Effective interactions in the quantum theory of molecular and condensed matter physics

Dissertation

zur Erlangung des Doktorgrades der Naturwissenschaften
(*Dr. rer. nat.*)

der

Naturwissenschaftlichen Fakultät II
Chemie, Physik und Mathematik

der Martin-Luther-Universität
Halle-Wittenberg

vorgelegt von

Herrn Falk Tandetzky

geb. am 26. Oktober 1986 in Schleiz

Contents

Chapter 1

Introduction

The aim of quantum chemistry and condensed matter physics is to describe the properties and the interaction of molecules, nano-structures and condensed matter. A detailed understanding is essential for the design of new molecules, materials or functional devices at a nano-scale. In principle the problem is easily outlined: We want to describe a system of electrons and nuclei on a quantum mechanical level. The mathematical description is given by the Schrödinger equation. It explains all desired properties to a high level of accuracy. However, the effort to solve this equation numerically is usually prohibitively large. Already representing the wave function on a computer is not possible in a straight forward way for most systems of interest, as the memory requirements increase exponentially with the number of particles under consideration. This is already an issue for small molecules. The number of degrees of freedom in a realistic system that resembles a solid is usually of the order of 10^{23} particles. To deal with this one often considers systems which are strictly periodic crystals as a first approximation to a real material. Using the Bloch theorem this reduces the number of degrees of freedom we have to deal with to something usually between ten and some thousands of degrees of freedom. Actually the Bloch theorem in its most well known form only applies to non-interacting particles. I will later come back on how to introduce an interaction.

There is another issue with the degrees of freedom we are dealing with. The mass of the nuclei differs by a factor of about one thousand to one hundred thousand to the mass of the electrons. This makes it very cumbersome to treat them on the same footing, at least numerically. This is usually fixed by the Born-Oppenheimer-Approximation. That is, one fixes the positions of the nuclei and solves the problem on how the electrons act in this fixed frame of nuclei. Roughly speaking this is like treating the nuclei as if they had infinite mass. In a second step one can then describe the behavior of the nuclei, but let us, for the moment, focus on excitations of the system of electrons. The question whether this separation of the system in electronic and nuclear degrees of freedom can be made is not an easy one. I will

come back to this. Finally, one has to decide to which extent relativistic effects are included. Often neglecting relativistic corrections can still provide reasonable results, at least when no heavy nuclei are present in the system of interest.

So we want to describe a non-relativistic periodic solid or a molecule in the Born-Oppenheimer-Approximation. For very small molecules, it is in fact possible to directly approximate the ground state wave function and to derive relevant properties from that. The most accurate methods for this are configuration interaction methods and extensions of them [1, 2]. This approach is limited to about ten to some hundreds of degrees of freedom, depending on the system, the desired accuracy and the properties one actually wants to characterize. Diffusion Monte Carlo methods can increase the number of particles one can deal with considerably, though at the cost of a reduced accuracy and making the access to most excitation properties much more difficult [2].

For larger systems the full wave function can no longer be approximated so well. The difficult part is the interaction among the particles. One of the first successful approaches to obtain ground state energies of atoms or molecules was the Hartree approximation, which replaces the interaction part of the Hamiltonian by its mean field potential. As a result the wave function becomes much simpler, namely a Slater determinant, which can be represented on a computer even for systems of moderate size. The Kohn-Sham system of density functional theory (DFT) [3, 4, 5] can be viewed as the exactification of this approach. It describes a way, how one could in principle find an effective local one-body potential that allows for the calculation of the ground state energy and its electron density. This approach has two important limitations. First, it relies on a suitable approximation of the so called exchange-correlation energy functional. The most famous and a very successful one is the local density approximation. It has been and still is an active field of research to find new such functionals in order to further improve on these approximations. The second fundamental limitation of Kohn-Sham DFT is the fact that it does not grant access to excited state properties of the system. It is constructed in such a way that the ground state density and total energy can be obtained. Properties that are related to excitations of the system are not immediately accessible. It is well known that even the exact exchange-correlation energy functional would not yield a Kohn-Sham system with the same excitation energies as the interacting system. Also ground state expectation values of relevant operators do not coincide in the Kohn-Sham system and the actual interacting system.

A theory that allows to predict excitation properties of a system is time-dependent DFT [6, 5]. In this case the interaction is replaced by a time-dependent exchange correlation potential. Unfortunately, finding good functionals becomes even more involved than in ordinary DFT. The functional not only depends on one density

distribution in space, but also on its history. There is no systematic way known to improve existing functionals.

Due to the limitations of the theories described above, I take another route. Towards this end let us consider a very successful tool in the field of quantum mechanics: perturbation theory. For that one splits the total Hamiltonian into a part that can be solved to high accuracy, like in this case the non-interacting system and a remaining part.

$$H = H_0 + \lambda H_{\mathtt{I}}, \tag{1.1}$$

where the λ is an artificially introduced real parameter. Then one expands the quantity of interest into a Taylor series with respect to λ. The first or the first few terms of this expansion can often be calculated relatively easily. If one neglects all other terms and sets $\lambda = 1$ one can hope to obtain a good approximation. This works well whenever the part that was treated perturbatively is, in a sense, small.

In the context of molecular or solid state physics the interaction among the particles is the Coulomb interaction. Unfortunately, this can almost never be regarded as small. In fact the very first term of this expansion, which is known as Hartree-Fock term, is not even well defined in a solid. The reason is the long ranged nature of the Coulomb interaction, which leads to an infinite energy density due to the repulsion of the electrons in the infinite solid. This is only compensated by the Coulomb interaction with the nuclei, which render the solid overall neutral. This issue is easy to deal with: one can change the reference system to be the one that includes the Hartree potential and regard the Coulomb interaction minus the Hartree potential as the perturbation. This renders all contributions finite.

Interestingly, such an approach introduces a completely new feature. It is well known that the Hartree or Hartree-Fock problem often has more than one self-consistent solution, even when following an aufbau-principle [7, 8, 9, 10, 11]. It is particularly interesting that even for systems in which the ground state is not degenerate, there may be several solutions to the Hartree-Fock problem, all with the lowest energy possible in this approximation. This has for instance been studied for cases like the stretched hydrogen molecule [12] or other diatomic molecules [11] where the lowest Hartree-Fock solution can be symmetry breaking.

