots dv_N \\ &= \int \phi_1^*(1) \ldots \phi_N^*(N) F \begin{vmatrix} \phi'_1(1) & . & \phi'_1(N) \\ . & . & . \\ \phi'_N(1) & . & \phi'_N(N) \end{vmatrix} dv_1 \ldots dv_N. \end{aligned} \tag{6.7}$$

To demonstrate it, one may observe that by expanding the first determinant, one obtains $N!$ terms that, once integrated, are identical.

6.1.2 Hartree-Fock equations

From the above property, and using the orthonormality of the orbitals, it is simple to write the expectation value of the Hamiltonian $E = \langle\psi|H|\psi\rangle$ in terms of matrix elements of one- and two-electron operators. The only non-zero terms are in fact

$$\langle\psi| \sum_i f_i |\psi\rangle = \sum_i \int \phi_i^*(1) f_1 \phi_i(1)\, dv_1 \tag{6.8}$$

and

$$\langle\psi|\sum_{\langle ij\rangle} g_{ij}|\psi\rangle = \sum_{\langle ij\rangle}\int \phi_i^*(1)\phi_j^*(2)g_{12}\left[\phi_i(1)\phi_j(2)-\phi_j(1)\phi_i(2)\right]\,dv_1dv_2. \quad (6.9)$$

Notice the sign and the index permutation in the second term in square brackets: both come from the antisymmetry. The integrals implicitly include summation over spin components. If we assume that g_{12} depends only upon coordinates, as in Coulomb interaction, and not upon spins, the second term is zero if i and j states have different spins. For this reason, it is convenient to make spin variables explicit. Eq.(6.9) can then be rewritten as

$$\langle\psi|\sum_{\langle ij\rangle} g_{ij}|\psi\rangle = \sum_{\langle ij\rangle}\int \phi_i^*(1)\phi_j^*(2)g_{12}\left[\phi_i(1)\phi_j(2)-\delta(\sigma_i,\sigma_j)\phi_j(1)\phi_i(2)\right]\,dv_1dv_2 \quad (6.10)$$

where σ_i is the spin of electron i, and:

$$\begin{aligned}\delta(\sigma_i,\sigma_j) &= 0 \text{ if } \sigma_i\neq\sigma_j\\ &= 1 \text{ if } \sigma_i=\sigma_j\end{aligned}$$

In summary:

$$\begin{aligned}\langle\psi|H|\psi\rangle &= \sum_i\int \phi_i^*(1)f_1\phi_i(1)\,dv_1 \quad (6.11)\\ &+ \sum_{\langle ij\rangle}\int \phi_i^*(1)\phi_j^*(2)g_{12}\left[\phi_i(1)\phi_j(2)-\delta(\sigma_i,\sigma_j)\phi_j(1)\phi_i(2)\right]\,dv_1dv_2\end{aligned}$$

Let us now apply the variational principle to Eq.(6.11). In principle we must impose normalization constraints such that all pairs ϕ_i, ϕ_j with same spin stay orthogonal, i.e., a (triangular) matrix ϵ_{ij} of Lagrange multipliers would be needed. It can be shown however (details e.g. on Slater's book, *Quantum theory of matter*) that it is always possible to find a transformation to a solution in which the matrix of Lagrange multipliers is diagonal. We assume that we are dealing with such a case and write the stationarity conditions as

$$\delta\left(\langle\psi|H|\psi\rangle - \sum_i \epsilon_i I_i\right) = 0 \quad (6.12)$$

where

$$I_i = \int \phi_i^*(1)\phi_i(1)\,dv_1 \quad (6.13)$$

are the constraints, ϵ_i are the Lagrange multipliers that enforce such constraints.

Let us vary only the orbital function ϕ_k. We find

$$\delta I_k = \int \delta\phi_k^*(1)\phi_k(1)\,dv_1 + \text{c.c.} \quad (6.14)$$

(the variations of all other normalization integrals will be zero) and, using the hermiticity of H as in Sec.4.1.1,

$$\begin{aligned}\delta\langle\psi|H|\psi\rangle &= \int \delta\phi_k^*(1)f_1\phi_k(1)\,dv_1 + \text{c.c.} \quad (6.15)\\ &+ \sum_j\int \delta\phi_k^*(1)\phi_j^*(2)g_{12}\left[\phi_k(1)\phi_j(2)-\phi_k(2)\phi_j(1)\delta(\sigma_k,\sigma_j)\right]\,dv_1dv_2 + \text{c.c.}\end{aligned}$$

Thus the variational principle takes the form

$$\int \delta\phi_k^*(1)\,\{f_1\phi_k(1) \quad + \quad \int \phi_j^*(2) g_{12} \sum_j [\phi_k(1)\phi_j(2) \tag{6.16}$$
$$- \quad \phi_k(2)\phi_j(1)\delta(\sigma_k,\sigma_j)]\; dv_2 - \epsilon_k\phi_k(1)\}\; dv_1 + \text{c.c.} = 0$$

i.e. the term between curly brackets (and its complex conjugate) must vanish so that the above equation is satisfied for any variation. This leads to the set of integro-differential *Hartree-Fock equations*:

$$f_1\phi_k(1) + \sum_j \int \phi_j^*(2) g_{12}\,[\phi_k(1)\phi_j(2) - \delta(\sigma_k,\sigma_j)\phi_j(1)\phi_k(2)]\; dv_2 = \epsilon_k\phi_k(1) \tag{6.17}$$

or, in more explicit form,

$$-\frac{\hbar^2}{2m_e}\nabla_1^2\phi_k(1) - \frac{Zq_e^2}{r_1}\phi_k(1) \quad + \quad \sum_j \int \phi_j^*(2)\frac{q_e^2}{r_{12}}\,[\phi_j(2)\phi_k(1) \tag{6.18}$$
$$- \quad \delta(\sigma_k,\sigma_j)\phi_k(2)\phi_j(1)]\; dv_2 = \epsilon_k\phi_k(1)$$

Eq.(6.18) has normally an infinite number of solutions, of which only the lowest-energy N will be occupied by electrons, the rest playing the role of excited states. The sum over index j runs only on occupied states.

6.1.3 Hartree and exchange potentials

Let us analyze the form of the Hartree-Fock equations. We re-write them as

$$-\frac{\hbar^2}{2m_e}\nabla_1^2\phi_k(1) - \frac{Zq_e^2}{r_1}\phi_k(1) + V_H(1)\phi_k(1) + (\hat{V}_x\phi_k)(1) = \epsilon_k\phi_k(1), \tag{6.19}$$

where we have introduced the *Hartree potential* V_H and the *exchange potential* $\hat{V}_x$. The Hartree potential is the average electrostatic potential felt by an electron in the field generated by all other electrons:

$$V_H(1) = \sum_j \int \phi_j^*(2)\frac{q_e^2}{r_{12}}\phi_j(2)dv_2 \equiv \int \rho(2)\frac{q_e^2}{r_{12}}dv_2, \tag{6.20}$$

where ρ is the charge density:

$$\rho(2) = \sum_j \phi_j^*(2)\phi_j(2). \tag{6.21}$$

It is straightforward to verify that $\rho(1)$ is equal to the probability to find an electron in (1):

$$\rho(1) = N\int |\psi(1,\ldots,N)|^2 dv_2 \ldots dv_N. \tag{6.22}$$

Note that both the Hartree potential and the charge density do not actually depend upon the spin.

The exchange potential reflects the effect of antisymmetry between electrons with the same spin:

$$(\hat{V}_x\phi_k)(1) = -\sum_j \delta(\sigma_k,\sigma_j)\int \phi_j(1)\phi_j^*(2)\frac{q_e^2}{r_{12}}\phi_k(2)\,dv_2. \tag{6.23}$$

While the Hartree potential is local (just multiplies the orbitals), the exchange potential is *non local*, acting on orbitals as

$$(\hat{V}_x\phi_k)(1) \equiv \int V_x(1,2)\phi_k(2)dv_2. \tag{6.24}$$

The careful observer will notice an alarming feature: the Hartree potential contains a "self-interaction" of an electron with itself. Such interaction is obviously spurious and unphysical. The same term with opposit sign is however present in the exchange potential, so it cancels out in the equations.

6.1.4 Hartree-Fock and correlation energy

The energy of the system, Eq. 6.11, can be expressed via the Hartree and exchange potential as

$$\begin{aligned} E &= \sum_i \int \phi_i^*(1) f_1 \phi_i(1)\,dv_1 + \frac{1}{2}\int \rho(1)V_H(1)dv_1 \\ &+ \frac{1}{2}\sum_k \int \phi_k(1)\hat{V}_x(1,2)\phi_k(2)dv_1dv_2. \end{aligned} \tag{6.25}$$

The factors 1/2 reflect the fact that the sums run over *pairs* of orbitals. The first term in the energy has a one-electron character and containg the kinetic and potential energy. The second term is called Hartree energy and is of electrostatic character; the last term is called exchange energy.

Alternatively, one may use the sum of eigenvalues of Eq.(6.18): by multiplying the two sides of Eq.(6.18) by $\phi_k^*(1)$ and integrating over dv_1, one finds

$$E = \sum_k \epsilon_k - \sum_{<jk>}\int \phi_k^*(1)\phi_j^*(2)g_{12}\left[\phi_k(1)\phi_j(2) - \delta(\sigma_j,\sigma_k)\phi_j(1)\phi_k(2)\right]dv_1dv_2, \tag{6.26}$$

or, in terms of the Hartree and exchange potential:

$$E = \sum_k \epsilon_k - \frac{1}{2}\int \rho(1)V_H(1)dv_1 - \frac{1}{2}\sum_k \int \phi_k(1)\hat{V}_x(1,2)\phi_k(2)dv_1dv_2. \tag{6.27}$$

Notice the sign of the Hartree and exchange energy terms: they compensate for the double counting of those terms that is present in the eigenvalue sum.

The Hartree-Fock energy is not the exact energy: it would be in a world where the exact wave function has the form of a Slater determinant. This is in general not true. The energy difference between the exact and Hartree-Fock solution is known as *correlation energy*.[1] This name reflects the fact

[1] Feynman called it *stupidity energy*, because the only physical quantity that it measures is our inability to find the exact solution!

that the Hartree-Fock approximation misses part of the "electron correlation": the effects of an electron on all others. This is present in Hartree-Fock via the exchange and electrostatic interactions, but more subtle effects are not accounted for, because they require a more general form of the wave function. For instance, the probability $P(r_1, r_2)$ to find an electron at distance r_1 and one at distance r_2 from the origin is not simply equal to $p(r_1)p(r_2)$, because electrons try to "avoid" each other. The correlation energy in the case of He atom is about 0.084 Ry: a small quantity relative to the energy ($\sim 1.5\%$), but not negligible.

6.1.5 Helium atom

Let us consider the simplest case of many-electron atom: He (Z=2). The ground state of He can be build with two orbitals, ϕ_1 and ϕ_2, having the same coordinate part $\phi(\mathbf{r})$, and opposite spins: $\phi_1 = \phi(\mathbf{r})v_+(\sigma)$, $\phi_2 = \phi(\mathbf{r})v_-(\sigma)$. In the exchange potential term, the contribution from different orbitals ϕ_1 and ϕ_2 vanishes, while the contribution from the same orbital cancels out with the same term in the Hartree potential. The Hartree-Fock equations, Eqs.(6.18), simplify to a single equation for $\phi(\mathbf{r}_1)$:

$$-\frac{\hbar^2}{2m_e}\nabla_1^2\phi(\mathbf{r}_1) - \frac{Zq_e^2}{r_1}\phi(\mathbf{r}_1) + \int \phi^*(\mathbf{r}_2)\frac{q_e^2}{r_{12}}\phi(\mathbf{r}_2)d\mathbf{r}_2\phi(\mathbf{r}_1) = \epsilon_1\phi(\mathbf{r}_1). \qquad (6.28)$$

The equation for the second orbital is obviously the same, and $\epsilon_2 = \epsilon_1$. One can recast such equation in a more transparent form:

$$\left(-\frac{\hbar^2}{2m_e}\nabla_1^2 + V_{scf}(\mathbf{r}_1)\right)\phi(\mathbf{r}_1) = \epsilon_1\phi(\mathbf{r}_1) \qquad (6.29)$$

where V_{scf} is the effective, *self-consistent*, potential felt by an electron:

$$V_{scf}(\mathbf{r}_1) = -\frac{Zq_e^2}{r_1} + \int \tilde{\rho}(\mathbf{r}_2)\frac{q_e^2}{r_{12}}d\mathbf{r}_2 \quad , \quad \tilde{\rho}(\mathbf{r}_2) = |\phi(\mathbf{r}_2)|^2. \qquad (6.30)$$

Note that $\tilde{\rho}$ is half of the charge density.

For closed-shell atoms, a further big simplification can be achieved: V_{scf} is a *central* field, i.e. it depends only on the distance r_1 between the electron and the nucleus. Even in open-shell atoms, this can be imposed as an approximation, by spherically averaging ρ. The simplification is considerable, since we know *a priori* that the orbitals will be factorized as in Eq.(2.9). The angular part is given by spherical harmonics, labelled with quantum numbers ℓ and m, while the radial part is characterized by quantum numbers n and ℓ. Of course the accidental degeneracy for different ℓ is no longer present. It turns out that even in open-shell atoms, this is an excellent approximation.

6.2 Code: helium_hf_radial

Code `helium_hf_radial.f90`[2] (or `helium_hf_radial.c`[3]) solves Hartree-Fock equations for the ground state of He atom. `helium_hf_radial` is based on code `hydrogen_radial` and uses the same integration algorithm based on Numerov's method. The new part is the implementation of the method of self-consistent field for finding the orbitals.

The calculation consists in solving the radial part of the Hartree-Fock equation (6.29) for the effective potential of Eq.(6.30), assumed to be radial:

$$V_{scf}(r) = -\frac{Zq_e^2}{r} + \tilde{V}_H(r) \quad , \quad \tilde{V}_H(r) \int \tilde{\rho}(r') \frac{q_e^2}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}'. \tag{6.31}$$

Note that $\tilde{V}_H$ is half of the Hartree potential. $V_{scf}(r)$ depends in turn upon the solution via $\tilde{\rho}(r)$. We start from an initial estimate for $\tilde{V}_H(r)$ (simply, $\tilde{V}_H(r) = 0$), calculated in routine `init_pot`. With the ground state $R(r)$ obtained from such potential, we calculate in routine `rho_of_r` the charge density (of the other electron) $\tilde{\rho}(r) = |R(r)|^2/4\pi$ (do not forget the angular part!). Routine `v_of_rho` re-calculates the new potential $\tilde{V}_H^{out}(r)$ by integration, using the Gauss theorem:

$$\tilde{V}_H^{out}(r) = V_0 + q_e^2 \int_{r_{max}}^{r} \frac{Q(s)}{s^2} ds, \quad Q(s) = \int_{r<s} \rho(r) 4\pi r^2 dr \tag{6.32}$$

where $Q(s)$ is the charge contained in the sphere of radius s; r_{max} is the outermost grid point, such that the potential has already assumed the asymptotic value $V_0 = q_e^2/r_{max}$, valid at large r.

The new potential is then re-introduced in the calculation not directly but as a linear combination of the old and the new potential. This is the simplest technique to ensure that the self-consistent procedure converges. It is not needed in this case, but in most cases it is: there is no guarantee that re-inserting the new potential in input will lead to convergence. We can write

$$\tilde{V}_H^{in,new}(r) = \beta \tilde{V}_H^{out}(r) + (1 - \beta) \tilde{V}_H^{in}(r), \tag{6.33}$$

where $0 < \beta \leq 1$. The procedure is iterated (repeated) until convergence is achieved. The latter is verified on the "variational correction" descrived below, but other choices are possible (e.g. the norm – the square root of the integral of the square – of $\tilde{V}_H^{out}(r) - \tilde{V}_H^{in}(r)$).

In output the code prints the eigenvalue ϵ_1 of Eq.6.29, plus various terms of the energy, with rather obvious meaning except the term *Variational correction*:

$$\delta E = \int (\tilde{V}_H^{in}(r) - \tilde{V}_H^{out}(r)) \rho(r) d^3r. \tag{6.34}$$

This term accounst for the fact that eigenvalues are calculated using the *input* potential, while other energy terms are calculated using the *output* potential. By adding this correction, the energy obtained from the eigenvalue sum as in

[2] http://www.fisica.uniud.it/%7Egiannozz/Didattica/MQ/Software/F90/helium_hf_radial.f90
[3] http://www.fisica.uniud.it/%7Egiannozz/Didattica/MQ/Software/C/helium_hf_radial.c

Eq.(6.27) is the same as the one obtained using Eq.(6.25), also far from self-consistency. The two values of the energy are printed side by side.

Also noticeable is the "virial check": for a Coulomb potential, the *virial theorem* states that $\langle T \rangle = -\langle V \rangle/2$, where the two terms are respectively the average values of the kinetic and the potential energy. It can be demonstrated that the Hartree-Fock solution obeys the virial theorem.

6.2.1 Laboratory

- Observe the behavior of self-consistency, verify that the energy (but not any single term of the energy!) decreases monotonically. Also note that the self-consistency error (the "variational correction" term) decreases exponentially with the number of iterations.

- Compare the energy obtained with this and with other methods: perturbation theory with hydrogen-like wave functions ($E = -5.5$ Ry, Sect. D.1), variational theory with effective Z ($E = -5.695$ Ry, Sect. D.2), best numerical Hartree(-Fock) result ($E = -5.72336$ Ry, as found in the literature), experimental result ($E = -5.8074$ Ry).

- Make a plot of orbitals (file `wfc.out`) for different n and ℓ. Note that the orbitals and the corresponding eigenvalues become more and more hydrogen-like for higher n. Can you explain why?

- If you do not know how to answer the previous question: make a plot of V_{scf} (file `pot.out`) and compare its behavior with that of the $-Zq_e^2/r$ term. What do you notice?

- Plot the $1s$ orbital together with those calculated by `hydrogen_radial` for Hydrogen ($Z = 1$), He^+ ($Z = 2$), and for a $Z = 1.6875$. See Sect. D.2 if you cannot make sense of the results.

Chapter 7

Molecules

In this chapter we consider the simplest case of a molecular system. From a formal point of view, a molecule is a system of interacting electrons and nuclei. We first introduce the *adiabatic*, or *Born-Oppenheimer*, approximation, enabling to split the starting problem into the solution of the problem of electrons under the field of nuclei, and the solution of the problem of nuclei under an effective potential - the *Potential Energy Surface*, or PES, depending upon the electronic state and energy.

The electronic problem can be solved, as a function of nuclear positions, using the methods introduced in the previous sections. Here the solution for the simplest molecule, H_2, is obtained using the Hartree-Fock method and a Gaussian basis set. The relation with the Molecular Orbital theory, as well as the limitation of the Hartree-Fock approximation, will become clear.