This feature is also present for DFT in the Kohn-Sham formulation [13, 14], which should not be surprising, as the trivial exchange correlation functional which vanishes identically, is the Hartree problem. In this context one can argue that the ambiguity in the solutions is introduced when the functional is approximated. In fact the exact functional is known to allow for one self-consistent solution only and convergence to this can be guaranteed with appropriate algorithms [15].

This point of view clearly supports the approach to try to improve DFT functionals. There are important limitations to this, however. The more practical one is that there is no clear guideline on how these can be improved further. The more fundamental issue is that DFT, in principle, does not provide access to excitation properties of the system. Instead the Kohn Sham system is an artificial mapping to a system of strictly non-interacting particles. Apart from the limitations on the approximation of the functional, this mapping is exact. As stressed before, the Coulomb interaction cannot be viewed as a small perturbation of the system. Hence one would not expect that the Kohn-Sham system has much in common with the real system. Despite of this DFT calculations are often used to directly compare to band structures, i.e. excitation properties, which have been obtained experimentally. The very concept of a band structure is actually based on a picture of non- or at least weakly interacting particles.

The key to understand this discrepancy is the concept of quasi particles [16, 17]. For the description of most of the properties of a material, only the low lying excitations are of interest. And as it turns out those can often be described by effective, weakly interacting particles. When an electron moves through a solid it repels other electrons, leading to a polarization cloud around it. This change of charge density in the vicinity plus the electron itself is called quasi electron. The polarization cloud effectively screens the charge of the electron, so that indeed two such quasi-particles interact less strongly than free electrons would.

The description of materials in terms of these quasi particles is surprisingly successful. Only with this background it can be understood, why band structures are actually meaningful concepts despite the Coulomb interaction among the electrons. This is also the reason why the Kohn Sham system is more than a mathematical tool. It often yields a decent approximation to the quasi particles just described [18].

If we want to get a description which is more accurate and keeps information on the remaining interactions among the quasi particles, we have to go beyond DFT. One path one can follow is many body theory. This is formulated in the language of Green's functions [19, 20, 17, 21, 22], which are a natural tool to describe the quasi particles of the system. The term Green's function is actually a bit misleading, as this is only a mathematical Green's function for non-interacting systems. An accurate approximation to the Green's function (GF) grants access to a number of experimentally relevant quantities, including not only ground state expectation values, but also linear response properties and excitation spectra [17]. In order to calculate the GF of an interacting system without using the wave function of the interacting system one can use many body perturbation theory. This approach can be viewed as a perturbative expansion of the Green's functions in terms of the interaction among the particles. It was originally developed in the context of

quantum electrodynamics and turned out to be a useful tool in solid state physics as well.

The mentioned expansions give rise to a number of interesting questions. It is well known that perturbative expansions usually have a convergence radius of zero [23, 24, 25, 26, 27], even for examples as simple as the harmonic oscillator perturbed by a cubic or quartic potential [28] or the Hydrogen atom perturbed by an electric or magnetic field (Stark and Zeeman-effect), which are closely related problems [29]. However, if the perturbation is small the first few terms decrease in magnitude and can nonetheless provide accurate results. At the same time one can often only calculate the first few terms in such an expansion.

These considerations show that it is very important to formulate the theory in such a way that the perturbation one expands in, is small. As already discussed this condition is not met for the Coulomb interaction. It would be desirable to describe the system using an effective interaction that more or less resembles the weaker interaction between the quasi particles described above. Then an expansion in terms of such an interaction can be hoped to provide good approximations, even if only a limited number of terms is evaluated. Remarkably such an approach was successfully implemented by Hedin, already half a century ago [30]. The equations he derived for this approach are now known as Hedin equations. They are formally exact equations that allowed Hedin to obtain the first terms of an expansion needed to calculate the Green's function. The expansion is formulated in terms of the so called screened interaction, which is the Coulomb interaction screened by the polarization of the medium, similar to the quasi-particle interaction described above [17]. Though the first few iterations have already been carried out by Hedin [30], he did not comment on whether or not this procedure converges. In a recent book by G. Stefanucci and R. van Leeuwen [22] it is stated that "[the authors] are not aware of any proof that such an iterative scheme converges, nor that it generates all possible skeleton diagrams for the self-energy". One result of this work is such a proof.

The original motivation for this work was a series of papers on the so-called one-point-model [31, 32, 33, 34] that treat Hedin's equations in a simplified but unphysical framework, where functions appearing in the equations are only considered for one point in space and time. Since those were solved explicitly, it was a natural question to ask, whether or not one can solve the equations for more than one point. One has to deal with a non-linear partial differential equation, for two or more points. Unfortunately, for non-linear partial differential equations already the question of existence and uniqueness of solutions is very complicated. Even long known and intensively studied issues such as the famous millennium problem on the existence and uniqueness questions for the Navier-Stokes equation are still unsolved due to the complications arising from chaotic or fractal behavior at small length

scales [35]. In the present case, it is also not obvious which boundary conditions would be natural. The discussion becomes easier, if the space of allowed solutions is restricted to analytic functions. For the equation at hand I was able to prove that there is at most one solution, and there is no freedom of choice for boundary conditions if one wants a solution that is analytic near the non-interacting case. This unique solution is of course identical to the sum of many body perturbation theory, whenever it converges. Finding closed expressions for this sum is of course a difficult task, and the result would depend on the precise way how the discretization takes place. The question whether one can find a closed expression in general, and perform the limit when the number of points considered tends to infinity remains open. As this would necessarily reproduce the exact solution to many body theory it would certainly be necessary to introduce approximations at some point.

The first parts of the present work focus on Hedin's equations. In the third chapter I discuss the convergence of the process of iterating Hedin's or similar equations symbolically. For this sake, I develop a theory to analyze equations formulated in terms of Feynman diagrams. In the famous 1965 article by Hedin [30] the first couple of iterations were illustrated and lead to expansions in terms of an effective interaction as described above. I clarify how this iteration process can be put into precise terms, by introducing another object which I call polarization vertex. This makes the iterative process more transparent. I then prove that such a process always converges to a unique formal diagrammatic expansion, independent from the starting set of diagrams. The theory developed to show this is general enough to cover many similar systems of equations. I will highlight some important such examples along the way towards Hedin's equations.

In chapter four, I consider the same equations, but rather than analyzing the diagrammatic expansions, I discuss the equations as partial differential equations. I provide a proof that the solution in the space of analytic functions is unique. I also investigate the most famous approximation based on Hedin's equations: the GW-approximation [36, 18, 37, 38]. To this end I analyze the fixed point structure of the emerging equations and study the convergence behavior both analytically and to a lesser extent also numerically.

In the fifth and final chapter I analyze the concept of the Berry phase in the framework of the exact factorization. This is about how one can lift the Born-Oppenheimer approximation and which effective interactions emerge for the system of nuclei when the electron degrees of freedom are integrated out. While this approach is not related to Green's function methods (yet), the underlying mathematical concepts employed are the same. In fact for both, the analysis of the solutions to Hedin's equations, and the analysis of the properties of the Berry phase in the exact factorization, the understanding of the nodal structure of polynomial equations is

essential. In the former case the polynomials emerge from coupled integral equations by discretization of the space. And in the latter case the space of parameter dependent hermitian operators with degeneracies turn out to be important. The latter can be characterized by nodes of the characteristic polynomial.

The main result of this last chapter is that the most common type of a molecular Berry phase, namely a topological Berry phase, never shows up in the exact factorization. I also show that for time reversal symmetric states (i.e. when the wave function can be chosen real) there is no Berry phase at all in this prescription. Finally, I show some no-go theorems about the nuclear wave function of the exact factorization, which provide important guidelines for further analytical studying and numerically implementing the theory of the exact factorization, both in the time-dependent and time-independent case.

Chapter 2

Basics

In this chapter I introduce some basic mathematical concepts and methods that we will need in the subsequent parts of this work.

Notation I try to be ISO-80000 conformal, i.e. I use italic letters for variables and upright letters for mathematical and physical constants and fixed symbols, including names. This makes it easier to read a formula and reduces the risk of misunderstandings. So for example the letters e and i would be italic, if they appear as variables, while they are written upright, if the Euler constant or the imaginary unit is referred to. Similarly a subscript like in G_{H} is written upright, when it refers to a name, like in this case Hartree, while an italic subscript, like in ϕ_R indicates that the object actually depends on a variable called R. Similarly, functions are usually italic ($f(x), g(y), \dots$), but fixed mathematical symbols like the gamma function or the delta distribution are written upright ($\Gamma(n), \delta(x), \dots$).

A convention in many body perturbation theory is to use numbers as variables, which can be a bit confusing at first sight. I will stick to this, but try limit any possible confusion, by putting italic numbers for that ($\mathit{1}, \mathit{2}, \mathit{3}, \mathit{4}, \dots$). As they are connected I will use the same italic numbers as labels in Feynman diagrams, though there they may be thought of as as just numbers or labels (as opposed to variables) assign to a vertex of a graph.

2.1 Systems of polynomial equations

Many physical problems lead to coupled non-linear equations. Often times approximations can transform them to systems of linear equations, which can then be solved by means of linear algebra. If one is not in such a fortunate situation, it might still be possible to reduce the problem to a system of polynomial equations. This section will provide a basic introduction on how such systems can be solved and what the

properties of such systems are. With the physical applications in mind, we only consider real or complex numbers for the space of solutions.

There are two limiting cases of polynomial equations that are easy to understand. One is the case that there is only one variable and one equation, the other is a system of linear equations. Let us consider the former first:

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 = 0, \tag{2.1}$$

where the a_k are complex (or real) coefficients and we seek for a complex solution x. Then the famous fundamental theorem of algebra tells us that the order of the polynomial n equals the number of solutions to the system. For specific coefficients it may happen that some of the solutions fall on top of another, effectively reducing the total number of distinct solutions. E.g. if $a_1 = a_0 = 0$ and $a_3 \neq 0$, then $x_0 = 0$ is a solution with multiplicity two, meaning that one can write the polynomial as a product of $(x - x_0)^2$ and another polynomial. This case is however the exception. Generically, all solutions are different. More specifically, if we choose all the coefficients at random (e.g. uniformly distributed on some interval), then the probability that there is a degeneracy is zero. The most obvious approach to find solutions to such an equation is to minimize the modulus of the polynomial. It is not hard to show that the only local minima are the roots of the polynomial, which guarantees that this minimization will always lead to solutions.

Now let us consider the other limiting case of systems of polynomial equations: A system of linear equations. We further restrict ourselves to the case that the number of equations equals the number of unknowns:

$$\begin{aligned} a_{11} x_1 + a_{12} x_2 + \cdots + a_{1n} x_n &= b_1 \\ &\vdots \\ a_{n1} x_1 + a_{n2} x_2 + \cdots + a_{nn} x_n &= b_n \end{aligned}$$

As is well known, the structure of the solution space depends on the coefficient matrix a_{ij}. If the determinant of it is not zero (the generic case) then there is exactly one solution to this system. If the determinant happens to be zero, there is either no solution at all, or a space of solutions of dimension n minus the rank of the matrix (which is the number of linear independent row-vectors of the matrix).

After this little warm up, let us consider the case of a system of polynomial equations. As before we restrict ourselves to the case that the number of equations

equals the number of variables.

$$\begin{aligned} P_1(x_1, x_2, \ldots, x_n) &= 0 \\ &\vdots \\ P_n(x_1, x_2, \ldots, x_n) &= 0, \end{aligned} \tag{2.2}$$

where P_i are polynomials in the denoted variables (i.e. linear combinations of products of these variables). As in the case of linear equations it may happen that some of the equations are automatically satisfied for all solutions of the other equations. Then the solutions may not only contain lines, planes and hyperplanes, but also objects like circles, parabolas, spheres, cones or hyper-spheres, to name but a few examples. However, in the generic case (meaning for random coefficients), the space of solutions is simply a finite set of isolated points [39].