7.1 Born-Oppenheimer approximation

Let us consider a system of interacting nuclei and electrons. In general, the Hamiltonian of the system will depend upon all nuclear coordinates, $\mathbf{R}_\mu$, and all electronic coordinates, $\mathbf{r}_i$. For a system of n electrons under the field of N nuclei with charge Z_μ, in principle one has to solve the following Schrödinger equation:

$$(T_I + V_{II} + V_{eI} + T_e + V_{ee})\,\Psi(\mathbf{R}_\mu, \mathbf{r}_i) = E\Psi(\mathbf{R}_\mu, \mathbf{r}_i) \tag{7.1}$$

where T_I is the kinetic energy of nuclei, V_{II} is the Coulomb repulsion between nuclei, V_{eI} is the Coulomb attraction between nuclei and electrons, T_e is the kinetic energy of electrons, V_{ee} is the Coulomb repulsion between electrons:

$$T_I = -\sum_{\mu=1,N} \frac{\hbar^2}{2M_\mu}\nabla^2_\mu, \quad T_e = -\sum_{i=1,n} \frac{\hbar^2}{2m}\nabla^2_i, \quad V_{II} = \frac{q_e^2}{2}\sum_{\mu\neq\nu} \frac{Z_\mu Z_\nu}{|\mathbf{R}_\mu - \mathbf{R}_\nu|},$$
$$V_{ee} = \frac{q_e^2}{2}\sum_{i\neq j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}, \quad V_{eI} = -q_e^2 \sum_{\mu=1,N}\sum_{i=1,n} \frac{Z_\mu}{|\mathbf{R}_\mu - \mathbf{r}_i|}. \tag{7.2}$$

This looks like an impressive problem. It is however possible to exploit the mass difference between electrons and nuclei to separate the global problem into an

electronic problem for fixed nuclei and a nuclear problem under an effective potential generated by electrons. Such separation is known as *adiabatic* or *Born-Oppenheimer* approximation. The crucial point is that the electronic motion is much faster than the nuclear motion: while forces on nuclei and electrons have the same order of magnitude, an electron is at least ~ 2000 times lighter than any nucleus. We can thus assume that at any time the electrons "follow" the nuclear motion, while the nuclei at any time "feel" an effective potential generated by electrons. Formally, we assume a wave function of the form

$$\Psi(\mathbf{R}_\mu, \mathbf{r}_i) = \Phi(\mathbf{R}_\mu)\psi^{(l)}_{\mathbf{R}}(\mathbf{r}_i) \tag{7.3}$$

where the electronic wave function $\psi^{(l)}_{\mathbf{R}}(\mathbf{r}_i)$ solves the following Schrödinger equation:

$$\left(T_e + V_{ee} + V_{eI}\right)\psi^{(l)}_{\mathbf{R}}(\mathbf{r}_i) = E^{(l)}_{\mathbf{R}}\psi^{(l)}_{\mathbf{R}}(\mathbf{r}_i). \tag{7.4}$$

The index $\mathbf{R}$ is a reminder that both the wave function and the energy depend upon the nuclear coordinates, via V_{eI}; the index l classifies electronic states. We now insert the wave function, Eq.(7.3), into Eq.(7.1) and notice that T_e does not act on nuclear variables. We will get the following equation:

$$\left(T_I + V_{II} + E^{(l)}_{\mathbf{R}}\right)\Phi(\mathbf{R}_\mu)\psi^{(l)}_{\mathbf{R}}(\mathbf{r}_i) = E\Phi(\mathbf{R}_\mu)\psi^{(l)}_{\mathbf{R}}(\mathbf{r}_i). \tag{7.5}$$

If we now neglect the dependency upon $\mathbf{R}$ of the electronic wave functions in the kinetic term:

$$T_I\left(\Phi(\mathbf{R}_\mu)\psi^{(l)}_{\mathbf{R}}(\mathbf{r}_i)\right) \simeq \psi^{(l)}_{\mathbf{R}}(\mathbf{r}_i)\left(T_I\Phi(\mathbf{R}_\mu)\right). \tag{7.6}$$

we obtain a Schrödinger equation for nuclear coordinates only:

$$\left(T_I + V_{II} + E^{(l)}_{\mathbf{R}}\right)\Phi(\mathbf{R}_\mu) = E\Phi(\mathbf{R}_\mu), \tag{7.7}$$

where electrons have "disappeared" into the eigenvalue $E^{(l)}_{\mathbf{R}}$. The term $V_{II}+E^{(l)}_{\mathbf{R}}$ plays the role of effective interaction potential between nuclei. Of course such potential, as well as eigenfunctions and eigenvalues of the nuclear problem, depends upon the particular electronic state.

The Born-Oppenheimer approximation is very well verified, except the special cases of *non-adiabatic* phenomena (that are very important, though). The main neglected term in Eq.7.6 has the form

$$\sum_\mu \frac{\hbar^2}{M_\mu}\left(\nabla_\mu\Phi(\mathbf{R}_\mu)\right)\left(\nabla_\mu\psi^{(l)}_{\mathbf{R}}(\mathbf{r}_i)\right) \tag{7.8}$$

and may if needed be added as a perturbation.

7.2 Potential Energy Surface

The Born-Oppenheimer approximation allows us to separately solve a Schrödinger equation for electrons, Eq.(7.4), as a function of atomic positions, and

a problem for nuclei only, Eq.(7.7). The latter is in fact a Schrödinger equation in which the nuclei interact via an *effective interatomic potential*, $V(\mathbf{R}_\mu) \equiv V_{II}+E^{(l)}$, a function of the atomic positions $\mathbf{R}_\mu$ and of the electronic state. The interatomic potential $V(\mathbf{R}_\mu)$ is also known as *potential energy surface* ("potential" and "potential energy" are in this context synonyms), or PES. It is clear that the nuclear motion is completely determined by the PES (assuming that the electron states does not change with time) since forces acting on nuclei are nothing but the gradient of the PES:

$$\mathbf{F}_\mu = -\nabla_\mu V(\mathbf{R}_\mu), \tag{7.9}$$

while equilibrium positions for nuclei, labelled with $\mathbf{R}_\mu^{(0)}$, are characterized by zero gradient of the PES (and thus of any force on nuclei):

$$\mathbf{F}_\mu = -\nabla_\mu V(\mathbf{R}_\mu^{(0)}) = 0. \tag{7.10}$$

In general, there can be many equilibrium points, either stable (a minimum: any displacement from the equilibrium point produces forces opposed to the displacement, i.e. the second derivative is positive everywhere) or unstable (a maximum or a saddle point: for at least some directions of displacement from equilibrium, there are forces in the direction of the displacement, i.e. there are negative second derivatives). Among the various minima, there will be a *global minimum*, the lowest-energy one, corresponding to the ground state of the nuclear system, for a given electronic state. If the electronic state is also the ground state, this will be the ground state of the atomic system. All other minima are *local minima*, that is, *metastable* states that the nuclear system can leave by overcoming a *potential barrier*.

7.3 Diatomic molecules

Let us consider now the simple case of diatomic molecules, and in particular, the molecule of H_2. There are 6 nuclear coordinates, 3 for the center of mass and 3 relative to it, but just one, the distance R between the nuclei, determines the effective interatomic potential $V(R)$ (the PES is in fact invariant, both translationally and rotationally). Given a distance R, we may solve Eq.(7.4) for electrons, find the l-th electronic energy level $E^{(l)}(R)$ and the corresponding interatomic potential $V(R) = E_{II}(R)+E^{(l)}(R)$. Note that the nuclear repulsion energy $E_{II}(R)$ is simply given by

$$E_{II}(R) = \frac{q_e^2 Z_1 Z_2}{R} \tag{7.11}$$

where Z_1 and Z_2 are nuclear charges for the two nuclei.

Let us consider the electronic ground state only for H_2 molecule. At small R, repulsion between nuclei is dominant, $V(R)$ becomes positive, diverging like q_e^2/R for $R \to 0$. At large R, the ground state becomes that of two neutral H atoms, thus $V(R) \simeq 2$Ry. At intermediate R, the curve has a minimum at about $R_0 = 0.74$Å, with $V(R_0) \simeq V(\infty) - 4.5$eV. Such value – the difference

between the potential energy of atoms at large distances and at the minimum, is known as *cohesive* or *binding energy*. The form of the interatomic potential is reminiscent of that of model potentials like Morse (in fact such potentials were proposed to model binding).

What is the electronic ground state for $R \sim R_0$? We can get an idea by using the method of *molecular orbitals*: an approximate solution in which single-electron states are formed as linear combinations of atomic states centered about the two nuclei. Combinations with the same phase are called *ligand*, as they tend to accumulate charge in the region of space between the two nuclei. Combinations with opposite phase are called *antiligand*, since they tend to deplete charge between the nuclei. Starting from molecular orbitals, one can build the wave function as a Slater determinant. Two ligand states of opposite spin, built as superpositions of $1s$ states centered on the two nuclei (σ orbitals), yield a good approximation to the ground state. By construction, the total spin of the ground state is zero.

Molecular orbitals theory can explain qualitatively the characteristics of the ground (and also excited) states, for the homonuclear dimer series (i.e. formed by two equal nuclei). It is however no more than semi-quantitative; for better results, one has to resort to the variational method, typically in the framework of Hartree-Fock or similar approximations. Orbitals are expanded on a basis set of functions, often atom-centered Gaussians or atomic orbitals, and the Hartree-Fock or similar equations are solved on this basis set.

7.4 Rothaan-Hartree-Fock equations

In general, one speaks of *Restricted Hartree-Fock* (RHF) for the frequent case in which all orbitals are present in pairs, formed by a same function of $\mathbf{r}$, multiplied by spin functions of opposite spin. In the following, this will always be the case.

The RHF equations can be recast into the following form:

$$\mathcal{F}\phi_k = \epsilon_k \phi_k, \quad k = 1, \ldots, N/2 \tag{7.12}$$

The index k labels the coordinate parts of the orbitals; for each k there is a spin-up and a spin-down orbital. $\mathcal{F}$ is called the *Fock operator*. This is of course a non-local operator which depends upon all orbitals ϕ_k. Explicitly, from Eqs.(6.18):

$$\begin{aligned} \mathcal{F}\phi_k(\mathbf{r}_1) \equiv & \frac{\hbar^2}{2m_e}\nabla_1^2\phi_k(\mathbf{r}_1) - \frac{Zq_e^2}{r_1}\phi_k(\mathbf{r}_1) + 2\sum_j \int \phi_j^*(\mathbf{r}_2)\frac{q_e^2}{r_{12}}\phi_j(\mathbf{r}_2)d\mathbf{r}_2\phi_k(\mathbf{r}_1) \\ & - \sum_j \int \phi_j(\mathbf{r}_1)\frac{q_e^2}{r_{12}}\phi_j^*(\mathbf{r}_2)d\mathbf{r}_2\phi_k(\mathbf{r}_2). \end{aligned} \tag{7.13}$$

Let us look now for a solution under the form of an expansion on a basis of functions: $\phi_k(\mathbf{r}) = \sum_1^M c_i^{(k)} b_i(\mathbf{r})$. We find the *Rothaan-Hartree-Fock* equations:

$$F\mathbf{c}^{(k)} = \epsilon_k S\mathbf{c}^{(k)} \tag{7.14}$$

where $\mathbf{c}^{(k)} = (c_1^{(k)}, c_2^{(k)}, \ldots, c_M^{(k)})$ is the vector of the expansion coefficients, S is the superposition matrix, F is the matrix of the Fock operator on the basis set functions:

$$F_{ij} = \langle b_i|\mathcal{F}|b_j\rangle, \qquad S_{ij} = \langle b_i|b_j\rangle. \tag{7.15}$$

that after some algebra can be written as

$$F_{ij} = f_{ij} + \sum_l \sum_m \left(2\sum_{k=1}^{N/2} c_l^{(k)*} c_m^{(k)} \right) \left(g_{iljm} - \frac{1}{2} g_{ijlm} \right), \tag{7.16}$$

where, with the notations introduced in this chapter:

$$f_{ij} = \int b_i^*(\mathbf{r}_1) f_1 b_j(\mathbf{r}_1) d^3r_1, \tag{7.17}$$

$$g_{iljm} = \int b_i^*(\mathbf{r}_1) b_j(\mathbf{r}_1) g_{12} b_l^*(\mathbf{r}_2) b_m(\mathbf{r}_2) d^3r_1 d^3r_2. \tag{7.18}$$

The sum over states between parentheses in Eq.(7.16) is called *density matrix*. The two terms in the second parentheses come respectively from the Hartree and the exchange potentials.

The problem of Eq.(7.14) is more complex than a normal secular problem solvable by diagonalization, since the Fock matrix, Eq.(7.16), depends upon its own eigenvectors. It is however possible to reconduct the solution to a self-consistent procedure, in which at each step a fixed matrix is diagonalized (or, for a non-orthonormal basis, a generalized diagonalization is performed at each step).

For matrix elements between Gaussians centered around the same centers, one can easily find analytical results, using the basic ingredients already used for code `hydrogen_gauss`. We need an expression for the g_{iljm} matrix elements introduced in Eq.(7.18). Using the properties of products of Gaussians, Eq.(5.16), these can be written in terms of the integral

$$I = \int e^{-\alpha r_1^2} e^{-\beta r_2^2} \frac{1}{r_{12}} d^3r_1 d^3r_2. \tag{7.19}$$

Let us look for a variable change that makes $(\mathbf{r}_1 - \mathbf{r}_2)^2$ to appear in the exponent of the Gaussians:

$$\alpha r_1^2 + \beta r_2^2 = \gamma \left[(\mathbf{r}_1 - \mathbf{r}_2)^2 + (a\mathbf{r}_1 + b\mathbf{r}_2)^2 \right] \tag{7.20}$$

$$= \frac{\alpha\beta}{\alpha+\beta} \left[(\mathbf{r}_1 - \mathbf{r}_2)^2 + \left(\sqrt{\frac{\alpha}{\beta}}\mathbf{r}_1 + \sqrt{\frac{\beta}{\alpha}}\mathbf{r}_2 \right)^2 \right]. \tag{7.21}$$

Let us now introduce a further variable change from $(\mathbf{r}_1, \mathbf{r}_2)$ to $(\mathbf{r}, \mathbf{s})$, where

$$\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2, \qquad \mathbf{s} = \sqrt{\frac{\alpha}{\beta}}\mathbf{r}_1 + \sqrt{\frac{\beta}{\alpha}}\mathbf{r}_2; \tag{7.22}$$

The integral becomes

$$I = \int e^{-\frac{\alpha\beta}{\alpha+\beta} r^2} \frac{1}{r} e^{-\frac{\alpha\beta}{\alpha+\beta} s^2} \left| \frac{\partial(\mathbf{r}_1, \mathbf{r}_2)}{\partial(\mathbf{r}, \mathbf{s})} \right| d^3r d^3s, \tag{7.23}$$

where the Jacobian is easily calculated as the determinant of the transformation matrix, Eq.(7.22):

$$\left|\frac{\partial(\mathbf{r}_1,\mathbf{r}_2)}{\partial(\mathbf{r},\mathbf{s})}\right| = \left(\frac{\sqrt{\alpha\beta}}{\alpha+\beta}\right)^3. \tag{7.24}$$

The calculation of the integral is trivial and provides the required result:

$$g_{iljm} = \frac{2\pi^{5/2}q_e^2}{\alpha\beta(\alpha+\beta)^{1/2}} \tag{7.25}$$

where $\alpha = \alpha_i + \alpha_j$, $\beta = \alpha_l + \alpha_m$.

7.4.1 Multi-center Gaussian integrals

Matrix elements involving Gaussians centered around different centers are less straightforward to compute. A complete derivation can be found at pages 77-81 of the book of Thijssen. The following is a quick summary.

We use a basis set of gaussian functions, in which the index now labels not only the coefficient of the Gaussian but also the center (one of the two nuclei in practice):

$$b_i(\mathbf{r}) = \exp\left(-\alpha_i(\mathbf{r}-\mathbf{R}_i)^2\right). \tag{7.26}$$

The following theorem for the product of two Gaussians:

$$\exp\left(-\alpha_i(\mathbf{r}-\mathbf{R}_i)^2\right) \times \exp\left(-\alpha_j(\mathbf{r}-\mathbf{R}_j)^2\right) = K_{ij}\exp\left[-(\alpha_i+\alpha_j)(\mathbf{r}-\mathbf{R}_{ij})^2\right], \tag{7.27}$$

where

$$K_{ij} = \exp\left[-\frac{\alpha_i\alpha_j}{\alpha_i+\alpha_j}|\mathbf{R}_i-\mathbf{R}_j|^2\right], \qquad \mathbf{R}_{ij} = \frac{\alpha_i\mathbf{R}_i+\alpha_j\mathbf{R}_j}{\alpha_i+\alpha_j} \tag{7.28}$$

allows to calculate the superposition integrals as follows:

$$S_{ij} = \int b_i(\mathbf{r})b_j(\mathbf{r})d^3r = \left(\frac{\pi}{\alpha_i+\alpha_j}\right)^{3/2} K_{ij}. \tag{7.29}$$

The kinetic contribution can be calculated using the Green's theorem:

$$T_{ij} = -\int b_i(\mathbf{r})\nabla^2 b_j(\mathbf{r})d^3r = \int \nabla b_i(\mathbf{r})\nabla b_j(\mathbf{r})d^3r \tag{7.30}$$

and finally

$$T_{ij} = \frac{\alpha_i\alpha_j}{\alpha_i+\alpha_j}\left[6 - 4\frac{\alpha_i\alpha_j}{\alpha_i+\alpha_j}|\mathbf{R}_i-\mathbf{R}_j|^2\right]S_{ij}. \tag{7.31}$$

The calculation of the Coulomb interaction term with a nucleus in $\mathbf{R}$ is more complex and requires to go through Laplace transforms. At the end one gets the following expression:

$$V_{ij} = -\int b_i(\mathbf{r})\frac{1}{|\mathbf{r}-\mathbf{R}|}b_j(\mathbf{r})d^3r = -S_{ij}\frac{1}{|\mathbf{R}_{ij}-\mathbf{R}|}\mathrm{erf}\left(\sqrt{\alpha_i+\alpha_j}|\mathbf{R}_{ij}-\mathbf{R}|\right). \tag{7.32}$$

In the case $\mathbf{R}_{ij} - \mathbf{R} = 0$ we use the limit $\mathrm{erf}(x) \to 2x/\sqrt{\pi}$ to obtain

$$V_{ij} = -\frac{2\pi}{\alpha_i + \alpha_j} K_{ij}, \qquad \mathbf{R}_{ij} - \mathbf{R} = 0 \tag{7.33}$$

which reduces for $\mathbf{R}_i = \mathbf{R}_j = \mathbf{R}$ to Eq.(5.29). The bi-electronic integrals introduced in the previous chapter, Eq.(7.18), can be calculated using a similar technique:

$$\begin{aligned} g_{iljm} &= \int b_i(\mathbf{r}) b_j(\mathbf{r}) \frac{1}{|\mathbf{r} - \mathbf{r}'|} b_l(\mathbf{r}') b_m(\mathbf{r}') d^3r d^3r' \\ &= S_{ij} S_{lm} \frac{1}{|\mathbf{R}_{ij} - \mathbf{R}_{lm}|} \mathrm{erf}\left(\sqrt{\frac{(\alpha_i + \alpha_j)(\alpha_l + \alpha_m)}{\alpha_i + \alpha_j + \alpha_l + \alpha_m}} |\mathbf{R}_{ij} - \mathbf{R}_{lm}| \right) \end{aligned} \tag{7.34}$$

(beware indices!).

Although symmetry is not used in the code, it can be used to reduce by a sizable amount the number of bi-electronic integrals g_{iljm}. They are obviously invariant under exchange of i, j and l, m indices. This means that if we have N basis set functions, the number of independent matrix elements is not N^4 but M^2, where $M = N(N+1)/2$ is the number of pairs of (i, j) and (l, m) indices. The symmetry of the integral under exchange of $\mathbf{r}$ and $\mathbf{r}'$ leads to another symmetry: g_{iljm} is invariant under exchange of the (i, j) and (l, m) pairs. This further reduces the independent number of matrix elements by a factor 2, to $M(M+1)/2 \sim N^4/8$.

7.5 Code: h2_hf_gauss

Code `h2_hf_gauss.f90`[1] (or `h2_hf_gauss.c`[2]) solves the Hartree-Fock equations for the ground state of the H_2 molecule, using a basis set of S Gaussians. The basis set is composed of two sets of Gaussians, one centered around nucleus 1 and one centered around nucleus 2. As a consequence, the overlap matrix and the matrix element of the Fock matrix contain terms (*multi-centered integrals*) in which the nuclear potential and the two basis functions are not centered on the same atom.