In order to solve such a system of polynomial equations, one can apply Buchberger's algorithm [40]. It is possible, though a bit more involved than in the linear case, to eliminate one variable from all but one equation. We will describe this process in a moment. Given that this is possible, it is natural to follow the strategy of the famous Gauss elimination method. The first step is to get rid of one of the variables in all but one of the equations. This way one obtains $(n-1)$ equations in $(n-1)$ variables, which can be simplified further by repeating this strategy. What one ends up with is a system of equations in triangle form

$$\begin{aligned} Q_1(x_1) &= 0 \\ Q_2(x_1, x_2) &= 0 \\ Q_3(x_1, x_2, x_3) &= 0 \\ &\vdots \\ Q_n(x_1, x_2, x_3, \ldots, x_n) &= 0, \end{aligned} \tag{2.3}$$

where the Q_k are again polynomials. If this system of equations is equivalent to the original one, it is a so-called Gröbner basis of the original system. This is straight forward to solve. The first equation is a polynomial equation in one variable. Once a solution to this is found, it can be inserted in the second equation. Then this is again a polynomial in one variable, which can be solved in turn. This can be continued until all equations are solved simultaneously. That way one can obtain all solutions to the system (though there are numerically much more stable algorithms).

Now we come back to how the elimination of variables actually works. Let us consider an example:

$$y^2 + x^2 - 2 = 0 \tag{2.4}$$

$$xy - 1 = 0 \tag{2.5}$$

The first equation describes a circle for real x and y. Since they are allowed to take arbitrary complex values, we notice that for purely imaginary x and real y, hyperbolas are described by this as well. In the same sense, the second equation describes hyperbolas as well as circles. In order to solve this system one could of course solve the second for one of the variables in terms of the other and plug the solution into the first equation. This would lead to a single equation in one variable. However, this approach only works for polynomials of low order. For higher order polynomials there are no explicit solution formulas known. Let us say we want to eliminate y from one of the equations. The strategy is to consider x as a parameter and remove the highest order of y first. In this example the highest order of y is two, which appears in the first term of the first equation. We can eliminate it by multiplying the first equation with $2x$ and the second equation with $-y$ and summing the resulting equations.

$$\left|\begin{array}{rl} y^2+x^2-2 &= 0 \\ xy-1 &= 0 \end{array}\right| \begin{array}{l} \cdot x \\ \uparrow \cdot(-y) \end{array} \tag{2.6}$$

$$\Rightarrow \left|\begin{array}{r} x^3-2x+y=0 \\ xy-1=0 \end{array}\right| \tag{2.7}$$

$$\Rightarrow \left|\begin{array}{rrl} y & +x^3-2x &= 0 \\ xy & -1 &= 0 \end{array}\right|, \tag{2.8}$$

where in the last step we rearranged the terms. Now there is no term quadratic in y left. We can now remove the term linear with respect to y.

$$\left|\begin{array}{rrl} y & +x^3-2x &= 0 \\ xy & -1 &= 0 \end{array}\right| \begin{array}{l} \cdot x \\ \uparrow \cdot(-1) \end{array} \tag{2.9}$$

$$\Rightarrow \left|\begin{array}{rl} x^4-2x^2+1 &= 0 \\ xy-1 &= 0 \end{array}\right| \tag{2.10}$$

Now the system has the triangular form of Eq. (2.3). The first equation has the solutions minus one and plus one, both with multiplicity two. If we insert one of those solutions into the second equation, we obtain the corresponding y. Hence the complete set of solutions of the system is

$$(x,y) \in \{(-1,-1),(1,1)\}.$$

Note that we need to check against the original equation, as in the process spurious extra solutions can appear. This example illustrates how the process works in general. If more variables and equations are present one can always regard the extra variables as parameters. Then one can remove one variable order by order,

beginning from the highest order, just as in the presented example. Eventually one will always end up with a system like 2.3. The order of the polynomials can increase exponentially with the number of equations and variables one starts with. As we already mentioned, generically the solutions to this system are then just isolated points. For the case of generic polynomials that is all that is to it. More specifically, if in each polynomial the constant term is chosen randomly, then the number of solutions to the system is finite with probability one [39]. Otherwise there may be a continuum of solutions.

2.2 Implicit function theorem

In the previous section, we considered systems of algebraic equations in which the number of equations equal the number of variables. Physically motivated equations do almost always contain some parameters, which may take different values. One might face an equation of the form

$$f(x, y) = 0, \tag{2.11}$$

where f is a known function, x are parameters of the problem and y are the unknowns one is interested in. For given parameters this equation should not allow for a continuum of solutions, hence the number of equations should equal the number of unknowns. So if y is an n-dimensional vector, f should be an n-dimensional vector as well, while there may be any number of parameters m.

$$x \in \mathbb{R}^m, \quad y \in \mathbb{R}^n, \quad f : x, y \mapsto f(x, y) \in \mathbb{R}^n. \tag{2.12}$$

One of course wants to solve the equations, i.e. one seeks a function $y(x)$, which describes the dependency of the unknowns on the parameters. Unfortunately, it is often not possible to solve for this function explicitly (meaning to express this function in terms of some known basic functions). Is is anyway possible to obtain properties of this solution, even if we cannot write it down explicitly. The main tool for this is the so-called implicit function theorem, which states that such a function exists under reasonable conditions and that if f is in a certain sense well behaved, then so is the function $y(x)$. A natural condition is of course that the equation does have a solution at all. The precise statement is [41, 42]

Theorem 1. *Let $f : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^n$ be a continuously differentiable function. And let x_0, y_0 be a solution to $f(x_0, y_0) = 0$. If the Jacobian matrix $J_{a,b} := \frac{\partial f_a}{\partial y_b}\Big|_{x_0,y_0}$ has non-zero determinant, then there exists a neighborhood $U = U' \times U'' \subset \mathbb{R}^{m+n}$ of*

$(x_0, y_0) \in U$ and a function $y : U' \to U'' : x \mapsto y(x)$ such that for all $x \in U'$ this function satisfies $f(x, y(x)) = 0$ and there is no other y in U'' that solves $f = 0$. Further more, if f is k times continuously differentiable in U, then so is $y(x)$ in U'. Similarly, if f is analytic in U, then so is $y(x)$ in U'.

It is important to notice that the statement about the unique solution is only valid in a neighborhood. One can see from simple examples that the provided conditions are not sufficient to allow for a global statement about the uniqueness of the solution [42].

Chapter 3

Feynman diagrams and Hedin's equations

Most works in the general context of solid state theory start with writing down a Hamiltonian, setting up the problem to solve. We will not do that. Firstly because the formalism we develop does not only apply to solid state physics, but also partly to other theories that can be described using diagrammatic techniques. These include for example the Feynman diagrams of other theories, like those of particle physics or the diagrams appearing in renormalization group theory [43]. The second reason is that this way, this chapter will be completely self contained, while we will need to refer to the literature on the theory of many body theory, when we introduce it.

We start with some definitions from graph theory. We will work out some important identities of Green's function theory without even defining what a Green's functions is in physical terms. Afterward we will of course clarify how all this relates to physical quantities of interest. We will then derive the analog of Hedin's equations in the language of Feynman diagrams, as alternative route to the original derivation, that uses the Martin-Schwinger functional derivative technique [30].

3.1 Graphs for QED Feynman diagrams

In this section we will formally define what we mean by expansions in terms of Feynman diagrams. We are doing this in a detailed way in order to be absolutely clear as to what we are talking about in the subsequent sections. For the reader familiar with Feynman diagrams there should not be any surprise here.

The most basic definition of a graph is a set of vertices with edges linking them. We will need a slightly more general notion of graphs, vertices and edges. We first define the basic building blocks

Definition 1. *A vertex is represented by a set of labels (e.g. $\{1\}$).*

Usually this set contains one element only. However, we will later also allow for several labels at one vertex.

Definition 2. *A directed edge is a symbol and an ordered pair of vertices. The interpretation is that the edge points from the second vertex to the first one.*

Definition 3. *An undirected edge is a symbol and a set of two vertices.*

The interpretation of the symbol will allow us to attach a physical meaning to the edge. Examples of symbols we will deal with are G_0, G_{H}, G, v and W, representing different Green's functions (GFs) and interactions, respectively.

Definition 4. *A graph is a triple of a finite set of vertices and, built from these, a set of directed edges and a set of undirected edges. The set of vertices is not empty and no label appears twice in the set of vertices.*

Definition 5. *A graph is called quantum electrodynamics Feynman diagram (or just diagram), if each vertex is linked to at most one ingoing, at most one outgoing and at most one undirected edge. We call vertices that are connected to three edges internal vertices, and all other external vertices.*

We will from now on call the directed edges *GF lines* and the undirected ones *interaction lines*. We will call external vertices that miss an ingoing (outgoing) GF line *open end for ingoing (outgoing) GFs*. Analogously we call external vertices that miss an interaction line *open end for interaction lines*.

Definition 6. *Two graphs are considered equal if one can relabel the internal vertices so that the graphs are equal in the sense of sets.*

In this sense diagrams are only *represented* by the sets introduced above. If we, for instance, talk about the set of all diagrams, we actually mean the set of all equivalence classes of diagrams.

An example of a graphical representation of a Feynman diagram is provided in Fig. 3.1. We represent interaction lines by wiggly lines and GF lines by straight lines with an arrow. We highlight vertices with open ends by drawing a short wiggly or directed line to it. These short lines are no GF - or interaction lines, but are rather guides for the eye. We will usually not do this if the vertex is an open end for both an interaction line and a GF line. We will, of course, label edges with their corresponding symbols whenever necessary.

Definition 7. *We call the open ends of two diagrams A and B compatible, if for each label, labeling an external vertex of diagram A, there is the same label appearing*

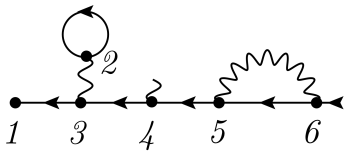
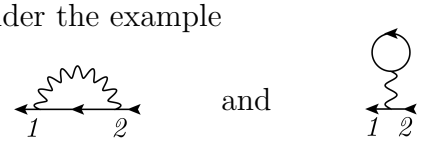


Figure 3.1: Example of a Feynman diagram with six vertices. Vertex *1* is an open end for both an outgoing GF and an interaction line. Vertex *4* and *6* are a open ends for an interaction line and an ingoing GF respectively. The small wiggly line at vertex *4* and the small ingoing line at vertex *6* are not representing an interaction - or GF line. They rather highlight the fact that these vertices are open ends.

in diagram B, labeling the same type of open end (one example of a type is an open end for an ingoing GF line and an interaction line) and vice versa.

A vertex, which has an open end for two (three) edges may have two (two or three) labels, each corresponding to one (one or two) of the edges. In this case diagrams are still considered compatible, if the labels can be matched up appropriately.

To illustrate this consider the example



. (3.1)

The first diagram has two open ends, one for an ingoing and one for an outgoing GF line. Each of them is labeled once. On the other hand, the second diagram has only one external vertex, labeled twice. If we now consider the labels of the second diagram as belonging to the outgoing and ingoing GF line respectively, then the diagrams are compatible in the sense of the definition above.

This compatibility is important: compatible diagrams can be linked to other diagrams in the same way. E.g. if one wants to link a GF line to two different diagrams, then they should, of course, both have an open end for it.

Eventually one can assign complex numbers to the diagrams defined here by means of the Feynman rules, which we will describe in the next chapter. It can make sense to build linear combinations of these complex numbers. One of the major results of many body perturbation theory (MBPT) is that one can approximate the Green's function by certain such linear combinations. It is very similar to approximating a function by linear combinations of monomials (better known as polynomials). In this chapter we are only interested in the coefficients of these linear combinations. The notion of formal power series allows one to discuss the coefficients appearing in a Taylor expansion of a function, even without assigning a number to the variable,

and even when the expansion does not even converge. We will adopt this strategy for the expansions of the functions of interest in MBPT.

Definition 8. *We define diagrammatic expansions in analogy to formal power series, as series of complex numbers, being interpreted as coefficients of the Feynman diagrams ordered in some way (How they are ordered precisely is not important. The only thing we need is that they can be enumerated somehow). In an expansion all Diagrams with non-zero prefactor should have compatible open ends. We consider two expansions equal, if the corresponding series are equal.*

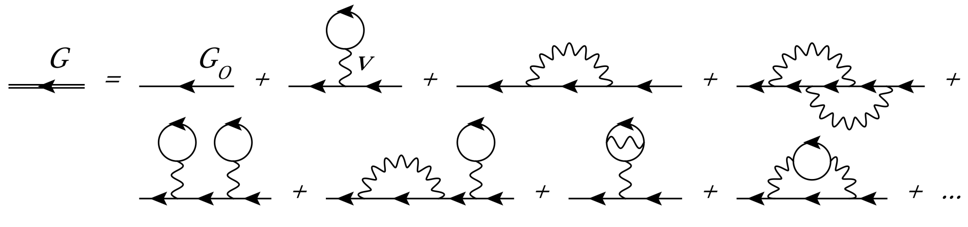
Whenever we consider linear combinations of diagrams, we also demand that they have compatible open ends. With that linear combinations of diagrams are of course equivalent to diagrammatic expansions with only finitely many nonzero coefficients. We will most of the time deal with diagrammatic expansions in which most of the coefficients are simply zero and the remaining ones are one. In this situation it is natural to represent these expansions just by drawing the diagrams that have a non-vanishing coefficient.