The code requires in input a set of Gaussians coefficients (with the same format as in `hydrogen_gauss`); then it solves the SCF equations at interatomic distances d_{min}, $d_{min} + \delta d$, $d_{min} + 2\delta d$, ..., d_{max} (the parameters d_{min}, d_{max}, δd have to be provided in input). It prints on output and on file `h2.out` the electronic energy (not to be confused with the Hartree-Fock eigenvalue), nuclear repulsive energy, the sum of the two (usually referred to as "total energy"), all in Ry; finally, the difference between the total energy and the energy of the isolated atoms at infinite distance, in eV.

Note that `h2_hf_gauss.f90` also solves the H_2^+ case if the variable `do_scf` is set to `.false.`. In this case the Schrödinger equation is solved, without any SCF procedure.

[1] http://www.fisica.uniud.it/%7Egiannozz/Didattica/MQ/Software/F90/h2_hf_gauss.f90
[2] http://www.fisica.uniud.it/%7Egiannozz/Didattica/MQ/Software/C/h2_hf_gauss.c

The self-consistent procedure is even simpler than in code `helium_hf_radial`: at each step, the Fock matrix is re-calculated using the density matrix at the preceding step, with no special tricks or algorithms, until energy converges within a given numerical threshold.

7.5.1 Laboratory

- Chosen a basis set that yields a good description of isolated atoms, find the equilibrium distance by minimizing the (electronic plus nuclear) energy. Verify how sensitive the result is with respect to the dimension of the basis set. Note that the "binding energy" printed on output is calculated assuming that the isolated H atom has an energy of -1 Ry, but you should verify what the actual energy of the isolated H atom is for your basis set.

- Make a plot of the ground state molecular orbital at the equilibrium distance along the axis of the molecule. For a better view of the binding, you may also try to make a two-dimensional contour plot on a plane containing the axis of the molecule. You need to write a matrix on a uniform two-dimensional `N x M` grid in the following format:

```
   x_0  y_0  \psi(x_0,y_0)
   ...
   x_N  y_0  \psi(x_N,y_0)
(blank line)
   x_0  y_1  \psi(x_0,y_1)
   ...
   x_N  y_1  \psi(x_N,y_1)
(blank line)
   ...
```

  and gnuplot commands `set contour; unset surface; set view 0, 90` followed by `splot` "file name" `u 1:2:3 w l`

- Plot the ground state molecular orbital, together with a ligand combination of $1s$ states centered on the two H nuclei (obtained from codes for hydrogen). You should find that slightly contracted Slater orbitals, corresponding to $Z = 1.24$, yield a better fit than the $1s$ of H. Try the same for the first excited state of H_2 and the antiligand combination of $1s$ states.

- Study the limit of superposed atoms ($R \to 0$) and compare with the results of code `hydrogen_gauss` and `helium_hf_radial`. The limit of isolated atoms ($R \to \infty$) will instead yield strange results. Can you explain why? What do you expect to be wrong in the Slater determinant in this limit?

- Can you estimate the vibrational frequency of H_2?

Chapter 8

Electrons in periodic potential

The computation of electronic states in a solid is a nontrivial problem. A great simplification can be achieved if the solid is a *crystal*, i.e. if it can be described by a regular, periodic, infinite arrangement of atoms: a *crystal lattice*. In this case, under suitable assumptions, it is possible to re-conduct the solution of the many-electron problem (really many: $\mathcal{O}(10^{23})$!) to the much simpler problem of an electron under a periodic potential. Periodicity can be mathematically formalized in a simple and general way in any number of dimensions, but in the following we will consider a one-dimensional model case, that still contains the main effects of periodicity on electrons.

8.1 Crystalline solids

Let us consider an infinite periodic system of "atoms", that will be represented by a potential, centered on the atomic position. This potential will in some way – why and how being explained in solid-state physics books – the effective potential (or *crystal potential*) felt by an electron in the crystal. We will consider only *valence* electrons, i.e. those coming from outer atomic shells. *Core* electrons, i.e. those coming from inner atomic shells, are tightly bound to the atomic nucleus: their state is basically atomic-like and only marginally influenced by the presence of other atoms. We assume that the effects of core electrons can be safely included into the crystal potential. The *pseudopotential* approach formalizes the neglect of core electrons.

The assumption that core electrons do not significantly contribute to the chemical binding and that their state does not significantly change with respect to the case of isolated atoms is known as *frozen-core approximation*. This is widely used for calculations in molecules as well and usually very well verified in practice.

We also consider independent electrons, assuming implicitly that the crystal potential takes into account the Coulomb repulsion between electrons. The aim of such simplification is to obtain an easily solvable problem that still captures the essence of the physical problem. With a judicious choice of the crystal potential, we can hope to obtain a set of electronic levels that can describe the main features of the crystal. A rigorous basis for such approach can be provided

by Hartree-Fock or Density-Functional theory. In the end, the basic step is to solve to the problem of calculating the energy levels in a periodic potential.

We haven't yet said anything about the composition and the periodicity of our system. Let us simplify further the problem and assume a one-dimensional array of atoms of the same kind, regularly spaced by a distance a. The atomic position of atom n will thus be given as $a_n = na$, with n running on all integer numbers, positive and negative. In the jargon of solid-state physics, a is the *lattice parameter*, while the a_n are the *vectors of the crystal lattice*. The system has a *discrete translational invariance*, that is: it is equal to itself if translated by a or multiples of a. Called $V(x)$ the crystal potential, formed by the superposition of atomic-like potentials: $V(x) = \sum_n V_n(x - a_n)$, the following symmetry holds: $V(x+a) = V(x)$. Such symmetry plays a very important role in determining the properties of crystalline solids. Our one-dimensional space (the infinite line) can be decomposed into finite regions (segments) of space, of length a, periodically repeated. A region having such property is called *unit cell*, and the smallest possible unit cell is called *primitive cell*. Its definition contains some degree of arbitrariness: for instance, both intervals $[0, a[$ and $]-a/2, +a/2]$ define a valid primitive cell in our case.

8.1.1 Periodic Boundary Conditions

Before starting to look for a solution, we must ask ourselves how sensible it is to apply such idealized modelling to a real crystal. The latter is formed by a macroscopically large (in the order of the Avogadro number or fractions of it) but *finite* number of atoms. We might consider instead a finite system containing N atoms with $N \to \infty$, but this is not a convenient way: translational symmetry is lost, due to the presence of surfaces (in our specific 1D case, the two ends). A much more convenient and formally correct approach is to introduce *periodic boundary conditions* (PBC). Let us consider the system in a box with dimensions $L = Na$ and let us consider solutions obeying to the condition $\psi(x) = \psi(x + L)$, i.e. periodic solutions with period $L >> a$. We can imagine our wave function that arrives at one end "re-enters" from the other side. In the one-dimensional case there is a simple representation of the system: our atoms are distributed not on a straight line but on a ring, with atom N between atom $N - 1$ and atom 1.

The advantage of PBC is that we can treat the system as finite (a segment of length L in the one-dimensional case) but macroscopically large (having N atoms, with N macroscopically large if a is a typical interatomic distance and L the typical size of a piece of crystal), still retaining the discrete translational invariance. Case $N \to \infty$ describes the so-called *thermodynamical limit*. It is to be noticed that a crystal with PBC has no surface. As a consequence there is no "inside" and "outside" the crystal: the latter is not contemplated. This is the price to pay for the big advantage of being able to use translational symmetry.

In spite of PBC and of translational symmetry, the solution of the Schrödinger equation for a periodic potential does not yet look like a simple problem. We will need to find a number of single-particle states equal to at least half the number of electrons in the system, assuming that the many-body wave function

is build as an anti-symmetrized product of single-electron states taken as spin-up and spin-down pairs (as in the case of He and H_2). Of course the resulting state will have zero magnetization ($S = 0$). The exact number of electrons in a crystal depends upon its atomic composition. Even if we assume the minimal case of one electron per atom, we still have N electrons and we need to calculate $N/2$ states, with $N \to \infty$. How can we deal with such a macroscopic number of states?

8.1.2 Bloch Theorem

At this point symmetry theory comes to the rescue under the form of the *Bloch theorem*. Let us indicate with $\mathcal{T}$ the discrete translation operator: $\mathcal{T}\psi(x) = \psi(x+a)$. What is the form of the eigenvalues and eigenvectors of $\mathcal{T}$? It can be easily verified (and rigorously proven) that $\mathcal{T}\psi(x) = \lambda\psi(x)$ admits as solution $\psi_k(x) = \exp(ikx)u_k(x)$, where k is a real number, $u_k(x)$ is a periodic function of period a: $u_k(x+a) = u_k(x)$. This result is easily generalized to three dimensions, where k is a vector: the *Bloch vector*. States ψ_k are called *Bloch states*. It is easy to verify that for Bloch states the following relation hold:

$$\psi_k(x+a) = \psi_k(x)e^{ika}. \tag{8.1}$$

Let us classify our solution using the Bloch vector k (in our case, a one-dimensional vector, i.e. a number). The Bloch vector is related to the eigenvalue of the translation operator (we remind that H and $\mathcal{T}$ are commuting operators). Eq.(8.1) suggests that all k differing by a multiple of $2\pi/a$ are equivalent (i.e. they correspond to the same eigenvalue of $\mathcal{T}$). It is thus convenient to restrict to the following interval of values for k: k: $-\pi/a < k \leq \pi/a$. Values of k outside such interval are brought back into the interval by a translation $G_n = 2\pi n/a$.

We must moreover verify that our Bloch states are compatible with PBC. It is immediate to verify that only values of k such that $\exp(ikL) = 1$ are compatible with PBC, that is, k must be an integer multiple of $2\pi/L$. As a consequence, for a finite number N of atoms (i.e. for a finite dimension $L = Na$ of the box), there are N admissible values of k: $k_n = 2\pi n/L$, con $n = -N/2, ..., N/2$ (note that $k_{-N/2} = -\pi/a$ is equivalent to $k_{N/2} = \pi/a$). In the thermodynamical limit, $N \to \infty$, these N Bloch vectors will form a dense set between $-\pi/a$ and π/a, in practice a continuum.

8.1.3 The empty potential

Before moving towards the solution, let us consider the case of the simplest potential one can think of: the non-existent potential, $V(x) = 0$. Our system will have plane waves as solutions: $\psi_k(x) = (1/\sqrt{L})\exp(ikx)$, where the factor ensure the normalization. k may take any value, as long as it is compatible with PBC, that is, $k = 2\pi n/L$, with n any integer. The energy of the solution with wave vector k will be purely kinetic, and thus:

$$\psi_k(x) = \frac{1}{\sqrt{L}}e^{ikx}, \qquad \epsilon(k) = \frac{\hbar^2 k^2}{2m}. \tag{8.2}$$

In order to obtain the same description as for a periodic potential, we simply "refold" the wave vectors k into the interval $-\pi/a < k \leq \pi/a$, by applying the translations $G_n = 2\pi n/a$. Let us observe the energies as a function of the "refolded" k, Eq.(8.2): for each value of k in the interval $-\pi/a < k \leq \pi/a$ there are many (actually infinite) states with energy given by $\epsilon_n(k) = \hbar^2(k+G_n)^2/2m$. The corresponding Bloch states have the form

$$\psi_{k,n}(x) = \frac{1}{\sqrt{L}} e^{ikx} u_{k,n}(x), \qquad u_{k,n}(x) = e^{iG_n x}. \tag{8.3}$$

The function $u_{k,n}(x)$ is by construction periodic. Notice that we have moved from an "extended" description, in which the vector k covers the entire space, to a "reduced" description in which k is limited between $-\pi/a$ and π/a. Also for the space of vectors k, we can introduce a "unit cell", $]-\pi/a, \pi/a]$, periodically repeated with period $2\pi/a$. Such cell is also called *Brillouin Zone* (BZ). It is immediately verified that the periodicity in k-space is given by the so-called *reciprocal lattice*: a lattice of vectors G_n such that $G_n \cdot a_m = 2\pi p$, where p is an integer.

8.1.4 Solution for the crystal potential

Let us now consider the case of a "true", non-zero periodic potential: we can think at it as a sum of terms centered on our "atoms"‘:

$$V(x) = \sum_n v(x - na), \tag{8.4}$$

but this is not strictly necessary. We observe first of all that the Bloch theorem allows the separation of the problem into independent sub-problems for each k. If we insert the Bloch form, Eq.(8.1), into the Schrödinger equation:

$$(T + V(x))e^{ikx} u_k(x) = E e^{ikx} u_k(x), \tag{8.5}$$

we get an equation for the periodic part $u_k(x)$:

$$\left[\frac{\hbar^2}{2m}\left(k^2 - 2ik\frac{d}{dx} - \frac{d^2}{dx^2}\right) + V(x) - E\right] u_k(x) = 0 \tag{8.6}$$

that has in general an infinite discrete series of solutions, orthogonal between them:

$$\int_{-L/2}^{L/2} u^*_{k,n}(x) u_{k,m}(x) dx = \delta_{nm} N \int_{-a/2}^{a/2} |u_{k,n}(x)|^2 dx, \tag{8.7}$$

where we have made usage of the periodicity of functions $u(x)$ to re-conduct the integral on the entire crystal (from $-L/2$ to $L/2$) to an integration on the unit cell only (from $-a/2$ to $a/2$). In the following, however, we are not going to use such equations. Notice that the solutions having different k are by construction orthogonal. Let us write the superposition integral between Bloch states for different k:

$$\begin{aligned} \int_{-L/2}^{L/2} \psi^*_{k,n}(x)\psi_{k',m}(x)dx &= \int_{-L/2}^{L/2} e^{i(k'-k)x} u^*_{k,n}(x) u_{k',m}(x) dx \qquad (8.8) \\ &= \left(\sum_p e^{ip(k'-k)a}\right) \int_{-a/2}^{a/2} e^{i(k'-k)x} u^*_{k,n}(x) u_{k',m}(x) dx, \end{aligned}$$

where the sum over p runs over all the N vectors of the lattice. The purely geometric factor multiplying the integral differs from zero only if k and k' coincide:

$$\sum_p e^{ip(k'-k)a} = N\delta_{k,k'}. \tag{8.9}$$

We have used Kronecker's delta, not Dirac's delta, because the k form a dense but still finite set (there are N of them). We note that the latter orthogonality relation holds irrespective of the periodic part $u(x)$ of Bloch states. There is no reason to assume that the periodic parts of the Bloch states at different k are orthogonal: only those for different Bloch states at the same k are orthogonal (see Eq.(8.7)).

8.1.5 Plane-wave basis set

Let us come back to the numerical solution. We look for the solution using a plane-wave basis set. This is especially appropriate for problems in which the potential is periodic. We cannot choose "any" plane-wave set, though: the correct choice is restricted by the Bloch vector and by the periodicity of the system. Given the Bloch vector k, the "right" plane-wave basis set is the following:

$$b_{n,k}(x) = \frac{1}{\sqrt{L}} e^{i(k+G_n)x}, \qquad G_n = \frac{2\pi}{a} n. \tag{8.10}$$

The "right" basis must in fact have a $\exp(ikx)$ behavior, like the Bloch states with Bloch vector k; moreover the potential must have nonzero matrix elements between plane waves. For a periodic potential like the one in Eq.(8.4), matrix elements:

$$\langle b_{i,k}|V|b_{j,k}\rangle = \frac{1}{L}\int_{-L/2}^{L/2} V(x)e^{-iGx}dx \tag{8.11}$$

$$= \frac{1}{L}\left(\sum_p e^{-ipGa}\right)\int_{-a/2}^{a/2} V(x)e^{-iGx}dx, \tag{8.12}$$

where $G = G_i - G_j$, are non-zero only for a discrete set of values of G. In fact, the factor $\sum_p e^{-ipGa}$ is zero except when Ga is a multiple of 2π, i.e. only on the reciprocal lattice vectors G_n defined above. One finally finds

$$\langle b_{i,k}|V|b_{j,k}\rangle = \frac{1}{a}\int_{-a/2}^{a/2} V(x)e^{-i(G_i-G_j)x}dx. \tag{8.13}$$

The integral is calculated in a single unit cell and, if expressed as a sum of atomic terms localized in each cell, for a single term in the potential. Note that the factor N cancels and thus the $N \to \infty$ thermodynamic limit is well defined.

In the simple case that will be presented, the matrix elements of the Hamiltonian, Eq.(8.13), can be analytically computed by straight integration. Another case in which an analytic solution is known is a crystal potential written as a sum of Gaussian functions:

$$V(x) = \sum_{p=0}^{N-1} v(x-pa), \qquad v(x) = Ae^{-\alpha x^2}. \tag{8.14}$$

This yields

$$\langle b_{i,k}|V|b_{j,k}\rangle = \frac{1}{a}\int_{-L/2}^{L/2} Ae^{-\alpha x^2}e^{-iGx}dx \tag{8.15}$$

The integral is known (it can be calculated using the tricks and formulae given in previous sections, extended to complex plane):

$$\int_{-\infty}^{\infty} e^{-\alpha x^2}e^{-iGx}dx = \sqrt{\frac{\pi}{\alpha}}e^{-G^2/4\alpha} \tag{8.16}$$

(remember that in the thermodynamical limit, $L \to \infty$).

Fast Fourier Transform

For a generic potential, one has to resort to numerical methods to calculate the integral. One advantage of the plane-wave basis set is the possibility to exploit the properties of Fourier Transforms (FT).

Let us *discretize* our problem in real space, by introducing a grid of n points $x_i = ia/n$, $i = 0, n-1$ in the unit cell. Note that due to periodicity, grid points with index $i \geq n$ or $i < 0$ are "refolded" into grid points in the unit cell (that is, $V(x_{i+n}) = V(x_i)$, and in particular, x_n is equivalent to x_0. Let us introduce the function f_j defined as follows:

$$f_j = \frac{1}{L}\int V(x)e^{-iG_j x}dx, \qquad G_j = j\frac{2\pi}{a}, \quad j = 0, n-1. \tag{8.17}$$

We can exploit periodicity to show that

$$f_j = \frac{1}{a}\int_0^a V(x)e^{-iG_j x}dx. \tag{8.18}$$

This is nothing but the FT $\tilde{f}(G_j)$, with a slightly different factor ($1/a$ instead of $1/\sqrt{a}$) with respect to the definition of Eq.(4.43). Note that the integration limits can be translated at will, again due to periodicity. Let us write now such integrals as a finite sum over grid points (with $\Delta x = a/n$ as finite integration step):

$$\begin{aligned} f_j &= \frac{1}{a}\sum_{m=0}^{n-1} V(x_m)e^{-iG_j x_m}\Delta x \\ &= \frac{1}{n}\sum_{m=0}^{n-1} V(x_m)e^{-iG_j x_m} \\ &= \frac{1}{n}\sum_{m=0}^{n-1} V_m \exp[-2\pi\frac{jm}{n}i], \qquad V_m \equiv V(x_m). \end{aligned} \tag{8.19}$$

Notice that the FT is now a *periodic* function in the variable G, with period $G_n = 2\pi n/a$! This shouldn't come as a surprise though: the FT of a periodic function is a discrete function, the FT of a discrete function is periodic.

It is easy to verify that the potential in real space can be obtained back from its FT as follows:

$$V(x) = \sum_{j=0}^{n-1} f_j e^{iG_j x}, \tag{8.20}$$

yielding the inverse FT in discretized form:

$$V_j = \sum_{m=0}^{n-1} f_m \exp[2\pi \frac{jm}{n} i], \quad j = 0, n-1. \tag{8.21}$$

The two operations of Eq.(8.19) and (8.21) are called *Discrete Fourier Transform*. One may wonder where have all the G vectors with negative values gone: after all, we would like to calculate f_j for all j such that $|G_j|^2 < E_c$ (for some suitably chosen value of E_c), not for G_j with j ranging from 0 to $n-1$. The periodicity of the discrete FT in both real and reciprocal space however allows to refold the G_j on the "right-hand side of the box", so to speak, to negative G_j values, by making a translation of $2\pi n/a$.

The discrete FT of a function defined on a grid of n points requires $\mathcal{O}(n^2)$ operations: a sum over n terms for each of the n points. There is however a recursive version of the algorithm, the Fast FT or FFT, which can do the transform in $\mathcal{O}(n \log n)$ operations. The difference may not seem so important but it is: the FFT is at the heart of many algorithms used in different fields.

An example of usage of FFTs is provided in codes `testfft.f90`[1] and `testfft.c`[2]. Their compilation and linking requires the FFTW[3] v.3 library. If properly installed on your system, it is sufficient to specify `-lfftw3`. In some cases you may need to specify `-I` followed by the directory where files to be included: `fftw3.f03`[4] for Fortran, `fftw3.h`[5] for C, can be found. Note that Intel MKL libraries contain, in addition to all BLAS and LAPACK, FFTW3 routines as well.

8.2 Code: periodicwell

Let us now move to the practical solution of a "true", even if model, potential: the periodic potential well, known in solid-state physics since the thirties under the name of *Kronig-Penney model*:

$$V(x) = \sum_n v(x - na), \quad v(x) = -V_0 \quad |x| \leq \frac{b}{2}, \quad v(x) = 0 \quad |x| > \frac{b}{2} \tag{8.22}$$

and of course $a \geq b$. Such model is exactly soluble in the limit $b \to 0$, $V_0 \to \infty$, $V_0 b \to$constant.

The needed ingredients for the solution in a plane-wave basis set are almost all already found in Sec.(4.3) and (4.4), where we have shown the numerical

[1] http://www.fisica.uniud.it/%7Egiannozz/Didattica/MQ/Software/F90/testfft.f90
[2] http://www.fisica.uniud.it/%7Egiannozz/Didattica/MQ/Software/C/testfft.c
[3] http://www.fftw.org
[4] http://www.fisica.uniud.it/%7Egiannozz/Didattica/MQ/Software/F90/fftw3.f03
[5] http://www.fisica.uniud.it/%7Egiannozz/Didattica/MQ/Software/C/fftw3.h

solution on a plane-wave basis set of the problem of a single potential well. Code `periodicwell.f90`[6] (or `periodicwell.c`[7]) is in fact a trivial extension of code `pwell`. Such code in fact uses a plane-wave basis set like the one in Eq.(8.10), which means that it actually solves the periodic Kronig-Penney model for $k = 0$. If we increase the size of the cell until this becomes large with respect to the dimension of the single well, then we solve the case of the isolate potential well.

The generalization to the periodic model only requires the introduction of the Bloch vector k. Our base is given by Eq.(8.10). In order to choose when to truncate it, it is convenient to consider plane waves up to a maximum (*cutoff*) kinetic energy:

$$\frac{\hbar^2(k+G_n)^2}{2m} \leq E_{cut}. \tag{8.23}$$

Bloch wave functions are expanded into plane waves:

$$\psi_k(x) = \sum_n c_n b_{n,k}(x) \tag{8.24}$$

and are automatically normalized if $\sum_n |c_n|^2 = 1$. The matrix elements of the Hamiltonian are very simple:

$$H_{ij} = \langle b_{i,k}|H|b_{j,k}\rangle = \delta_{ij}\frac{\hbar^2(k+G_i)^2}{2m} + \widetilde{V}(G_i - G_j), \tag{8.25}$$

where $\widetilde{V}(G)$ is the Fourier transform of the crystal potential, defined as in Eq.(8.18). Code `pwell` may be entirely recycled and generalized to the solution for Bloch vector k. It is convenient to introduce a cutoff parameter E_{cut} for the truncation of the basis set. This is preferable to setting a maximum number of plane waves, because the convergence depends only upon the modulus of $k+G$. The number of plane waves, instead, also depends upon the dimension a of the unit cell.

Code `periodicwell` requires in input the well depth, V_0, the well width, b, the unit cell length, a. Internally, a loop over k points covers the entire BZ (that is, the interval $[-\pi/a, \pi/a]$ in this specific case), calculates $E(k)$, writes the lowest $E(k)$ values to files `bands.out` in an easily plottable format.

8.2.1 Laboratory

- Plot $E(k)$, that goes under the name of *band structure*, or also *dispersion*. Note that if the potential is weak (the so-called *quasi-free electrons* case), its main effect is to induce the appearance of intervals of forbidden energy (i.e.: of energy values to which no state corresponds) at the boundaries of the BZ. In the jargon of solid-state physics, the potential *opens a gap*. This effect can be predicted on the basis of perturbation theory.

- Observe how $E(k)$ varies as a function of the periodicity and of the well depth and width. As a rule, a band becomes wider (more *dispersed*, in

[6] http://www.fisica.uniud.it/%7Egiannozz/Didattica/MQ/Software/F90/periodicwell.f90
[7] http://www.fisica.uniud.it/%7Egiannozz/Didattica/MQ/Software/C/periodicwell.c

the jargon of solid-state physics) for increasing superposition of the atomic states.

- Plot for a few low-lying bands the Bloch states in real space (borrow and adapt the code from `pwell`). Remember that Bloch states are complex for a general value of k. Look in particular at the behavior of states for $k = 0$ and $k = \pm\pi/a$ (the "zone boundary"). Can you understand their form?

Chapter 9

Pseudopotentials

In general, the band structure of a solid will be composed both of more or less extended states, coming from outer atomic orbitals, and of strongly localized (core) states, coming from deep atomic levels. Extended states are the interesting part, since they determine the (structural, transport, etc.) properties of the solid. The idea arises naturally to get rid of core states by replacing the true Coulomb potential and core electrons with a *pseudopotential* (or *effective core potential* in Quantum Chemistry parlance): an effective potential that "mimics" the effects of the nucleus and the core electrons on valence electrons. A big advantage of the pseudopotential approach is to allow the usage of a plane-wave basis set in realistic calculations.

9.1 Three-dimensional crystals

Let us consider now a more realistic (or slightly less unrealistic) model of a crystal. The description of periodicity in three dimensions is a straightforward generalization of the one-dimensional case, although the resulting geometries may look awkward to an untrained eye. The lattice vectors, $\mathbf{R}_n$, can be written as a sum with integer coefficients, n_i:

$$\mathbf{R}_n = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3 \tag{9.1}$$

of three primitive vectors, $\mathbf{a}_i$. There are 14 different types of lattices, known as *Bravais lattices*. The nuclei can be found at all sites $\mathbf{d}_\mu + \mathbf{R}_n$, where $\mathbf{d}_\mu$ runs on all atoms in the unit cell (that may contain from 1 to thousands of atoms!). It can be demonstrated that the volume Ω of the unit cell is given by $\Omega = \mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)$, i.e. the volume contained in the parallelepiped formed by the three primitive vectors. We remark that the primitive vectors are in general linearly independent (i.e. they do not lye on a plane) but not orthogonal.

The crystal is assumed to be contained into a box containing a macroscopic number N of unit cells, with PBC imposed as follows:

$$\psi(\mathbf{r} + N_1\mathbf{a}_1 + N_2\mathbf{a}_2 + N_3\mathbf{a}_3) = \psi(\mathbf{r}). \tag{9.2}$$

Of course, $N = N_1 \cdot N_2 \cdot N_3$ and the volume of the crystal is $V = N\Omega$.

A reciprocal lattice of vectors $\mathbf{G}_m$ such that $\mathbf{G}_m \cdot \mathbf{R}_n = 2\pi p$, with p integer, is introduced. It can be shown that

$$\mathbf{G}_m = m_1\mathbf{b}_1 + m_2\mathbf{b}_2 + m_3\mathbf{b}_3 \tag{9.3}$$

with m_i integers and the three vectors $\mathbf{b}_j$ given by

$$\mathbf{b}_1 = \frac{2\pi}{\Omega}\mathbf{a}_2 \times \mathbf{a}_3, \quad \mathbf{b}_2 = \frac{2\pi}{\Omega}\mathbf{a}_3 \times \mathbf{a}_2, \quad \mathbf{b}_3 = \frac{2\pi}{\Omega}\mathbf{a}_1 \times \mathbf{a}_2 \tag{9.4}$$

(note that $\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi\delta_{ij}$). The Bloch theorem generalizes to

$$\psi(\mathbf{r} + \mathbf{R}) = e^{i\mathbf{k}\cdot\mathbf{R}}\psi(\mathbf{r}) \tag{9.5}$$

where the Bloch vector $\mathbf{k}$ is any vector obeying the PBC. Bloch vectors are usually taken into the three-dimensional Brillouin Zone (BZ), that is, the unit cell of the reciprocal lattice. It is a straightforward exercise in vector algebra to show that the volume Ω_{BZ} of the Brillouin Zone is related to the volume of the unit cell by $\Omega_{BZ} = 8\pi^3/\Omega$.

It can be shown that there are N Bloch vectors compatible with the box defined in Eq.(9.2); in the thermodynamical limit $N \to \infty$, the Bloch vector becomes a continuous variable as in the one-dimensional case. We remark that this means that at each k-point we have to "accommodate" ν electrons, where ν is the number of electrons in the unit cell. For a nonmagnetic, spin-unpolarized insulator, this means $\nu/2$ filled states. In semiconductor physics, occupied states are called "valence bands", while empty states are called "conduction bands". We write the electronic states as $\psi_{\mathbf{k},i}$ where $\mathbf{k}$ is the Bloch vector and i is the band index.

9.2 Plane waves, core states, pseudopotentials

For a given lattice, the plane wave basis set for Bloch states of vector $\mathbf{k}$ is

$$b_{n,\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{V}}e^{i(\mathbf{k}+\mathbf{G}_n)\cdot\mathbf{r}} \tag{9.6}$$

where $\mathbf{G}_n$ are reciprocal lattice vector. A finite basis set can be obtained, as seen in the previous section, by truncating the basis set up to some cutoff on the kinetic energy:

$$\frac{\hbar^2(\mathbf{k}+\mathbf{G}_n)^2}{2m} \leq E_{cut}. \tag{9.7}$$

In realistic crystals, however, E_{cut} must be very large in order to get a good description of the electronic states. The reason is the very localized character of the core, atomic-like orbitals, and the extended character of plane waves. Let us consider core states in a crystal: their orbitals will be very close to the corresponding states in the atoms and will exhibit the same strong oscillations. Moreover, these strong oscillations will be present in valence (i.e. outer) states as well, because of orthogonality (for this reason these strong oscillations are referred to as "orthogonality wiggles"). Reproducing highly localized functions

that vary strongly in a small region of space requires a large number of delocalized functions such as plane waves.

Let us estimate how large is this large number using Fourier analysis. In order to describe features which vary on a length scale δ, one needs Fourier components up to $q_{max} \sim 2\pi/\delta$. In a crystal, our wave vectors $\mathbf{q} = \mathbf{k} + \mathbf{G}$ have discrete values. There will be a number N_{PW} of plane waves approximately equal to the volume of the sphere of radius q_{max}, divided by the volume Ω_{BZ} of the unit cell of the reciprocal lattice, i.e. of the BZ:

$$N_{PW} \simeq \frac{4\pi q_{max}^3}{3\Omega_{BZ}}. \tag{9.8}$$

A simple estimate for diamond is instructive. The $1s$ orbital of the carbon atom has its maximum around 0.3 a.u., so $\delta \simeq 0.1$ a.u. is a reasonable value. Diamond has a "face-centered cubic" (fcc) lattice with lattice parameter $a_0 = 6.74$ a.u. and primitive vectors:

$$\mathbf{a}_1 = a_0 \left(0, \frac{1}{2}, \frac{1}{2}\right), \quad \mathbf{a}_2 = a_0 \left(\frac{1}{2}, 0, \frac{1}{2}\right), \quad \mathbf{a}_3 = a_0 \left(\frac{1}{2}, \frac{1}{2}, 0\right). \tag{9.9}$$

The unit cell has a volume $\Omega = a_0^3/4$, the BZ has a volume $\Omega_{BZ} = (2\pi)^3/(a_0^3/4)$. Inserting the data, one finds $N_{PW} \sim 250,000$ plane wave, clearly too much for practical use.

It is however possible to use a plane wave basis set in conjunction with *pseudopotentials*: an effective potential that "mimics" the effects of the nucleus and the core electrons on valence electrons. The true electronic valence orbitals are replaced by "pseudo-orbitals" that do not have the orthogonality wiggles typical of true orbitals. As a consequence, they are well described by a much smaller number of plane waves.

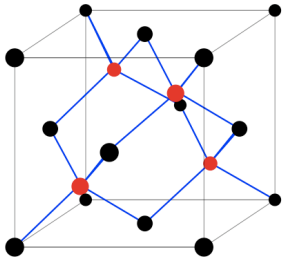
Pseudopotentials have a long history, going back to the 30's. Born as a rough and approximate way to get decent band structures, they have evolved into a sophisticated and exceedingly useful tool for accurate and predictive calculations in condensed-matter physics.

9.3 Code: cohenbergstresser

Code `cohenbergstresser.f90`[1] (or `cohenbergstresser.c`[2]) implements the calculation of the band structure in Si using the pseudopotentials published by M. L. Cohen and T. K. Bergstresser, Phys. Rev. **141**, 789 (1966). These are "empirical" pseudopotentials, i.e. devised to reproduce available experimental data, and not derived from first principles.

[1] http://www.fisica.uniud.it/%7Egiannozz/Didattica/MQ/Software/F90/cohenbergstresser.f90
[2] http://www.fisica.uniud.it/%7Egiannozz/Didattica/MQ/Software/C/cohenbergstresser.c

Si has the same crystal structure as Diamond: a face-centered cubic lattice with two atoms in the unit cell. In the figure, the black and red dots identify the two sublattices. The side of the cube is the lattice parameter a_0. In the Diamond structure, the two sublattices have the same composition; in the zincblende structure (e.g. GaAs), they have different composition.



The origin of the coordinate system is arbitrary; typical choices are one of the atoms, or the middle point between two neighboring atoms. We use the latter choice because it yields inversion symmetry. The Si crystal can thus be described [3] by three primitive vectors as in Eq.(9.9) with lattice parameter $a_0 = 10.26$ a.u., and two atoms in the unit cell at positions $\mathbf{d}_1 = -\mathbf{d}$, $\mathbf{d}_2 = +\mathbf{d}$, where

$$\mathbf{d} = a_0 \left(\frac{1}{8}, \frac{1}{8}, \frac{1}{8}\right). \tag{9.12}$$

The reciprocal lattice of the fcc lattice is a "body-centered cubic" (bcc) lattice, whose primitive vectors are

$$\mathbf{b}_1 = \frac{2\pi}{a_0}(1,1,-1), \quad \mathbf{b}_2 = \frac{2\pi}{a_0}(1,-1,1), \quad \mathbf{b}_3 = \frac{2\pi}{a_0}(-1,1,1) \tag{9.13}$$

(again, the choice is not unique).

Let us re-examine the matrix elements between plane waves of a potential V, given by a sum of spherically symmetric potentials V_μ centered around atomic positions:

$$V(\mathbf{r}) = \sum_n \sum_\mu V_\mu(|\mathbf{r} - \mathbf{d}_\mu - \mathbf{R}_n|) \tag{9.14}$$

With some algebra, one finds:

$$\langle b_{i,\mathbf{k}}|V|b_{j,\mathbf{k}}\rangle = \frac{1}{\Omega}\int_\Omega V(\mathbf{r}) e^{-i\mathbf{G}\cdot\mathbf{r}} d\mathbf{r} = V_{Si}(G)\cos(\mathbf{G}\cdot\mathbf{d}), \tag{9.15}$$

where $\mathbf{G} = \mathbf{G}_i - \mathbf{G}_j$. The cosine term is a special case of a geometrical factor known as *structure factor*, while $V_{Si}(G)$ is known as the *atomic form factor*:

$$V_{Si}(G) = \frac{1}{\Omega}\int_\Omega V_{Si}(r) e^{-i\mathbf{G}\cdot\mathbf{r}} d\mathbf{r}. \tag{9.16}$$

Cohen-Bergstresser pseudopotentials are given as atomic form factors for a few values of $|\mathbf{G}|$, corresponding to the smallest allowed modules: $G^2 = 0, 3, 4, 8, 11, ...$, in units of $(2\pi/a_0)^2$.

[3] We remark that the face-centered cubic lattice can also be described as a simple-cubic lattice:

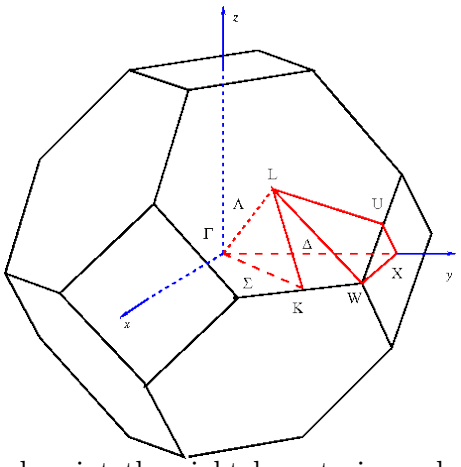
$$\mathbf{a}_1 = a_0(1,0,0), \quad \mathbf{a}_2 = a_0(0,1,0), \quad \mathbf{a}_3 = a_0(0,0,1) \tag{9.10}$$

with four atoms in the unit cell, at positions:

$$\mathbf{d}_1 = a_0\left(0, \frac{1}{2}, \frac{1}{2}\right), \quad \mathbf{d}_2 = a_0\left(\frac{1}{2}, 0, \frac{1}{2}\right), \quad \mathbf{d}_3 = a_0\left(\frac{1}{2}, \frac{1}{2}, 0\right), \quad \mathbf{d}_4 = (0,0,0). \tag{9.11}$$

The code requires on input the cutoff (in Ry) for the kinetic energy of plane waves, and a list of vectors $\mathbf{k}$ in the Brillouin Zone. Traditionally these points are chosen along high-symmetry lines, joining high-symmetry points shown in figure and listed below:

$$
\begin{aligned}
\Gamma &= (0,0,0), \\
X &= \frac{2\pi}{a_0}(1,0,0), \\
W &= \frac{2\pi}{a_0}(1,\frac{1}{2},0), \\
K &= \frac{2\pi}{a_0}(\frac{3}{4},\frac{3}{4},0), \\
L &= \frac{2\pi}{a_0}(\frac{1}{2},\frac{1}{2},\frac{1}{2}),
\end{aligned}
$$



On output the code prints for each k-point the eight lowest eigenvalues, corresponding to the four[4] valence (occupied) bands and the four conduction (empty) bands.

9.3.1 Laboratory

- Verify which cutoff is needed to achieve converged results for $E(\mathbf{k})$.
- Understand and reproduce the results in the original paper for Si, Ge, Sn. [5] You may either try to plot the band structure along a few high-symmetry lines, or compare some selected energy differences.
- Try to figure out what the charge density would be by plotting the sum of the four lowest wave functions squared at the Γ point. It is convenient to plot along the (110) plane (that is: one axis along (1,1,0), the other along (0,0,1)).
- In the zincblende lattice, the two atoms are not identical. Cohen and Bergstresser introduce a "symmetric" and an "antisymmetric" contribution, corresponding respectively to a cosine and a sine times the imaginary unit in the structure factor:

$$\langle b_{i,\mathbf{k}}|V|b_{j,\mathbf{k}}\rangle = V_s(G)\cos(\mathbf{G}\cdot\mathbf{d}) + iV_a(G)\sin(\mathbf{G}\cdot\mathbf{d}). \tag{9.17}$$

  What do you think is needed in order to extend the code to cover the case of Zincblende lattices?