An illustration of the previous definitions and at the same time an important object is provided by

Definition 9. *The interacting GF (or just the GF) $G(1,2)$ is the diagrammatic expansion representing the sum of all connected diagrams with an open end labeled 1 and an open end labeled 2 with the properties*

1. *all GF lines have the symbol* G_0 .
2. *all interaction lines have the symbol* v.
3. *the vertex labeled 1 is only linked to a GF line that points towards this vertex.*
4. *the vertex labeled 2 is only linked to a GF line that points away from this vertex.*

The term connected is understood in the same way as it is common in graph theory. It means for any pair of vertices there is a path built from edges that links them. We use the labels *1* and *2* in this definition only as placeholders. We will of course call this expansion a GF even when other labels are used. Symbolically this definition can be summarized as



(3.2)

which becomes unambiguous if we state that the RHS consists of all connected diagrams. It is clear according to our definition that the diagrams on the RHS have two external vertices, that are supposed to be labeled. For brevity, we drop those labels, whenever there is no ambiguity.

We will also use different symbols for edges that correspond to different symbols (G_0, G, v etc.). Note that we no longer draw the vertices explicitly and that we sometimes drop the arrows on lines, when this does not lead to ambiguities, as e.g. in the last two terms written out in Eq. (3.2).

We now define several operations of diagrammatic expansions or linear combinations of diagrams (the latter can be regarded a special case of the former in which only finitely many non-zero coefficients appear).

Definition 10.

The linear combination *of two or more diagrammatic expansions is defined as the coefficient-wise linear combination. We allow this operation only on diagrams, where the open ends are compatible.*

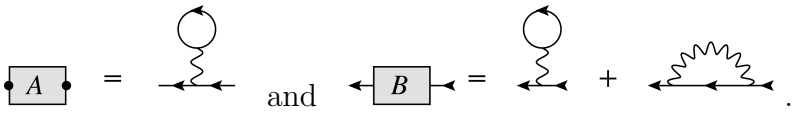
The contraction *of two open ends of one diagram is obtained by replacing two external vertices by one that is linked to the edges that were linked to the original vertices (Note that this operation can map different diagrams to the same one and that the resulting vertex can be internal or external). We will denote this either by directly drawing the corresponding vertices on top of each other, or by linking the two vertices with a dashed line. This is extended to linear combinations by applying the operation for each diagram individually (i.e. the contraction commutes with multiplication and addition). For expansions this is extended in a natural way, i.e. the coefficient of a given diagram is obtained as the sum of the coefficients of the diagrams leading to this given diagram via contraction. Note that this definition is only meaningful, after we demanded that expansions and linear combinations always represent sums of diagrams with compatible open ends.*

The combination *of two diagrams is understood as the (disconnected) diagram obtained from union of the vertices and the union of the edges of each diagram. Graphically this means just drawing the two diagrams next to each other. We demand that the sets of vertices of the combined diagrams are disjoint. We generalize this to linear combinations, by summing over the result from the combination of each pair of diagrams, where the prefactor is obtained as the product of the prefactors of the diagrams used as building blocks (i.e. the operation "combination" commutes with multiplication and distributes over addition). The extension to expansions shall be understood in the natural way, i.e. the coefficient of a given diagram is the sum of the product the coefficients of each pair of diagrams leading to the given diagram. Note that this sum is always a finite sum, so there is no ambiguity related to choosing*

the order of summation.

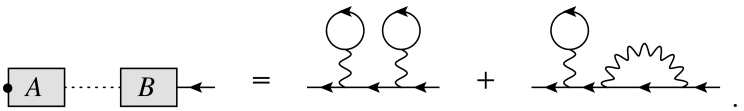
The contraction of two expansions *is understood as the corresponding contraction of the single expansion obtained from the combination of the two expansions.*

Now we provide examples illustrating the previous definitions. Let us define the two expansions (and at the same time linear combinations)



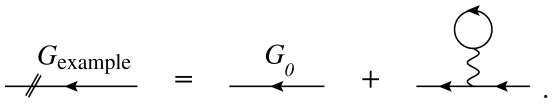
(3.3)

Note the difference in A and the first term in B. The diagram A consists of two more bare GF lines and has an open end for both an ingoing GF line and an interaction line, and an open end for an outgoing GF line and an interaction line, while the short arrows in B do not represent GF lines, but rather highlight the fact that there are open ends for ingoing or outgoing GFs. Now we can combine A and B to a single diagram with four open ends and contract the resulting diagram, or equivalently contract A and B. The result is an expansion with two open ends. This can again be contracted with a bare GF line, yielding



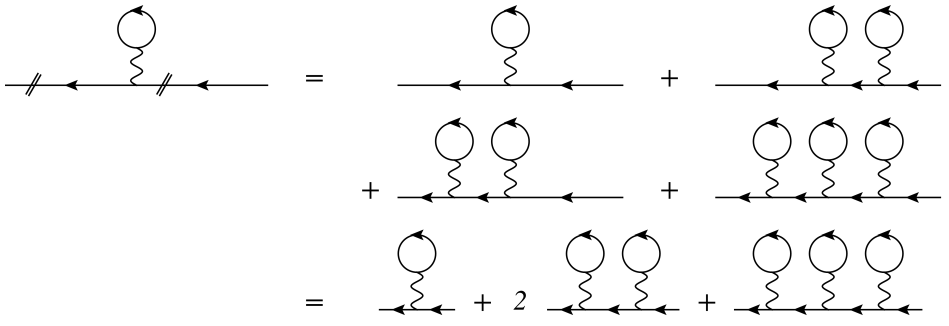
(3.4)

As another example, we define the linear combination



(3.5)

We can contract this expansion with itself and another diagram to obtain



(3.6)

Note that in this example one of the diagrams appeared twice. Such a multiple appearance of a diagram is often referred to as double counting.