[4] The Si atom, whose electronic configuration is $1s^2 2s^2 2p^6 3s^2 3p^2$, has 10 core and 4 valence electrons, thus crystal Si has 8 valence electrons per unit cell and 4 valence bands.

[5] Remember that absolute values of $E(\mathbf{k})$ have no physical meaning: the zero of the energy is not defined for a crystal with PBC, since there is no reference level, no "inside" and "outside". In the paper, $E = 0$ is set at the top of the valence band.

Chapter 10

Exact diagonalization of quantum spin models

Systems of interacting spins are used since many years to model magnetic phenomena. Their usefulness extends well beyond the field of magnetism, since many different physical phenomena can be mapped, exactly or approximately, onto spin systems. Many models exists and many techniques can be used to determine their properties under various conditions. In this chapter we will deal with the exact solution (i.e. finding eigenvalues and eigenvectors) for the *Heisenberg model*, i.e. a quantum spin model in which spins centered at lattice sites interact via the *exchange interaction*. The hyper-simplified model we are going to study is sufficient to give an idea of the kind of problems one encounters when trying to solve many-body systems without resorting to mean-field approximations (i.e. reducing the many-body problem to that of a single spin under an effective field generated by the other spins). Moreover it allows to introduce two very important concepts in numerical analysis: *iterative diagonalization* and *sparseness* of a matrix.

10.1 The Heisenberg model

Let us consider a set of atoms in a crystal, each atom having a magnetic moment, typically due to localized, partially populated states such as $3d$ states in transition metals and $4f$ states in rare earths. The energy of the crystal may in general depend upon the orientation of the magnetic moments. In many cases [1] these magnetic moments tend to spontaneously orient (at sufficiently low temperatures) along a given axis, in the same direction. This phenomenon is known as *ferromagnetism*. Other kinds of ordered structures are also known, and in particular *antiferromagnetism*: two or more sublattices of atoms are formed, having opposite magnetization. Well before the advent of quantum mechanics, it was realized that these phenomena could be quite well modeled by a system of interacting magnetic moments. The origin of the interaction was however mysterious, since direct dipole-dipole interaction is way too small to justify the

[1] but not for our model: it can be demonstrated that the magnetization vanishes at $T \neq 0$, for all 1-d models with short-range interactions only

observed behavior. The microscopic origin of the interaction was later found in the antisymmetry of the wave functions and in the constraints it imposes on the electronic structure (this is why it is known as *exchange interaction*).

One of the phenomenological models used to study magnetism is the Heisenberg model. This consists in a system of quantum spins $\mathbf{S}_i$, localized at lattice sites i, described by a *spin Hamiltonian*:

$$H = -\sum_{<ij>} \left(J_x(ij)S_x(i)S_x(j) + J_y(ij)S_y(i)S_y(j) + J_z(ij)S_z(i)S_z(j)\right) \quad (10.1)$$

The sum runs over all pairs of spins.

In the following, we will restrict to the simpler case of a single isotropic interaction energy J between nearest neighbors only:

$$H = -J \sum_{<ij>} \mathbf{S}(i) \cdot \mathbf{S}(j). \quad (10.2)$$

The restriction to nearest-neighbor interactions only makes physical sense, since in most physically relevant cases the exchange interaction is short-ranged. We will also restrict ourselves to the case $S = 1/2$.

10.2 Hilbert space in spin systems

The ground state of a spin system can be exactly found in principle, since the Hilbert space is finite: it is sufficient to diagonalize the Hamiltonian over a suitable basis set of finite dimension. The Hilbert space of spin systems is in fact formed by all possible linear combinations of products:

$$|\mu\rangle = |\sigma_\mu(1)\rangle \otimes |\sigma_\mu(2)\rangle \otimes \ldots \otimes |\sigma_\mu(N)\rangle \quad (10.3)$$

where N is the number of spins and the $\sigma_\mu(i)$ labels the two possible spin states ($\sigma = -1/2$ or $\sigma = +1/2$) for the $i-$th spin. The Hilbert space has dimension $N_h = 2^N$ (or $N_h = (2S+1)^N$ for spin S), thus becoming quickly intractable for N as small as a few tens (e.g. for $N = 30$, $N_h \sim 1$ billion). It is however possible to reduce the dimension of the Hilbert space by exploiting some symmetries of the system, or by restricting to states of given total magnetization. For a system of N spins, n up and $N - n$ down, it can be easily demonstrated that $N_h = N!/n!/(N-n)!$. For 30 spins, this reduces the dimension of the Hilbert space to "only" 155 millions at most. The solution "reduces" (so to speak) to the diagonalization of the $N_h \times N_h$ Hamiltonian matrix $H_{\mu,\nu} = \langle\mu|H|\nu\rangle$, where μ and ν run on all possible N_h states.

For a small number of spins, up to 12-13, the size of the problem may still tractable with today's computers. For a larger number of spin, one has to resort to techniques exploiting the *sparseness* of the Hamiltonian matrix. The number of nonzero matrix elements is in fact much smaller than the total number of matrix elements. Let us re-write the spin Hamiltonian under the following form:

$$H = -\frac{J}{2} \sum_{<ij>} \left(S_+(i)S_-(j) + S_-(i)S_+(j) + 2S_z(i)S_z(j)\right). \quad (10.4)$$

The only nonzero matrix elements for the two first terms are between states $|\mu\rangle$ and $|\nu\rangle$ states such that $\sigma_\mu(k) = \sigma_\nu(k)$ for all $k \neq i, j$, while for $k = i, j$:

$$\langle\alpha(i)| \otimes \langle\beta(j)|S_+(i)S_-(j)|\beta(i)\rangle \otimes |\alpha(j)\rangle \tag{10.5}$$

$$\langle\beta(i)| \otimes \langle\alpha(j)|S_-(i)S_+(j)|\alpha(i)\rangle \otimes |\beta(j)\rangle \tag{10.6}$$

where $\alpha(i), \beta(i)$ mean $i-$th spin up and down, respectively. The term $S_z(i)S_z(j)$ is diagonal, i.e. nonzero only for $\mu = \nu$.

Sparseness, in conjunction with symmetry, can be used to reduce the Hamiltonian matrix into blocks of much smaller dimensions that can be diagonalized with a much reduced computational effort.

10.3 Iterative diagonalization

In addition to sparseness, there is another aspect that can be exploited to make the calculation more tractable. Typically one is interested in the ground state and in a few low-lying excited states, not in the entire spectrum. Calculating just a few eigenstates, however, is just marginally cheaper than calculating all of them, with conventional (LAPACK) diagonalization algorithms: an expensive tridiagonal (or equivalent) step, costing $\mathcal{O}(N_h^3)$ floating-point operations, has to be performed anyway. It is possible to take advantage of the smallness of the number of desired eigenvalues, by using *iterative* diagonalization algorithms. Unlike conventional algorithms, they are based on successive refinement steps of a trial solution, until the required accuracy is reached. If an approximate solution is known, the convergence may be very quick. Iterative diagonalization algorithms typically use as basic ingredients $H\psi$, where ψ is the trial solution. Such operations, in practice matrix-vector products, require $\mathcal{O}(N_h^2)$ floating-point operations. Sparseness can however be exploited to speed up the calculation of $H\psi$ products. In some cases, the special structure of the matrix can also be exploited (this is the case for one-electron Hamiltonians in a plane-wave basis set). It is not just a problem of speed but of storage: even if we manage to store into memory vectors of length N_h, storage of a $N_h \times N_h$ matrix is impossible.

Among the many algorithms and variants, described in many thick books, a time-honored one that stands out for its simplicity is the *Lanczos* algorithm. Starting from $|v_0\rangle = 0$ and from some initial guess $|v_1\rangle$, normalized, we generate the following chain of vectors:

$$|w_{j+1}\rangle = H|v_j\rangle - \alpha_j|v_j\rangle - \beta_j|v_{j-1}\rangle, \qquad |v_{j+1}\rangle = \frac{1}{\beta_{j+1}}|w_{j+1}\rangle, \tag{10.7}$$

where

$$\alpha_j = \langle v_j|H|v_j\rangle, \qquad \beta_{j+1} = \left(\langle w_{j+1}|w_{j+1}\rangle\right)^{1/2}. \tag{10.8}$$

The first condition in Eq.(10.8) enforces orthogonality of $|w_{j+1}\rangle$ to $|v_j\rangle$; the second, to vector $|v_{j-1}\rangle$. In fact, $\langle v_{j-1}|w_{j+1}\rangle = 0$ implies

$$\beta_j = \langle v_{j-1}|H|v_j\rangle, \tag{10.9}$$

but from Eq.(10.7) we have that $\langle v_{j-1}|H = \langle w_j| + \alpha_{j-1}\langle v_{j-1}| + \beta_{j-1}\langle v_{j-2}|$, demonstrating the equivalence of Eq.(10.9) with the definition of β_j in Eq.(10.8).

A remarkable property of Lanczos chains is that each new vector of the chain is automatically orthogonal to *all* previous vectors, not just to the previous two. This can be demonstrated by induction. Let us consider $\langle w_{j+1}|v_i\rangle = (\langle v_j|H - \alpha_j\langle v_j| - \beta_j\langle v_{j-1}|)|v_i\rangle$ for $i < j-1$. If we assume that $\langle v_j|v_i\rangle = 0$ for all $i < j$, we have $\langle w_{j+1}|v_i\rangle = \langle v_j|H|v_i\rangle$. Since however $H|v_i\rangle$ has components only on $|v_{i+1}\rangle$, $|v_i\rangle$ and $|v_{i-1}\rangle$ (see Eq.(10.7)), $\langle v_j|H|v_i\rangle = 0$ for all $i < j-1$.

In summary, vectors $|v_j\rangle$ form an orthonormal set: $\langle v_i|v_j\rangle = \delta_{ij}$. In the basis of the $|v_j\rangle$ vectors, the Hamiltonian has a tridiagonal form (see Eqs.(10.8) and (10.9)), with α_j elements on the diagonal, β_j on the subdiagonal. After n steps:

$$H_t = \begin{pmatrix} \alpha_1 & \beta_2 & 0 & \dots & 0 \\ \beta_2 & \alpha_2 & \beta_3 & 0 & \vdots \\ 0 & \beta_3 & \alpha_3 & \ddots & 0 \\ \vdots & 0 & \ddots & \ddots & \beta_n \\ 0 & \dots & 0 & \beta_n & \alpha_n \end{pmatrix}. \tag{10.10}$$

If $n = N_h$, this transformation becomes exact: $H_t = H$, and constitutes an alternative tridiagonalization algorithm. In practice, the Lanczos recursion tends to be unstable and may lead to loss of orthogonality between states. If however we limit to a few steps, we observe that the lowest eigenvalues, and especially the lowest one, of matrix H_t converge very quickly to the corresponding ones of H. Since the diagonalization of a tridiagonal matrix is a very quick and easy operation, this procedure gives us a convenient numerical framework for finding a few low-lying states of large matrices. If moreover it is possible to exploit sparseness (or other properties of the matrix) to quickly calculate $H|v\rangle$ products *without storing the entire matrix*, the advantage over conventional diagonalization becomes immense.

10.4 Code: `heisenberg_exact`

Code `heisenberg_exact.f90`[2] (or `heisenberg_exact.c`[3]) finds the ground state energy of the 1-dimensional Heisenberg model, using Periodic Boundary Conditions:

$$H = -J\sum_{i=1}^{N} \mathbf{S}(i)\cdot\mathbf{S}(i+1), \quad \mathbf{S}(N+1) = \mathbf{S}(1). \tag{10.11}$$

In the code, energies are in units of $|J|$, spins are adimensional. If $J > 0$ a ferromagnetic ground state, with all spins oriented along the same direction, will be favored, while the $J < 0$ case will favor an antiferromagnetic ordering. The sign of J is set in the code (to change it, edit the code and recompile).

For the totally magnetized (ferromagnetic) case, the solution is trivial: there is just one state with all spins up (let us call it $|F\rangle$), yielding $E_0 = \langle F|H|F\rangle =$

[2] http://www.fisica.uniud.it/%7Egiannozz/Didattica/MQ/Software/F90/heisenberg_exact.f90
[3] http://www.fisica.uniud.it/%7Egiannozz/Didattica/MQ/Software/C/heisenberg_exact.c

$-NJ/4$. Also the case with $N-1$ spins up can be exactly solved. We have N states with $N-1$ spins up, that we label as $|n\rangle = S_-(n)|F\rangle$. Exploiting translational symmetry, one introduces Bloch-like states

$$|k\rangle = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} e^{ikn}|n\rangle, \qquad k = 2\pi m/N, \quad m = 0, ..., N-1. \tag{10.12}$$

It can then be shown that these are the eigenvectors of H with eigenvalues $E(k) = E_0 + J(1 - \cos k)$. Careful readers will recognize *spin waves* in this solution.

In the antiferromagnetic case, the ground state has $S_z = 0$ for even $S_z = 1/2$ for odd number of spins. In the limit of infinite chains, the ground-state energy is known: $E_0 = -NJ(\log 2 - \frac{1}{4})$, and the gap between E_0 and the first excited state E_1 decreases as $1/N$. A general exact solution (for other similar spin problems as well) can be found by means of the *Bethe Ansatz*, a highly nontrivial technique.

The code requires the number `N` of spins and the number `nup` of up spins, computes the dimension `nhil` of the Hilbert space. It then proceeds to the labelling of the states, using a trick often employed for spin-1/2 systems: an integer index `k`, running from 1 to 2^N-1, contains in its $i-$th bit the information (0=down, 1=up) for the $i-$th spin. Of course this works only up to 32 spins, for default integers (or 64 spins for `INTEGER(8)`). The integer `k` is stored into array `states` for the states in the required Hilbert space.

The Hamiltonian matrix is then filled (the code does not takes advantage of sparseness) and the number of nonzero matrix elements counted. For the S_+S_- and S_-S_+ terms in the Hamiltonian, only matrix elements as in 10.5 and 10.6, respectively, are calculated. We remark that the line

```
k = states(ii)+2**(j-1)-2**(i-1)
```

is a quick-and-dirty way to calculate the index for the state obtained by flipping down spin `i` and flipping up spin `j` in state `states(ii)`.[4]

We then proceed to the generation of the Lanczos chain. The number `nl` of chain steps (should not exceed `nhil`) is prompted for and read from terminal. The starting vector is filled with random numbers. Note the new BLAS routines `dnrm2` and `dgemv`: the former calculates the module of a vector, the latter a matrix-vector product and is used to calculate $H|v\rangle$.

The Hamiltonian in tridiagonal form (contained in the two arrays `d` and `e`) is then diagonalized by the LAPACK routine `dsterf`, that actually finds only the eigenvalues. The lowest eigenvalues is then printed for increasing values of the dimension of the tridiagonal matrix, up to `nl`, so that the convergence of the Lanczos chain can be estimated. You can modify the code to print more eigenvalues.

As a final check, the matrix is diagonalized using the conventional algorithm (routine `dspev`). Note how much slower this final step is than the rest of the

[4] A more elegant but hardly more transparent way would be to directly manipulate the corresponding bits.

calculation! Once you are confident that the Lanczos chain works, you can speed up the calculations by commenting out the exact diagonalization step. The limiting factor will become the size of the Hamiltonian matrix.

10.4.1 Computer Laboratory

- Examine the convergence of the Lanczos procedure to the ground-state energy. Examine excited states as well: notice that they will invariably "fall down" towards lower-energy states after a sufficiently large number of Lanczos steps is performed, due to loss of orthogonality.

- For the antiferromagnetic case, verify that the ground state has zero magnetization (for even N) or magnetization $1/2$ (for odd N). Plot the ground-state energy E_0 and the first excited state E_1 (where can you find it? why?) as a function of N, try to verify if the gap $E_1 - E_0$ has a $1/N$ dependence. Note that all energies are at least doubly degenerate for odd N, as a consequence of time-reversal symmetry and Kramer's theorem.

- For the ferromagnetic case, verify that the ground state has all spins aligned. Note that the ground state will have the same energy no matter which total magnetization you choose! This is a consequence of the rotational invariance of the Heisenberg Hamiltonian. Verify that the case with $N - 1$ spins up corresponds to the spin-wave solution, Eq.(10.12). You will need to print all eigenvalues.

Possible code extensions:

- Modify the code in such a way that open boundary conditions (that is: the system is a finite chain) are used instead of periodic boundary conditions. You may find the following data useful to verify your results: $E/N = -0.375$ for $N = 2$, $E/N = -1/3$ for $N = 3$, $E/N = -0.404$ per $N = 4$, $E/N \to -0.44325$ per $N \to \infty$

- Modify the code in such a way that the entire Hamiltonian matrix is no longer stored. There are two possible ways to do this:

  - Calculate the $H\psi$ product "on the fly", without ever storing the matrix;
  - Store only nonzero matrix elements, plus an index keeping track of their position.

  Of course, you cannot any longer use `dspev` to diagonalize the Hamiltonian. Note that diagonalization packages for sparse matrices, such as ARPACK, exist.

Chapter 11

Density-Functional Theory

Density-Functional Theory (DFT) provides an effective alternative to the paradigm of "Hartree-Fock plus corrections". Unlike the latter, or more sophisticated MBPT (Many-Body Perturbation Theory) methods, DFT is not based on the wave function: it focuses instead on the *charge density* as the fundamental quantity. Atfer a slow start DFT has become widespread in electronic-structure calculations for materials, especially in the implementation based on plane waves and pseudopotentials.

We will see here in action a basic DFT self-consistent code, using a simple form (Applebaum-Hamann) of atomic pseudopotentials for Si.

11.1 Hohenberg-Kohn theorem

Density-Functional Theory (DFT) is based on the Hohenberg-Kohn theorem (1964). This states that the ground-state charge density, $\rho(\mathbf{r})$, defined as

$$\rho(\mathbf{r}) = N \int |\Psi(\mathbf{r}, \mathbf{r}_2, ..., \mathbf{r}_N)|^2 d^3r_2 ... d^3r_N \tag{11.1}$$

for a system of N electrons[1] with ground-state many-electron wave-function Ψ, uniquely determines the external (e.g. nuclear) potential V acting on electrons, and as a consequence the many-body Hamiltonian: $H = T + V + U$, where T is the kinetic energy, U is the electron-electron repulsion.

While it is quite obvious that V determines ρ, the opposite is much less obvious. The Hohenberg-Kohn theorem demonstrates just this, by showing that no two potentials V and V' can have the same ρ as ground-state charge density (unless they differ by a constant).

A straightforward consequence of the Hohenberg-Kohn theorem is that the energy of the ground state is a *functional* of the charge density:

$$E = \langle \Psi | H | \Psi \rangle = \langle \Psi | T + V + U | \Psi \rangle = F[\rho] + \int \rho(\mathbf{r}) v(\mathbf{r}) d^3r \tag{11.2}$$

where $F[\rho] = \langle \Psi | T + U | \Psi \rangle$ is a *universal* functional, i.e. independent upon the external potential V, and we have assumed that the potential V acts locally on the electrons: $V \equiv \sum_i v(\mathbf{r}_i)$.