Note that our notation for graphs and the notation for combining and contracting expansions is identical. Hence expressions like for instance the LHS of Eq. (3.6) can be understood either as a single Feynman diagram, where in this case some GF lines are labeled $G_{\texttt{example}}$, or as contractions of expansions of diagrams.

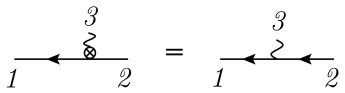
3.2 From equations for graphs to Hedin's equations

3.2.1 The vertex insertion operator

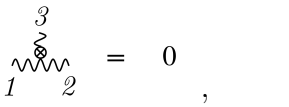
Building on the basic definitions made in the previous section, we will now introduce further concepts to manipulate graphs and develop a theory to analyze the resulting equations. Based on these concepts, we will not only derive Hedin's equations, but also characterize the solution structure to those and similar equations.

Definition 11. *We define the vertex insertion operator "$\sim\!\otimes$" as a map from a label and an expansions of diagrams to an expansion of diagrams in the following way*

1. *The action on a single link is given by*



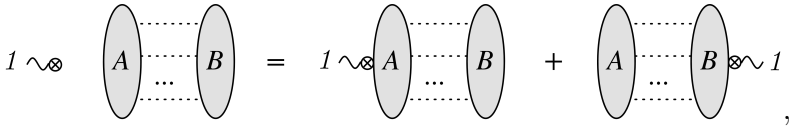
(3.7)



, (3.8)

   i.e. a bare GF is replaced by two that are linked to a new vertex and a bare interaction line is mapped to zero.

2. *It satisfies a product rule*



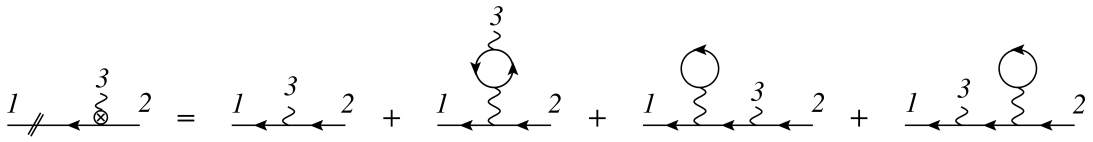
, (3.9)

   where A and B represent any two diagrams, each with any number of open ends. They may be linked by any number of contractions, as suggested by the ellipsis (zero is allowed as well). In the LHS of the equation we mean the action of the operator on the diagram obtained from contracting A and B, while on the RHS in each term the diagram obtained from acting with the operator is linked to the other diagram.

3. *The action of the operator on a linear combination is the linear combination of the action on each term (the operator commutes with multiplication and distributes over addition). This also applies to expansions (viewed as "infinite linear combinations") in an analogous fashion.*

Whenever we provide a graphical representation of this operator acting on diagrams, we consider it acting before any combination or contraction is done (The only exception is the LHS of Eq. (3.9)).

This definition means that the operator maps a diagram that consists of n bare GFs to the sum of n diagrams, where in each term, one of the GFs is replaced by two that are linked to an external vertex. We can illustrate this by acting with this operator on the GF of the example of the previous section (see Eq. (3.5))

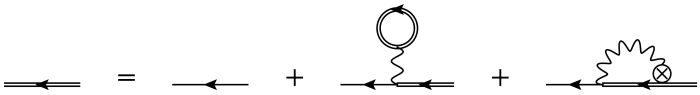


(3.10)

3.2.2 Graph equations

We can now state our first theorem

Theorem 2. *The unique solution* $G =$ *to the equation*



(3.11)

is the sum of all connected diagrams, i.e. the full GF, we defined in Def. 9 (p. 19).

We will base all our derivations on this equation. Before we provide the proof, we show the existence and uniqueness in a more general context. Eq. (3.11) is our first example of

Definition 12. *A grapherential equation for an unknown expansion* X *is an equation of the form* $0 = D[X]$*, where* $D[X]$ *stands symbolically for any linear combination of expansions, each built from combinations and contractions of any of* X*, the vertex insertion operator acting on* X*, GF lines and interaction lines (see Def. 10 (p. 20) and Def. 11 (p. 22)).*

We call the grapherential equation nice, if this linear combination includes the diagram X *alone (without extra GF lines, interaction lines, or the vertex insertion*

operator acting) and does not include a term in which X is contracted exclusively with itself.

We call the grapherential equation an algegraphic equation, if the vertex insertion operator does not appear.

Everything we will state for grapherential equations will also apply to algegraphic equations. For that reason it is convenient to regard the latter as a special case of the former.

A nice grapherential equation can be written as

$$X = A + D[X], \tag{3.12}$$

where A is a linear combination of diagrams that do not involve X, and D is a linear combination of diagrams that do not involve any diagrams without X appearing (this part was put in A) and not the object X alone (that part was pulled to the LHS of the equation).

Lemma 1. *The solution of a nice grapherential equation*

$$X = A + D[X], \tag{3.13}$$

exists and is unique. It can be obtained by iteration of

$$X^{n+1} = A + D[X^n], \tag{3.14}$$

as the limiting expansion (in the sense of coefficient-wise convergence), independent on the starting expansion X^0.

We emphasize that coefficient-wise convergence, though useful to construct a solution, is a rather weak notion of convergence, as opposed to e.g. uniform convergence. To illustrate this, imagine we iterate Eq. (3.11), starting from a disconnected graph as initial guess for G (we would do this explicitly, but the number of graphs increases so fast that it is impractical to provide them explicitly). Throughout the iteration process one would generate more and more disconnected diagrams, which one would not obtain, if one started with a connected diagram (or with zero) as an initial guess. However, this is in no contradiction to the provided lemma, as for any given disconnected diagram the coefficient differs from zero only for finitely many iterations and hence the limit of each such coefficient is indeed zero.

Proof. Of Lemma 1.