[1] Note that the original DFT applies to *spinless* fermions

A further consequence is that the energy functional $E[\rho]$ is *minimized* by the ground-state charge density. This suggests a very interesting algorithm to find the ground state: finding a three-dimensional function that minimizes the energy functional is a much easier task than finding the $3N-$dimensional wave function that minimizes the same quantity. Unfortunately it is not easy to minimize an unknown functional, since all we know at this stage is that it exists.[2]

11.2 Kohn-Sham equations

The transformation of the Hohenberg-Kohn theorem from a curiosity into a useful tool takes place via the Kohn-Sham (KS) approach and via a simple approximation known as Local-Density Approximation, LDA (1965). Since the Hohenberg-Kohn theorem holds irrespective of the electron-electron repulsion U, we may introduce an auxiliary system of *non-interacting* electrons having the same density as the true system:

$$\rho(\mathbf{r}) = \sum_i |\psi_i(\mathbf{r})|^2, \tag{11.3}$$

where the ψ_i (*Kohn-Sham orbitals*) are single-electron wavefunctions, to be determined by the condition that $E[\rho]$ is minimized, under orthonormality constraints $\langle\psi_i|\psi_j\rangle = \delta_{ij}$. We still do not know the functional to be minimized, but let us write it as a sum of known terms, large and easy to compute, and the rest:

$$E = T_s[\rho] + E_H[\rho] + \int \rho(\mathbf{r})v(\mathbf{r})d^3r + E_{xc}[\rho], \tag{11.4}$$

where T_s is the kinetic energy of the non-interacting electrons:

$$T_s = -\frac{\hbar^2}{2m}\sum_i \int \psi_i^*(\mathbf{r})\nabla^2\psi_i(\mathbf{r})d^3r, \tag{11.5}$$

(note that in general $T_s \neq \langle\Psi|T|\Psi\rangle$), E_H is the electrostatic (Hartree) energy:

$$E_H[\rho] = \frac{q_e^2}{2}\int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d^3r d^3r', \tag{11.6}$$

the third term is the interaction energy with the external potential, and all the rest is hidden into the E_{xc} term. The latter is known as *exchange-correlation energy*, for historical reasons coming from Hartree-Fock terminology: in principle, E_{xc} contains *both* the exchange energy of the Hartree-Fock method, *and* the correlation energy that is missing in it.

By imposing the condition that the KS orbitals ψ_i minimize the energy, we find the *Kohn-Sham equations* to which KS orbitals obey:

$$\left(-\frac{\hbar^2\nabla^2}{2m} + V_{KS}(\mathbf{r})\right)\psi_i(\mathbf{r}) = \epsilon_i\psi_i(\mathbf{r}), \tag{11.7}$$

[2]We actually know a lot more than this about the properties of the exact functional $F[\rho]$, but there is no way to write it down explicitly and in a simple form

where the effective, or Kohn-Sham potential, $V_{KS} = v(\mathbf{r}) + V_H(\mathbf{r}) + V_{xc}(\mathbf{r})$, is a functional of the charge density:

$$V_H(\mathbf{r}) = q_e^2 \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d^3r', \qquad V_{xc}(\mathbf{r}) = \frac{\delta E_{xc}[\rho]}{\delta \rho(\mathbf{r})}. \tag{11.8}$$

KS equations are reminiscent of Hartree-Fock equations, Eq.(6.19), with the exchange potential replaced by the exchange-correlation potential. Note that the latter is a *local* potential, while the exchange potential is non-local. The energy can be rewritten using the sum over KS eigenvalues ϵ_i. It is straightforward to prove that the energy functional can be written as

$$E = \sum_i \epsilon_i - E_H[\rho] + E_{xc} - \int \rho(\mathbf{r}) V_{xc}(\mathbf{r}) d^3r. \tag{11.9}$$

11.3 Approximated functionals

Not much progress seems to be done yet: E_{xc} is still an unknown functional, and so is V_{xc}. There is however a long tradition, pre-dating DFT, of using homogeneous electron gas results to approximate similar functions. The most famous historical method is Slater's local approximation to the exchange potential:

$$V_x(\mathbf{r}) \simeq -\frac{3q_e^2}{2\pi} (3\pi^2 \rho(\mathbf{r}))^{(1/3)}. \tag{11.10}$$

Kohn and Sham extend and improve upon such ideas by introducing the *local density approximation* (LDA): they re-write the the energy functional as

$$E_{xc} = \int \rho(\mathbf{r}) e_{xc}(\mathbf{r}) d^3r, \tag{11.11}$$

using for the exchange-correlation energy density $e_{xc}(\mathbf{r})$ the result for the homogeneous electron gas of density n, $\epsilon_{xc}(n)$, computed in each point *at the local charge density*: $e_{xc}(\mathbf{r}) \equiv \epsilon_{xc}(\rho(\mathbf{r}))$. The function $\epsilon_{xc}(n)$ can be computed with high accuracy and fitted a some simple analytical form, as e.g. in the following parameterization (Perdew-Zunger) of the Monte-Carlo results by Ceperley and Alder. In Ry atomic units:

$$\begin{aligned} \epsilon_{xc}(n) = &\quad -\tfrac{0.9164}{r_s} - \tfrac{0.2846}{(1+1.0529\sqrt{r_s}+0.3334 r_s)}, && r_s \geq 1 \\ = &\quad -\tfrac{0.9164}{r_s} - 0.0960 + 0.0622 \ln r_s - 0.0232 r_s + 0.0040 r_s \ln r_s, && r_s \leq 1. \end{aligned} \tag{11.12}$$

Here $r_s = (3/4\pi n)^{1/3}$, a parameter traditionally used in the theory of metals. One recognizes in the first term Hartree-Fock exchange energy, so the remaining terms are refereed to as "correlation". The exchange-correlation potential can be computed as functional derivative of the energy, Eq.(11.8), that in this case reduces to simple derivation:

$$V_{xc}(\mathbf{r}) = \left(\epsilon_{xc}(n) + \rho \frac{d\epsilon_{xc}(n)}{dn} \right)_{n=\rho(\mathbf{r})}. \tag{11.13}$$

In spite of its simplicity, and of its derivation from an electron gas model that wouldn't look suitable to describe real, highly inhomogeneous materials, LDA gives surprising good results for several properties (e.g. bond lengths, crystal structures, vibrational frequencies) of a large class of materials (e.g. *sp* bonded semiconductors). It also gives (unsurprisingly) bad results for other properties (e.g. band gap) and for other cases (e.g. transition metals). A number of functionals have been proposed with various degrees of sophistication, extending DFT to a much larger class of properties and of materials. The search for better functionals is currently a very active field of research.

11.4 Structure of a DFT code

The basic[3] algorithm for the solution of the DFT problem, as implemented in code `ah.f90`[4] or `ah.c`[5], consists of a self-consistent loop in which, at iteration n:

1. the KS potential $V_{KS}^{(n)}$ is computed from $\rho^{(n)}$.
2. KS equations are solved, yielding KS orbitals $\psi_i^{(n)}$;
3. the output charge density $\rho_{out}^{(n)}$ is calculated by summing the square of all occupied KS orbitals: $\rho_{out}^{(n)} = \sum_i |\psi_i^{(n)}|^2$;
4. the new charge density $\rho^{(n+1)}$ is computed from a linear combination of previous input $\rho^{(n)}$ and output $\rho_{out}^{(n)}$ charge densities.

The loop is started with some suitable initial choice of the charge density, is iterated until self-consistency is achieved, i.e. $\rho_{out}^{(n)} = \rho^{(n)}$ according to some pre-defined criterion (see Ch.6 for an introduction to self-consistency). At the end of the loop, the DFT energy can be computed, using either Eq.(11.4) or Eq.(11.9). The total energy will be given by the sum of the DFT energy and of the nuclear repulsion energy.

Let us focus on a periodic system with a plane-wave basis set. We consider a simple but nontrivial case: Si crystal using Appelbaum-Hamann (Phys. Rev. **B 8**, 1777 (1973)) pseudopotentials. We use the definitions of lattice, reciprocal lattice, Bloch vector, plane-wave basis set as in code `cohenbergstresser`, introduced in Ch. 9, as well as conventional (LAPACK) diagonalization and the "simple mixing" algorithm of Sec.6.2:

$$\rho^{(n+1)} = \beta\rho_{out}^{(n)} + (1-\beta)\rho^{(n)}, \qquad 0 < \beta \leq 1. \tag{11.14}$$

for achieving self-consistency.[6]

New algorithms in code `ah` deal with the calculation of

[3] But not the only one: it is also possible to directly minimize the energy functional

[4] http://www.fisica.uniud.it/%7Egiannozz/Didattica/MQ/Software/F90/ah.f90

[5] http://www.fisica.uniud.it/%7Egiannozz/Didattica/MQ/Software/C/ah.c

[6] Serious codes use iterative diagonalization, similar to the Lanczos method of Ch. 10, and more sophisticated algorithms for self-consistency.

1. matrix elements of the pseudopotential;
2. charge density from KS orbitals;
3. self-consistent potential from the charge density.

In the following we examine in some detail these cases.

11.4.1 Matrix elements of the potential

In order to compute matrix elements of the Hamiltonian, we need the pseudopotential form factors, Eq.(9.16). This is done in function `form_factor`. The Appelbaum-Hamann pseudopotential for Si is given as the sum of two terms, one long-ranged, one short-ranged. The former is the electrostatic potential v_{lr} generated by a charge density distribution ρ_{at}:

$$v_{lr}(r) = -q_e^2 \int \frac{\rho_{at}(r')}{|\mathbf{r}-\mathbf{r}'|} d^3r', \qquad \rho_{at}(r) = Z_v \left(\frac{\alpha}{\pi}\right)^{3/2} e^{-\alpha r^2}. \tag{11.15}$$

Note that both ρ_{at} and v_{lr} are spherically symmetric, and so are their Fourier transforms. You may want to verify that ρ_{at} integrates to $Z_v = 4$ electrons. Appelbaum-Hamann pseudopotentials describe in fact a Si^{4+} (pseudo-)ion, interacting with the four valence electrons of Si.

The short-ranged potential has the form

$$v_{sr}(r) = e^{-\alpha r^2}(v_1 + v_2 r^2). \tag{11.16}$$

α, v_1, v_2 are adjustable parameters provided in the paper. The form factor of the electrostatic term:

$$\tilde{v}_{lr}(G) = \frac{1}{\Omega}\int v_{lr}(r) e^{-i\mathbf{G}\cdot\mathbf{r}} d^3r = -\frac{q_e^2}{\Omega}\int \left(\int \frac{\rho_{at}(r')}{|\mathbf{r}-\mathbf{r}'|} d^3r'\right) e^{-i\mathbf{G}\cdot\mathbf{r}} d^3r \tag{11.17}$$

can be computed by rearranging the integral:

$$\tilde{v}_{lr}(G) = -\frac{q_e^2}{\Omega}\int \left(\int \frac{1}{|\mathbf{r}-\mathbf{r}'|} e^{-i\mathbf{G}\cdot(\mathbf{r}-\mathbf{r}')} d^3r\right) \rho_{at}(r') e^{-i\mathbf{G}\cdot\mathbf{r}'} d^3r'. \tag{11.18}$$

The integral between brakets can be brought by a change of variable to the following known result:

$$\int \frac{1}{r} e^{-i\mathbf{q}\cdot\mathbf{r}} d^3r = \frac{4\pi}{q^2}. \tag{11.19}$$

Finally:

$$\tilde{v}_{lr}(G) = -4\pi q_e^2 \frac{\tilde{\rho}_{at}(G)}{G^2}, \quad \tilde{\rho}_{at}(G) = \frac{1}{\Omega}\int \rho_{at}(r) e^{-i\mathbf{G}\cdot\mathbf{r}} d^3r. \tag{11.20}$$

The equation on the left has a general validity: it is the solution of the Poisson equation in Fourier space and is used also to compute the Hartree potential from the charge. The Fourier transform of a Gaussian is known (see Eq.(8.16)). One finally finds

$$\tilde{v}_{lr}(G) = -4\pi Z_v q_e^2 \frac{e^{-G^2/4\alpha}}{G^2}, \tag{11.21}$$

and for the short-range term:

$$\tilde{v}_{sr}(G) = q_e^2 \left(\frac{\pi}{\alpha}\right)^{3/2} \left(v_1 + \frac{v_2}{\alpha}\left(\frac{3}{2} - \frac{G^2}{4\alpha}\right)\right) e^{-G^2/4\alpha}. \tag{11.22}$$

The careful reader will notice that the $G = 0$ term diverges as $-4\pi Z_v q_e^2/G^2$. This is due to the presence of the long-range Coulomb term. The divergence, however, cancels out, at least in neutral systems, with the divergence of opposite sign coming from the Hartree potential of electrons. The $G = 0$ term can be evaluated in practice [7] by taking the $G \to 0$ limit and throwing away the divergent term $-4\pi Z_v q_e^2/G^2$.

11.4.2 FFT and FFT grids

The three-dimension generalization of the Discrete Fourier-Transform algorithm introduced in Sec. 8.1.5 is needed. We define a real-space grid of points $\mathbf{r}_{jmn}$, spanning the unit cell, as

$$\mathbf{r}_{jmn} = \frac{j}{n_1}\mathbf{a}_1 + \frac{m}{n_2}\mathbf{a}_2 + \frac{n}{n_3}\mathbf{a}_3, \tag{11.23}$$

where the integer indices j, m, n run from 0 to n_1-1, n_2-1, n_3-1, respectively; and a corresponding reciprocal-space grid of vectors $\mathbf{G}_{hkl}$, as

$$\mathbf{G}_{hkl} = h\mathbf{b}_1 + k\mathbf{b}_2 + l\mathbf{b}_3, \tag{11.24}$$

where the integer indices h, k, l run from 0 to $n_1 - 1, n_2 - 1, n_3 - 1$ and are known as *Miller's indices*. These are stored into array `mill`, while array `indg` returns the index of the $\mathbf{G}$ vector as a function of Miller indices. The factors n_1, n_2, n_3 are chosen big enough to include all Fourier components (see next section). These grids are referred to as "FFT grids".

It can be easily verified that the discretized version of the three-dimensional Fourier Transform:

$$\tilde{F}(\mathbf{G}) = \frac{1}{\Omega}\int F(\mathbf{r}) e^{-i\mathbf{G}\cdot\mathbf{r}} d^3r, \tag{11.25}$$

where Ω is the volume of the unit cell, can be written as follows:

$$\tilde{F}_{hkl} = \frac{1}{n_1 n_2 n_3} \sum_{j=0}^{n_1-1} e^{-2\pi i hj/n_1} \sum_{m=0}^{n_2-1} e^{-2\pi i km/n_2} \sum_{n=0}^{n_3-1} e^{-2\pi i ln/n_3} F_{jmn}. \tag{11.26}$$

while the corresponding inverse transform:

$$F(\mathbf{r}) = \sum_{\mathbf{G}} \tilde{F}(\mathbf{G}) e^{i\mathbf{G}\cdot\mathbf{r}} \tag{11.27}$$

(valid for periodic functions) can be written as

$$F_{jmn} = \sum_{h=0}^{n_1-1} e^{2\pi i hj/n_1} \sum_{k=0}^{n_2-1} e^{2\pi i km/n_2} \sum_{l=0}^{n_3-1} e^{2\pi i ln/n_3} \tilde{F}_{hkl}. \tag{11.28}$$

[7] Handle with care! there are many subtleties about divergent terms in periodic system, but they are beyond the scope of these lecture notes

Routine `fft3d` implements Eq.(11.28) or Eq.(11.26) (with the factor $1/n_1n_2n_3$) if the first argument (the sign of the exponential) has value +1 or -1, respectively. Note that Eq.(11.26) is also called "forward" FT, Eq.(11.28) "backward" FT. Note that as in Sec.8.1.5, both the real- and the reciprocal-space grids are periodic, so $\mathbf{G}-$vectors with negative indices hkl appear "at the other end of the box". Also note that the "inverse" transform is really the inverse: if you apply a FT to a function and then the inverse FT, or vice versa, you get exactly the starting function.

Why is FFT important? because it allows to quickly jump between real to reciprocal space, performing the required operations in the space where it is more convenient. Such "dual-space" technique is fundamental in modern DFT codes based on plane waves to achieve high performances in terms of speed.

11.4.3 Computing the charge density

The calculation of the charge density requires a sum (actually, an integral) over an infinitely dense set of Bloch vectors (or "k-points") covering the entire Brillouin Zone. This apparently hopeless task can in fact be accomplished by approximating the integral with a discrete sum over a finite grid of Bloch vectors. For insulators and semiconductors, quite accurate results can be obtained with relatively coarse grids of points. This method is often referred to as "special points technique". In our sample code, we use a really "special" k-point:

$$\mathbf{k_0} = \frac{2\pi}{a_0}(0.6223, 0.2953, 0) \tag{11.29}$$

fully exploiting the fcc lattice symmetry, also known as *mean-value point*[8]. The calculation of the charge density reduces to

$$\rho(\mathbf{r}) = \sum_{\nu=1}^{4} |\psi_{\mathbf{k_0},\nu}(\mathbf{r})|^2. \tag{11.30}$$

In spite of its simplicity, this approximation is remarkably good. Since the expensive part of the calculation is typically the diagonalization of the KS Hamiltonian, that must be done for each k-point, this choice reduces the computational burden to the strict minimum.

The actual calculation of the charge density is performed in real space using FFT's (see subroutine `sum_charge`), but it is instructive to look at how the same calculation would appear if performed in reciprocal space. With straightforward algebra:

$$\rho(\mathbf{G}) = \sum_{\nu,\mathbf{k}} \sum_{\mathbf{G'}} \psi_{\mathbf{k}_0,\nu}(\mathbf{G} - \mathbf{G'})\psi_{\mathbf{k}_0,\nu}(\mathbf{G'}). \tag{11.31}$$

As a consequence is that the largest $\mathbf{G}$-vector appearing in $\rho(\mathbf{G})$ has modulus twice as large as the largest $\mathbf{G}$-vector appearing in $\psi(\mathbf{G})$. This gives a prescription to choose the n_1, n_2, n_3 factors defining the FFT grid: they must be large enough to accommodate $\mathbf{G}$-vectors up to a maximum cutoff $E_\rho = 4E_{cut}$, where E_{cut} is the cutoff for the plane waves basis set, Eq.(9.7). This choice guarantees that no Fourier components are "lost" when computing the charge density.

[8] A. Baldereschi, Phys. Rev. B **7**, 5212 (1973)

11.4.4 Computing the potential

The self-consistent potential appearing in the Kohn-Sham equations, Eq.(11.8), consists of two terms, the Hartree and the exchange-correlation term. Pseudopotentials are build to work in conjunction with a specific kind of exchange-correlation functional. Appelbaum-Hamann pseudopotentials work together with "Slater exchange", Eq.(11.10), one of the simplest and less accurate functionals.

The self-consistent potential is computed in subroutine `v_of_rho`.

Calculation of XC potential The exchange-correlation potential (exchange only in our case) can be convenienty and directly computed on the real-space grid. In the following few lines, `rho` is the charge density, `vr` the potential $V_{xc}(\mathbf{r})$:

```
do n3=1,nr3
   do n2=1,nr2
      do n1=1,nr1
         vr(n1,n2,n3) = - alphax * 3.0_dp / 2.0_dp * e2 * &
                ( 3.0_dp*rho(n1,n2,n3)/pi )**(1.0_dp/3.0_dp)
      end do
   end do
end do
```

where `alphax` $= 2/3$ is the adjustable parameter of the so-called $X\alpha$ method. Since we need $V(\mathbf{G})$ to fill the Hamiltonian matrix, we Fourier-transform the potential, storing it in vector `vg`.

Hartree potential The Hartree potential can be conveniently computed in reciprocal space. Straightforward algebra shows that $V_H(\mathbf{G})$ can be written as

$$V_H(\mathbf{G}) = 4\pi q_e^2 \frac{\rho(\mathbf{G})}{G^2}. \tag{11.32}$$

This is nothing but the solution of the Poisson equation in Fourier space. The diverging $G = 0$ term is compensated (in neutral systems) by the same term coming from pseudopotentials. The lines

```
do ng =1, ngm
   if ( g2(ng) > eps ) vg(ng) = vg(ng) + fpi*e2*rhog(ng)/g2(ng)
end do
```

add, for $G \neq 0$, to `vg` the Hartree potential directly computed from `rhog` (charge in reciprocal space) and `g2`, square modulus of G.

11.4.5 Laboratory

In order to compile the `ah` code, you will need routines from the BLAS, LAPACK and FFTW3 libraries. You may follow the same procedure as for compilation of code `testfft` code, Ch.8.

- Verify which cutoff is needed to achieve converged results for $E(\mathbf{k})$.
- Compute the band structure. To this end you will need compute and store the final self-consistent potential, then to perform a non-self-consistent calculation with the previously computed self-consistent potential, on a suitably chosen set of k-points.
- Examine how the band structure, and especially the band gap, changes by changing `alphax` from 2/3 to 1.
- Plot the charge density along the (111) direction, along the bond between two atoms, and on the (110) plane.
- Examine how the charge density changes if a denser grid of k-points is used instead of the "mean-value point" in the self-consistent calculation. A suitable grid can be generated as follows:

$$\mathbf{k}_n = \frac{n_1}{N_1}\mathbf{b}_1 + \frac{n_2}{N_2}\mathbf{b}_2 + \frac{n_3}{N_3}\mathbf{b}_3, \tag{11.33}$$

where $n_1 = 0, 1, ..., N_1 - 1$ and so forth. You may start from $N_1 = N_2 = N_3 = 2$. Notice that $\mathbf{k}$ and $-\mathbf{k}$ are equivalent: you may use one of the two with double weight.

Appendix A

Real-space two- and three-dimensional grids

Let us consider the one-dimensional time-independent Schrödinger equation, with the simplest discretization of the Laplacian operator for a uniform grid of N points x_i, as in Eq.(1.27):

$$\frac{d^2\psi}{dx^2} \equiv \frac{\psi_{i-1} + \psi_{i+1} - 2\psi_i}{(\Delta x)^2}, \qquad \psi_i \equiv \psi(x_i, t). \tag{A.1}$$

where ψ is the vector formed by the N values ψ_i. The resulting discretized Schrödinger equation:

$$-\frac{\hbar^2}{2m}\frac{\psi_{i-1} + \psi_{i+1} - 2\psi_i}{(\Delta x)^2} + V_i\psi_i = E\psi_i, \qquad V_i \equiv V(x_i) \tag{A.2}$$

can be recast under the form of a matrix equation:

$$\sum_{j=1}^{N} H_{ij}\psi_j = E\psi_i, \quad i = 1, N \tag{A.3}$$

where the $N \times N$ matrix H is *tridiagonal*:

$$H = \begin{pmatrix} H_{1,1} & H_{1,2} & 0 & \dots & 0 \\ H_{2,1} & H_{2,2} & H_{2,3} & 0 & \vdots \\ 0 & H_{3,2} & H_{3,3} & \ddots & 0 \\ \vdots & 0 & \ddots & \ddots & H_{N-1,N} \\ 0 & \dots & 0 & H_{N,N-1} & H_{N,N} \end{pmatrix} \tag{A.4}$$

and the nonzero terms are given by

$$H_{i,i} = \frac{\hbar^2}{m(\Delta x)^2} + V_i, \quad H_{i,i+1} = H_{i+1,i} = -\frac{\hbar^2}{2m(\Delta x)^2}. \tag{A.5}$$

The resemblance with the secular equation as obtained from the variational principle, Ch.4, is quite obvious, and can be made more explicit by introducing

a set of "basis functions" $b_i(x)$ so defined:

$$b_i(x) = \frac{1}{\sqrt{\Delta x}} \qquad x_i - \frac{\Delta x}{2} < x < x_i + \frac{\Delta x}{2} \tag{A.6}$$

$$b_i(x) = 0 \qquad x < x_i - \frac{\Delta x}{2}, x > x_i + \frac{\Delta x}{2}. \tag{A.7}$$

It is straightforward to verify that these functions are orthonormal. The H_{ij} elements are the "matrix elements" of the Hamiltonian. The potential is represented by its value in the grid point: $V_{ii} = \langle b_i|V|b_i\rangle \simeq V(x_i)$ and is thus *diagonal* in this basis. The kinetic energy is less obvious: our basis functions are not differentiable, but the matrix elements for the kinetic energy:

$$T_{ij} = \langle b_i|T|b_j\rangle = -\frac{\hbar^2}{2m}\int b_i(x)\frac{d^2}{dx^2}b_j(x)dx \tag{A.8}$$

can be specified via the discretized form of the second derivative, as in Eq.(A.1).

The solution of the Schrödinger equation via the secular equation has no real advantage with respect to the numerical integration procedure of Ch.1–3 in one dimension, but it can be generalized to more dimensions in a straightforward way: for instance, we can introduce a uniform grid in the following way:

$$\vec{r}_{i,j,k} = (x_i, y_j, z_k), \qquad x_i = i\Delta x, \quad y_j = j\Delta y, \quad z_k = k\Delta z. \tag{A.9}$$

Assuming to solve our problem in a parallelepiped spanned by N_1, N_2, N_3 points along x, y, z respectively, the total number of grid points, $N = N_1 \cdot N_2 \cdot N_3$, defines the "computational complexity" of our problem. Obviously, this may quickly become intractable, but we should not forget that the Hamiltonian matrix is *sparse*: the potential is present only in the diagonal terms, while the kinetic energy has nonzero matrix elements only for a few points close to the one considered. The discretization of Eq.(A.1) generalizes in three dimensions to

$$\begin{aligned}\nabla^2\psi \equiv & \frac{\psi_{i-1,j,k} + \psi_{i+1,j,k} - 2\psi_{i,j,k}}{(\Delta x)^2} \\ + & \frac{\psi_{i,j-1,k} + \psi_{i,j+1,k} - 2\psi_{i,j,k}}{(\Delta y)^2} \\ + & \frac{\psi_{i,j,k-1} + \psi_{i,j,k+1} - 2\psi_{i,j,k}}{(\Delta z)^2},\end{aligned} \tag{A.10}$$

where $\psi_{i,j,k} = \psi(\vec{r}_{i,j,k})$ with $\vec{r}_{i,j,k}$ as in Eq.(A.9). More accurate, higher-order finite-difference formulae for the Laplacian can be in principle, and are in practice, used, but they all share the same characteristics: they involve only a few points around the one in which we want to calculate the Laplacian. Sparseness is the key to fast solution.

Appendix B

Solution of time-dependent Schrödinger equations

When the potential does not contain any explicit dependence upon time, the solution of time-dependent Schrödinger equation:

$$i\hbar\frac{\partial\psi(x,t)}{\partial t} = H\psi(x,t), \qquad H = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V(x) \tag{B.1}$$

for the one-dimensional case can be obtained via variable separation and written down in terms of eigenvectors and eigenvalues of the time-independent Schrödinger equation, $H\psi_n(x) = E_n\psi_n(x)$, as

$$\psi(x,t) = \sum_n c_n e^{-iE_nt/\hbar}\psi_n(x), \qquad c_n = \int \psi(x,0)\psi_n^*(x)dx. \tag{B.2}$$

The coefficients c_n guarantee that the system at $t = 0$ is in the desired state. One can formally write time evolution via an operator $U(t)$:

$$\psi(x,t) = U(t)\psi_n(x,0), \qquad U(t) = e^{-iHt/\hbar}, \tag{B.3}$$

where the "exponential of an operator" is defined via the series expansion

$$e^A = \sum_{n=0}^{\infty}\frac{A^n}{n!}. \tag{B.4}$$

The time-evolution operator is *unitary*: $U^\dagger(t) = U^{-1}(t)$, and $U(-t) = U^\dagger(t)$ hold. These properties reflect important features of time evolution in quantum mechanics: it is *invariant under time reversal* and *conserves the norm.*

If the potential has an explicit dependence upon the time, variable separation is no longer possible. Analytical solutions of course exist only for a small number of cases. Let us consider the numerical solution of the time-dependent Schrödinger equation. This can be useful also for time-independent potentials, actually, since it may be more convenient to compute the time evolution for a given initial state than to project it over all eigenfunctions.

B.1 Discretization in time: Crank-Nicolson algorithm

Let us consider for simplicity the one-dimensional case. We use the same discretization on a uniform grid as in Appendix A. The right-hand side of the Schrödinger equation:

$$i\hbar\frac{\partial\psi_i}{\partial t} = -\frac{\hbar^2}{2m}\frac{\psi_{i-1}+\psi_{i+1}-2\psi_i}{(\Delta x)^2} + V_i\psi_i, \qquad V_i \equiv V(x_i) \tag{B.5}$$

can be recast into a matrix-vector product:

$$i\hbar\frac{\partial\psi}{\partial t} = H\psi \tag{B.6}$$

where ψ is the vector formed by the N values ψ_i and the tridiagonal $N \times N$ matrix H is given in Eq.(A.4).

Let us now proceed to discretize in time, with "time step" Δt. The simplest algorithm one can think of is *Euler* discretization. One writes for the first derivative:

$$\frac{\partial\psi_i}{\partial t} \simeq \frac{\psi_i(t+\Delta t)-\psi_i(t)}{\Delta t} \tag{B.7}$$

and derives the following equation ("forward" Euler) for the wave function at time $t+\Delta t$, given the wave function at time t:

$$\psi(t+\Delta t) = \psi(t) - \frac{i\Delta t}{\hbar}H\psi(t). \tag{B.8}$$

From wave functions at each time step, one directly obtains wave functions at the following time step: the algorithm is thus *explicit*. Unfortunately it turns out that such algorithm is also numerically unstable: it has solutions that grow exponentially with time, even when there aren't any such solutions in the original equation.

A numerically stable algorithm is instead provided by the "backward" Euler discretization (notice the different time at the right-hand side):

$$\psi(t+\Delta t) = \psi(t) - \frac{i\Delta t}{\hbar}H\psi(t+\Delta t). \tag{B.9}$$

This algorithm is called *implicit* because the wave function at the next time step cannot be simply obtained from the wave function at the previous one (note that the former appear in both sides of the equation). The discretized equations can however be recast into the form of a *linear system*:

$$A\psi(t+\Delta t) \equiv \left(1+\frac{i\Delta t}{\hbar}H\right)\psi(t+\Delta t) = \psi(t). \tag{B.10}$$

Since the matrix A is tridiagonal, its solution can be found very quickly, in order $\mathcal{O}(N)$ operations. The implicit Euler algorithm can be demonstrated to be numerically stable, but it still suffers from a major drawback: it breaks, with an error proportional to the time step, the unitariety of the time evolution.

The better *Crank-Nicolson* algorithm is obtained by combining explicit and implicit Euler discretization:

$$\left(1 + \frac{i\Delta t}{2\hbar} H\right) \psi(t + \Delta t) = \left(1 - \frac{i\Delta t}{2\hbar} H\right) \psi(t). \tag{B.11}$$

Notice the factor 2 dividing the time step: you can obtain the above formula by performing an explicit step, followed by an implicit one, both with time step $\Delta t/2$. This often used algorithm conserves unitariety of the time evolution up to order $\mathcal{O}((\Delta t)^2))$ and is numerically stable.

B.2 Direct discretization of the time evolution operator

As mentioned above, even for time-independent problems there are cases in which the solution via eigenvalues and eigenvectors is not practical, either because too many states are needed, or because their calculation is expensive. In these cases, the direct discretization of the time evolution operator $U(t)$, Eq.(B.3) may provide a convenient alternative.

Let us first decompose the Hamiltonian into kinetic and a potential energy, $H = T + V$. In general, one cannot decompose the exponential of the sum of two noncommuting operators into the product of two exponentials for each operator:

$$e^{(A+B)t} \neq e^{At} e^{Bt}, \qquad [A, B] \neq 0. \tag{B.12}$$

One can however use the *Trotter* formula, stating that

$$\lim_{n\to\infty} \left(e^{(A+B)t/n}\right)^n = \left(e^{At/n} e^{Bt/n}\right)^n. \tag{B.13}$$

For finite n, the error is $\mathcal{O}\left((\Delta t)^2\right)$, where $\Delta t = t/n$. The discretization of the time evolution operator follows naturally:

$$e^{-iHt/\hbar} \simeq \left(U_T(\Delta t) U_V(\Delta t)\right)^n, \tag{B.14}$$

where the right-hand side is a sequence of applications of the time evolution operator, each consisting of the application of a purely potential term, followed by a purely kinetic one,

$$U_T(\Delta t) = e^{-iT\Delta t/\hbar}, \quad U_V(\Delta t) = e^{-iV\Delta t/\hbar}. \tag{B.15}$$

The Trotter approximation would be of limited use without a simple way to calculate the two operators U_T and U_V. The former is hardly a problem, since the potential is a *diagonal* operator in real space (at least for a potential having the usual $V(r)$ form, i.e. a local potential). The application of U_V to a wave function on a real-space grid amounts to a simple multiplication:

$$U_V(\Delta t)\psi_i(t) = e^{-iV_i\Delta t/\hbar}\psi_i(t). \tag{B.16}$$

In real space and in one dimension, the kinetic term is represented by a tridiagonal matrix like the one in Eq.(A.4) (without the potential contribution in

the diagonal). The exponential of such an operator is far from simple. One can however take advantage of the Fast Fourier-Transform (FFT) algorithm to switch back and forth between real and Fourier (or "reciprocal") space. In reciprocal space, the kinetic energy is diagonal and the calculation of $U(t)$ is as simple as for the potential term in real space:

$$U_T(\Delta t)\widetilde{\psi}_j(t) = e^{-i\hbar q_j^2 \Delta t/2m}\widetilde{\psi}_j(t) \tag{B.17}$$

where $\widetilde{\psi}_j(t)$ is the Fourier transform of $\psi_i(t)$ of wave-vector q_j.

Appendix C

Derivation of Van der Waals interaction

The Van der Waals attractive interaction can be described in semiclassical terms as a dipole-induced dipole interaction, where the dipole is produced by a charge fluctuation. A more quantitative and satisfying description requires a quantum-mechanical approach. Let us consider the simplest case: two nuclei, located in $\mathbf{R}_A$ and $\mathbf{R}_B$, and two electrons described by coordinates $\mathbf{r}_1$ and $\mathbf{r}_2$. The Hamiltonian for the system can be written as

$$\begin{aligned} H &= -\frac{\hbar^2}{2m}\nabla_1^2 - \frac{q_e^2}{|\mathbf{r}_1 - \mathbf{R}_A|} - \frac{\hbar^2}{2m}\nabla_2^2 - \frac{q_e^2}{|\mathbf{r}_2 - \mathbf{R}_B|} \\ &\quad - \frac{q_e^2}{|\mathbf{r}_1 - \mathbf{R}_B|} - \frac{q_e^2}{|\mathbf{r}_2 - \mathbf{R}_A|} + \frac{q_e^2}{|\mathbf{r}_1 - \mathbf{r}_2|} + \frac{q_e^2}{|\mathbf{R}_A - \mathbf{R}_B|}, \end{aligned} \tag{C.1}$$

where ∇_i indicates derivation with respect to variable $\mathbf{r}_i$, $i = 1, 2$. Even this "simple" Hamiltonian is a really complex object, whose general solution will be the subject of several chapters of these notes. We shall however concentrate on the limit case of two Hydrogen atoms at a large distance R, with $\mathbf{R} = \mathbf{R}_A - \mathbf{R}_B$. Let us introduce the variables $\mathbf{x}_1 = \mathbf{r}_1 - \mathbf{R}_A$, $\mathbf{x}_2 = \mathbf{r}_2 - \mathbf{R}_B$. In terms of these new variables, we have $H = H_A + H_B + \Delta H(R)$, where H_A (H_B) is the Hamiltonian for a Hydrogen atom located in $\mathbf{R}_A$ ($\mathbf{R}_B$), and ΔH has the role of "perturbing potential":

$$\Delta H = -\frac{q_e^2}{|\mathbf{x}_1 + \mathbf{R}|} - \frac{q_e^2}{|\mathbf{x}_2 - \mathbf{R}|} + \frac{q_e^2}{|\mathbf{x}_1 - \mathbf{x}_2 + \mathbf{R}|} + \frac{q_e^2}{R}. \tag{C.2}$$

Let us expand the perturbation in powers of $1/R$. The following expansion, valid for $R \to \infty$, is useful:

$$\frac{1}{|\mathbf{x} + \mathbf{R}|} \simeq \frac{1}{R} - \frac{\mathbf{R} \cdot \mathbf{x}}{R^3} + \frac{3(\mathbf{R} \cdot \mathbf{x})^2 - x^2 R^2}{R^5}. \tag{C.3}$$

Using such expansion, it can be shown that the lowest nonzero term is

$$\Delta H \simeq \frac{2q_e^2}{R^3}\left(\mathbf{x}_1 \cdot \mathbf{x}_2 - 3\frac{(\mathbf{R} \cdot \mathbf{x}_1)(\mathbf{R} \cdot \mathbf{x}_2)}{R^2}\right). \tag{C.4}$$

The problem can now be solved using perturbation theory. The unperturbed ground-state wave function can be written as a product of $1s$ states centered around each nucleus: $\Phi_0(\mathbf{x}_1, \mathbf{x}_2) = \psi_{1s}(\mathbf{x}_1)\psi_{1s}(\mathbf{x}_2)$ (it should actually be anti-symmetrized but in the limit of large separation it makes no difference). It is straightforward to show that the first-order correction, $\Delta^{(1)}E = \langle\Phi_0|\Delta H|\Phi_0\rangle$, to the energy, vanishes because the ground state is even with respect to both $\mathbf{x}_1$ and $\mathbf{x}_2$. The first nonzero contribution to the energy comes thus from second-order perturbation theory:

$$\Delta^{(2)}E = -\sum_{i>0} \frac{|\langle\Phi_i|\Delta H|\Phi_0\rangle|^2}{E_i - E_0} \tag{C.5}$$

where Φ_i are excited states and E_i the corresponding energies for the unperturbed system. Since $\Delta H \propto R^{-3}$, it follows that $\Delta^{(2)}E = -C_6/R^6$, the well-known behavior of the Van der Waals interaction.[1] The value of the so-called C_6 coefficient can be shown, by inspection of Eq.(C.5), to be related to the product of the polarizabilities of the two atoms.

[1]Note however that at very large distances a correct electrodynamical treatment yields a different behavior: $\Delta E \propto -1/R^7$. In practice such asymptotic behavior becomes dominant when in is too small to be of any relevance.

Appendix D

The Helium atom

D.1 Perturbative treatment for Helium atom

The Helium atom is characterized by a Hamiltonian operator

$$H = -\frac{\hbar^2\nabla_1^2}{2m_e} - \frac{Zq_e^2}{r_1} - \frac{\hbar^2\nabla_2^2}{2m_e} - \frac{Zq_e^2}{r_2} + \frac{q_e^2}{r_{12}} \tag{D.1}$$

where $r_{12} = |\mathbf{r}_2 - \mathbf{r}_1|$ is the distance between the two electrons. The last term corresponds to the Coulomb repulsion between the two electrons and makes the problem non separable.

As a first approximation, let us consider the interaction between electrons:

$$V = \frac{q_e^2}{r_{12}} \tag{D.2}$$

as a perturbation to the problem described by

$$H_0 = -\frac{\hbar^2\nabla_1^2}{2m_e} - \frac{Zq_e^2}{r_1} - \frac{\hbar^2\nabla_2^2}{2m_e} - \frac{Zq_e^2}{r_2} \tag{D.3}$$

which is easy to solve since it is separable into two independent problems of a single electron under a central Coulomb field, i.e. a Hydrogen-like problem with nucleus charge $Z = 2$. The ground state for this system is given by the wave function described in Eq.(2.29) ($1s$ orbital):

$$\phi^0(\mathbf{r}_i) = \frac{Z^{3/2}}{\sqrt{\pi}}\mathrm{e}^{-Zr_i} \tag{D.4}$$

in a.u.. We note that we can assign to both electrons the same wave function, as long as their spin is opposite. The total unperturbed wave function (coordinate part) is simply the product

$$\psi^0(\mathbf{r}_1, \mathbf{r}_2) = \frac{Z^3}{\pi}\mathrm{e}^{-Z(r_1+r_2)} \tag{D.5}$$

which is a symmetric function (antisymmetry being provided by the spin part). The energy of the corresponding ground state is the sum of the energies of the two Hydrogen-like atoms:

$$E_0 = -2Z^2\mathrm{Ry} = -8\mathrm{Ry} \tag{D.6}$$

since $Z = 2$. The electron repulsion will necessarily raise this energy, i.e. make it less negative. In first-order perturbation theory,

$$E - E_0 = \langle\psi_0|V|\psi_0\rangle \tag{D.7}$$

$$= \frac{Z^6}{\pi^2}\int \frac{2}{r_{12}} \mathrm{e}^{-2Z(r_1+r_2)} d^3\mathbf{r}_1 d^3\mathbf{r}_2 \tag{D.8}$$

$$= \frac{5}{4} ZRy. \tag{D.9}$$

For $Z = 2$ the correction is equal to 2.5 Ry and yields $E = -8 + 2.5 = -5.5\,\mathrm{Ry}$. The experimental value is $-5.8074\,\mathrm{Ry}$. The perturbative approximation is not accurate but provides a reasonable estimate of the correction, even if the "perturbation", i.e. the Coulomb repulsion between electrons, is of the same order of magnitude of all other interactions. Moreover, he ground state assumed in perturbation theory is usually qualitatively correct: the exact wave function for He will be close to a product of two $1s$ functions.

D.2 Variational treatment for Helium atom

The Helium atom provides a simple example of application of the variational method. The independent-electron solution, Eq.(D.5), is missing the phenomenon of *screening*: each electron will "feel" a nucleus with partially screened charge, due to the presence of the other electron. In order to account for this phenomenon, we may take as our trial wave function an expression like the one of Eq.(D.5), with the true charge of the nucleus Z replaced by an "effective charge" Z_e, presumably smaller than Z. Let us find the optimal Z_e variationally, i.e. by minimizing the energy. We assume

$$\psi(\mathbf{r}_1, \mathbf{r}_2; Z_e) = \frac{Z_e^3}{\pi} \mathrm{e}^{-Z_e(r_1+r_2)} \tag{D.10}$$

and we re-write the Hamiltonian as:

$$H = \left[-\frac{\hbar^2\nabla_1^2}{2m_e} - \frac{Zq_e^2}{r_1} - \frac{\hbar^2\nabla_2^2}{2m_e} - \frac{Zq_e^2}{r_2}\right] + \left[-\frac{(Z-Z_e)q_e^2}{r_1} - \frac{(Z-Z_e)q_e^2}{r_2} + \frac{q_e^2}{r_{12}}\right] \tag{D.11}$$

We now calculate

$$E(Z_e) = \int \psi^*(\mathbf{r}_1, \mathbf{r}_2; Z_e) H \psi(\mathbf{r}_1, \mathbf{r}_2; Z_e)\, d^3\mathbf{r}_1 d^3\mathbf{r}_2 \tag{D.12}$$

The contribution to the energy due to the first square bracket in Eq.(D.11) is $-2Z_e^2$ a.u.: this is in fact a Hydrogen-like problem for a nucleus with charge Z_e, for two non-interacting electrons. By expanding the remaining integrals and using symmetry we find

$$E(Z_e) = -2Z_e^2 - \int |\psi|^2 \frac{4(Z-Z_e)}{r_1} d^3\mathbf{r}_1 d^3\mathbf{r}_2 + \int |\psi|^2 \frac{2}{r_{12}} d^3\mathbf{r}_1 d^3\mathbf{r}_2 \tag{D.13}$$

(in a.u.) with

$$|\psi|^2 = \frac{Z_e^6}{\pi^2} \mathrm{e}^{-2Z_e(r_1+r_2)} \tag{D.14}$$

Integrals can be easily calculated and the result is

$$E(Z_e) = -2Z_e^2 - 4(Z - Z_e)Z_e + 2\frac{5}{8}Z_e = 2Z_e^2 - \frac{27}{4}Z_e \tag{D.15}$$

where we explicitly set $Z = 2$. Minimization of $E(Z_e)$ with respect to Z_e immediately leads to

$$Z_e = \frac{27}{16} = 1.6875 \tag{D.16}$$

and the corresponding energy is

$$E = -\frac{729}{128} = -5.695\,\mathrm{Ry} \tag{D.17}$$

This result is definitely better that the perturbative result $E = -5.50\,\mathrm{Ry}$, even if there is still a non-negligible distance with the experimental result $E = -5.8074\,\mathrm{Ry}$.

It is possible to improve the variational result by extending the set of trial wave functions. Sec.(6.1) shows how to produce the best single-electron functions using the Hartree-Fock method. Even better results can be obtained using trial wave functions that are more complex than a simple product of single-electron functions. For instance, let us consider trial wave functions like

$$\psi(\mathbf{r}_1, \mathbf{r}_2) = [f(\mathbf{r}_1)g(\mathbf{r}_2) + g(\mathbf{r}_1)f(\mathbf{r}_2)]\,, \tag{D.18}$$

where the two single-electron functions, f and g, are Hydrogen-like wave function as in Eq.(D.4) with different values of Z, that we label Z_f and Z_g. By minimizing with respect to the two parameters Z_f and Z_g, one finds $Z_f = 2.183$, $Z_g = 1.188$, and an energy $E = -5.751$ Ry, much closer to the experimental result than for a single effective Z. Note that the two functions are far from being similar!

D.3 Beyond-HF treatment for Helium atom

Let us make no explicit assumption on the form of the ground-state wave function of He. We assume however that the total spin is zero and thus the coordinate part of the wave function is symmetric. The wave function is expanded over a suitable basis set, in this case a symmetrized product of two single-electron gaussians. The lower-energy wave function is found by diagonalization. Such approach is of course possible only for a very small number of electrons.

Code `helium_gauss.f90`[1] (or `helium_gauss.c`[2]) looks for the ground state of the He atom, using an expansion into Gaussian functions, already introduced in the code *hydrogen_gauss*. We assume that the solution is the product of a symmetric coordinate part and of an antisymmetric spin part, with total spin $S = 0$. The coordinate part is expanded into a basis of symmetrized products of gaussians, B_k:

$$\psi(\mathbf{r}_1, \mathbf{r}_2) = \sum_k c_k B_k(\mathbf{r}_1, \mathbf{r}_2). \tag{D.19}$$

[1] http://www.fisica.uniud.it/%7Egiannozz/Didattica/MQ/Software/F90/helium_gauss.f90
[2] http://www.fisica.uniud.it/%7Egiannozz/Didattica/MQ/Software/C/helium_gauss.c

If the b_i functions are S-like gaussians as in Eq.(5.24), we have:

$$B_k(\mathbf{r}_1,\mathbf{r}_2) = \frac{1}{\sqrt{2}}\left(b_{i(k)}(\mathbf{r}_1)b_{j(k)}(\mathbf{r}_2) + b_{i(k)}(\mathbf{r}_2)b_{j(k)}(\mathbf{r}_1)\right) \tag{D.20}$$

where k is an index running over $n(n+1)/2$ pairs $i(k), j(k)$ of gaussian functions. The overlap matrix $\widetilde{S}_{kk'}$ may be written in terms of the S_{ij} overlap matrices, Eq.(5.27), of the hydrogen-like case:

$$\widetilde{S}_{kk'} = \langle B_k|B_{k'}\rangle = \left(S_{ii'}S_{jj'} + S_{ij'}S_{ji'}\right). \tag{D.21}$$

The matrix elements, $\widetilde{H}_{kk'}$, of the Hamiltonian:

$$\widetilde{H}_{kk'} = \langle B_k|H|B_{k'}\rangle, \qquad H = -\frac{\hbar^2\nabla_1^2}{2m_e} - \frac{Zq_e^2}{r_1} - \frac{\hbar^2\nabla_2^2}{2m_e} - \frac{Zq_e^2}{r_2} + \frac{q_e^2}{r_{12}} \tag{D.22}$$

can be written using matrix elements $H_{ij} = H_{ij}^K + H_{ij}^V$, obtained for the hydrogen-like case with $Z = 2$, Eq.(5.28) and (5.29):

$$\widetilde{H}_{kk'} = \left(H_{ii'}S_{jj'} + H_{ij'}S_{ji'} + S_{ii'}H_{jj'} + H_{ij'}S_{ji'}\right) + \langle B_k|V_{ee}|B_{k'}\rangle, \tag{D.23}$$

and the matrix element of the Coulomb electron-electron interaction V_{ee}:

$$\begin{aligned}\langle B_k|V_{ee}|B_{k'}\rangle &= \int b_{i(k)}(\mathbf{r}_1)b_{j(k)}(\mathbf{r}_2)\frac{q_e^2}{r_{12}}b_{i(k')}(\mathbf{r}_1)b_{j(k')}(\mathbf{r}_2)d^3r_1d^3r_2 \\ &+ \int b_{i(k)}(\mathbf{r}_1)b_{j(k)}(\mathbf{r}_2)\frac{q_e^2}{r_{12}}b_{j(k')}(\mathbf{r}_1)b_{i(k')}(\mathbf{r}_2)d^3r_1d^3r_2.\end{aligned} \tag{D.24}$$

These matrix elements can be written, using Eq.(7.25), as

$$\langle B_k|V_{ee}|B_{k'}\rangle = \frac{q_e^2\pi^{5/2}}{\alpha\beta(\alpha+\beta)^{1/2}} + \frac{q_e^2\pi^{5/2}}{\alpha'\beta'(\alpha'+\beta')^{1/2}}, \tag{D.25}$$

where

$$\alpha = \alpha_{i(k)} + \alpha_{i(k')},\ \beta = \alpha_{j(k)} + \alpha_{j(k')},\ \alpha' = \alpha_{i(k)} + \alpha_{j(k')},\ \beta' = \alpha_{j(k)} + \alpha_{i(k')}. \tag{D.26}$$

In an analogous way one can calculate the matrix elements between symmetrized products of gaussians formed with P-type gaussian functions (those defined in Eq.5.25). The combination of P-type gaussians with $L = 0$ has the form:

$$B_k(\mathbf{r}_1,\mathbf{r}_2) = \frac{1}{\sqrt{2}}(\mathbf{r}_1\cdot\mathbf{r}_2)\left(b_{i(k)}(\mathbf{r}_1)b_{j(k)}(\mathbf{r}_2) + b_{i(k)}(\mathbf{r}_2)b_{j(k)}(\mathbf{r}_1)\right) \tag{D.27}$$

It is immediately verified that the product of a S-type and a P-type gaussian yields an odd function that does not contribute to the ground state.

In the case with S-type gaussians only, the code writes to file "gs-wfc.out" the function:

$$P(r_1,r_2) = (4\pi r_1r_2)^2|\psi(r_1,r_2)|^2, \tag{D.28}$$

where $P(r_1, r_2)dr_1dr_2$ is the joint probability to find an electron between r_1 and $r_1 + dr_1$, and an electron between r_2 and $r_2 + dr_2$. The probability to find an electron between r and $r + dr$ is given by $p(r)dr$, with

$$p(r) = 4\pi r^2 \int |\psi(r, r_2)|^2 4\pi r_2^2 dr_2 = \int P(r, r_2) dr_2. \tag{D.29}$$

It is easy to verify that for a wave function composed by a product of two identical functions, like the one in (D.5), the joint probability is the product of single-electron probabilities: $P(r_1, r_2) = p(r_1)p(r_2)$. This is not true in general for the exact wave function.

D.4 Laboratory

- observe the effect of the number of basis functions, to the choice of coefficients λ of the gaussians, to the inclusion of P-type gaussians
- compare the obtained energy with the one obtained by other methods: perturbation theory with hydrogen-like wave functions, (Sec.D.1), variational theory with effective Z (Sec.D.2), exact result (-5.8074 Ry).
- Make a plot of the probability $P(r_1, r_2)$ and of the difference $P(r_1, r_2) - p(r_1)p(r_2)$, using for instance `gnuplot` and the following commands:

  ```
  set view 0, 90
  unset surface
  set contour
  set cntrparam levels auto 10
  splot [0:4][0:4] "gs-wfc.out" u 1:2:3 w l
  ```

  Note that the probability $P(r_1, r_2)$ (column 3 in "splot") is not exactly equal to the product $p(r_1)p(r_2)$ (column 4; column 5 is the difference between the two).

Appendix E

More about pseudopotentials

E.1 An early idea

In one of the early methods to compute the band structure in crystals, a basis set of *orthogonalized plane waves*, or OPW, was introduced:

$$\tilde{b}_{i,\mathbf{k}}(\mathbf{r}) = b_{i,\mathbf{k}}(\mathbf{r}) - \sum_c \phi_c(\mathbf{r})\langle\phi_c|b_{i,\mathbf{k}}\rangle, \tag{E.1}$$

where the $b_{n,\mathbf{k}}$ are plane waves and the ϕ_c are atomic core states[1]. By construction, $\langle\phi_c|\tilde{b}_{i,\mathbf{k}}\rangle = 0$. The solution of the Schrödinger equation for the crystal potential can be obtained by expanding the valence wavefunctions into OPWs:

$$\psi = \sum_j c_j \tilde{b}_{j,\mathbf{k}}(\mathbf{r}) \tag{E.2}$$

One obtains a generalized secular equation (OPWs are not orthonormal) whose solution yields the valence wavefunctions, while core states retain their atomic-like character. Alternatively, one writes the Schrödinger equation $(H-\epsilon)\psi = 0$, projects it over plane waves:

$$\langle b_{i,\mathbf{k}}|(H-\epsilon)|\psi\rangle = 0 \tag{E.3}$$

obtaining

$$\sum_j c_j \langle b_{i,\mathbf{k}}|(H-\epsilon)\left(|b_{j,\mathbf{k}}(\mathbf{r})\rangle - \sum_c |\phi_c\rangle\langle\phi_c|b_{j,\mathbf{k}}\rangle\right) = 0. \tag{E.4}$$

Under the assumption that the crystal potential is well approximated by a sum of atomic-like potentials, and that atomic core states ϕ_c retain their atomic-like character, we may write $H\phi_c \simeq \epsilon_c\phi_c$, where ϵ_c is the eigenvalue corresponding to ϕ_c in the atom. We may thus rewrite Eq.(E.4) as

$$\sum_j c_j \langle b_{i,\mathbf{k}}|\left(H - \epsilon + \sum_c |\phi_c\rangle(\epsilon-\epsilon_c)\langle\phi_c|\right)|b_{j,\mathbf{k}}\rangle = 0. \tag{E.5}$$

[1] more exactly: Bloch sums of atomic core states

Eq.(E.5) is a secular equation for the plane-wave part only, for the effective potential

$$\tilde{V} = V + \sum_c |\phi_c\rangle(\epsilon - \epsilon_c)\langle\phi_c|. \tag{E.6}$$

By averaging or neglecting the dependence upon the energy, we obtain a smooth potential $\tilde{V}$ that has smooth solutions: the prototype of a pseudo-potential.

The big problem of such approach is the so-called *orthogonality hole*: it is apparent that the pseudo-wavefunctions $\tilde{\psi}$ obtained from the plane-wave part only:

$$\tilde{\psi} = \sum_j c_j b_{j,\mathbf{k}}(\mathbf{r}), \tag{E.7}$$

do not have the correct normalization, due to the lack of orthogonalization with the core states.

E.2 A modern view

The ideal pseudo-potential has smooth valence "pseudo-wavefunctions", that can be expanded into PWs and do not exhibit the "wiggles" due to orthogonalization to core states. However, they must be equal to the true wavefunctions far from the core region.

In the Projector Augmented Wave (PAW) method, a mapping is introduced between the complete wavefunction and the pseudo-wavefunction via a suitable linear operator:

$$|\tilde{\phi}_l\rangle = (1 + \mathcal{T})|\phi_l\rangle \tag{E.8}$$

where the functions $\tilde{\phi}_l$ are solutions, regular at the origin but not necessarily bound, for the atom; the functions ϕ_l are corresponding pseudo-functions. These are constructed as smooth functions in the core region that join smoothly to the $\tilde{\phi}_l$ beyond the core region, where we may set $\mathcal{T} = 0$.

In the core region, we write a pseudo-wavefunction ψ for our molecular or solid-state system as a sum over the atomic pseudo-waves ϕ_l:

$$|\psi\rangle = \sum_l c_l|\phi_l\rangle \tag{E.9}$$

By applying the operator $(1 + \mathcal{T})$ to both sides of the above expansion we find

$$|\tilde{\psi}\rangle = \sum_l c_l|\tilde{\phi}_l\rangle \tag{E.10}$$

where $\tilde{\psi}$ is the all-electron wavefunction. The above result can be recast into the form

$$|\tilde{\psi}\rangle = |\psi\rangle + \sum c_l \left(|\tilde{\phi}_l\rangle - |\phi_l\rangle\right). \tag{E.11}$$

It remains to define the c_l coefficients. Let us introduce the projectors β_l with the following properties:

$$\langle\beta_l|\phi_m\rangle = \delta_{lm}, \qquad \sum_l |\phi_l\rangle\langle\beta_l| = I. \tag{E.12}$$

It is easy to verify that $c_l = \langle\beta_l|\psi\rangle$ and that we can write

$$|\tilde{\psi}\rangle = |\psi\rangle + \sum_l \langle\beta_l|\psi\rangle \left(|\tilde{\phi}_l\rangle - |\phi_l\rangle\right) \quad (E.13)$$

$$= \left[I + \sum_l \left(|\tilde{\phi}_l\rangle - |\phi_l\rangle\right)\langle\beta_l|\right]|\psi\rangle. \quad (E.14)$$

The quantity between square brackets is our $1+\mathcal{T}$ operator. This replaces the pseudo-states ϕ from the pseudo-wavefunctions around the atoms and replaces them with the all-electron states $\tilde{\phi}$. The $1+\mathcal{T}$ operator is a purely atomic quantity that is obtained from a judicious choice of the $\tilde{\phi}_l$ all-electron atomic states, the corresponding pseudo-states ϕ_l, and the projectors β_l.

The equations to solve in the PAW method are then obtained by inserting the above form for $\tilde{\psi}$ in the energy functional and by finding its minimum with respect to the variation of the smooth part only, ψ. Rather cumbersome expressions results. An important feature of the resulting equations is that the charge density is no longer given simply by the square of the orbitals, but it contains in general an additional (*augmentation*) term:

$$n(\mathbf{r}) = \sum_i |\psi_i(\mathbf{r})|^2 + \sum_i \sum_{lm} \langle\psi_i|\beta_l\rangle q_{lm}(\mathbf{r})\langle\beta_m|\psi_i\rangle \quad (E.15)$$

where

$$q_{lm}(\mathbf{r}) = \tilde{\phi}_l(\mathbf{r})\tilde{\phi}_m(\mathbf{r}) - \phi_l(\mathbf{r})\phi_m(\mathbf{r}) \quad (E.16)$$

(using the completeness relation, Eq.(E.12)). Conversely the pseudo-wavefunctions are no longer orthonormal, but obey instead a generalized orthonormality relation:

$$\langle\psi_i|S|\psi_j\rangle = \delta_{ij}, \qquad S = I + \sum_{lm} |\beta_l\rangle Q_{lm}\langle\beta_m| \qquad Q_{lm} = \int q_{lm}(\mathbf{r})d\mathbf{r}, \quad (E.17)$$

where the integral is performed over the core regione.

The careful reader will notice some similarity between the PAW approach and the approach of Sec.E.1 based on the OPW method. In the PAW approach the orthogonality hole is "plugged" by defining the charge density in the correct way.

The PAW method can be used for the solution of the electronic structure problem in molecules or crystals. In addition, it gives a formal and clear foundation to the theory of pseudopotentials: all modern pseudopotentials ("ultrasoft" or "norm-conserving") can be shown to be special cases of PAW.