Step 1 We consider a nice grapherential equation

$$X = A + D[X], \tag{3.15}$$

where $D[X]$ involves only diagrams in which X appears at least once. Applying the vertex insertion operator or adding GFs or interaction lines (while keeping the same set of open ends) increases the number of vertices. Hence, if Y consist of diagrams, each with at least n vertices, then $D[Y]$ contains only diagrams that contain at least $n+1$ vertices (remember that we excluded the case that $D[Y]$ includes terms in which Y is contracted with itself only (see Def. 12 (p. 23)). Without that, the diagram consisting of one vertex only would spoil this property). What happens, if we add an expansion O_k to Y, assuming that O_k consists of diagrams, each with at least k vertices? Due to the commutation and distribution relations the allowed operations satisfy, it is clear that we get the same diagrams as before plus diagrams that are built using diagrams appearing in O_k. These extra diagrams contain at least $k+1$ vertices, by the same reasoning as before. We conclude

$$D[Y + \mathrm{O}_k] = D[Y] + \mathrm{O}_{k+1}, \tag{3.16}$$

where O_{k+1} is an expansion of diagrams, each with at least $k+1$ vertices (In analogy to the big-O-notation, we use this symbol for any expansion, i.e. O_k can refer to a different expansion in each equation).

Step 2 Now we define a series X^n via

$$X^0 = 0 \tag{3.17}$$

$$X^{n+1} = A + D[X^n] \tag{3.18}$$

and prove by induction that

$$X^{n+1} = X^n + \mathrm{O}_{n+1}, \tag{3.19}$$

i.e. X^n and X^{n+1} do only differ by diagrams with more than n vertices. The inductive base is trivial, even if we allow for any other initial expansion X^0. For the inductive step we assume that Eq. (3.19) is valid for $n-1$, i.e.

$$X^n = X^{n-1} + \mathrm{O}_n \tag{3.20}$$

and show that it follows for n by

$$X^{n+1} = A + D[X^n] \tag{3.21}$$
$$X^{n+1} \overset{(3.20)}{=} A + D[X^{n-1} + \mathrm{O}_n] \tag{3.22}$$
$$X^{n+1} \overset{(3.16)}{=} A + D[X^{n-1}] + \mathrm{O}_{n+1} \tag{3.23}$$
$$X^{n+1} \overset{(3.18)}{=} X^n + \mathrm{O}_{n+1}. \tag{3.24}$$

This concludes the inductive proof that Eq. (3.19) is valid in general.

Step 3 From Eq. (3.19) we conclude that for any given FD, its coefficient in the expansion X^n converges to a final value as n tends to infinity (if a diagram has k vertices, the coefficient is constant for all $n \geq k$). Hence the series has a limiting series X^∞, to which it converges coefficient-wise. Now we want to prove that X^∞ is the unique solution to Eq. (3.15). First we prove that it is a solution. By construction

$$X^n = X^\infty + \mathrm{O}_{n+1} \tag{3.25}$$

We plug this in the LHS and RHS of Eq. (3.18) and get

$$X^\infty = A + D[X^\infty + \mathrm{O}_n] + \mathrm{O}_{n+1} \tag{3.26}$$
$$X^\infty = A + D[X^\infty] + \mathrm{O}_{n+1}, \tag{3.27}$$

where we used Eq. (3.16) to pull out the O_n and the extra terms were absorbed in the O_{n+1}. This means that Eq. (3.15) is true, if we consider only the diagrams with at most n vertices. However this is valid for any value of n and hence actually for all coefficients. So X^∞ is a solution to Eq. (3.15).

What is left to be proven is that there is no other solution. So let us assume that both X and Y satisfy Eq. (3.15). If we further assume that

$$Y = X + \mathrm{O}_n, \tag{3.28}$$

i.e. X and Y have the same coefficients in front of the diagrams with at most $n-1$ vertices ($n = 1$ is allowed), then

$$Y = A + D[Y] \tag{3.29}$$
$$Y \overset{(3.28)}{=} A + D[X + \mathrm{O}_n] \tag{3.30}$$
$$Y \overset{(3.16)}{=} A + D[X] + \mathrm{O}_{n+1} \tag{3.31}$$
$$Y \overset{(3.15)}{=} X + \mathrm{O}_{n+1}. \tag{3.32}$$

Hence X and Y have the same coefficients even in front of all diagrams with up to n vertices. By induction it follows that $X = Y$ and hence that the solution is unique. □

This Lemma can be generalized to systems of grapherential equations.

Definition 13. *A nice system of grapherential equations for the unknown expansions $X_1, X_2, \ldots, X_N$ is a set of equations of the form*

$$\begin{aligned} X_1 &= A_1 + D_1[X_1, X_2, \ldots, X_N] \\ X_2 &= A_2 + D_2[X_1, X_2, \ldots, X_N] \\ &\ldots \\ X_N &= A_N + D_N[X_1, X_2, \ldots, X_N], \end{aligned}$$

where each A_k is a linear combination of diagrams that do not involve any X_l, and each $D_k[X_1, \ldots, X_N]$ is a linear combination of diagrams that include at least one of the X_l with the property

$$D_k[X_1 + \mathrm{O}_n, \ldots, X_N + \mathrm{O}_n] = D_k[X_1, \ldots, X_N] + \mathrm{O}_{n+1}, \tag{3.33}$$

where in each argument O_n can stand for any expansion of diagrams, each with at least n internal vertices.

The property of Eq. (3.33) is satisfied, if in all terms of D_k there is at least one contraction or one vertex insertion operator acting. Note that we specialized the meaning of O_n slightly, in that n now refers to the number of internal vertices rather than to the total number of vertices. This distinction was not important in the case of a single equation, but here it allows for equations that do increase the number of internal vertices while leaving the total number of vertices unchanged. Note, however that we could equally well deal with equations in which all of the appearing D_k do increase the total number, rather than the number of internal vertices.

Theorem 3. *The solution of a nice system of grapherential equations exists and is unique. It can be obtained by iteration of*

$$X_k^{n+1} = A_k + D_k[X_1^n, \ldots, X_N^n], \tag{3.34}$$

as the limiting expansions (again in the sense of coefficient-wise convergence), independent on the starting expansions X_k^0.

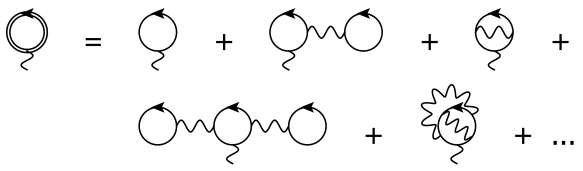
Proof. The generalization of the previous proof is straight forward. One just needs to replace the object X by the vector composed of the X_k and replace the term vertices by internal vertices. Eq. (3.33) is the generalization of Eq. (3.16). □

This Theorem 3 will be the basis for almost all existence and uniqueness statements that will come up. Every equation we will find in this chapter will be derived from the one we stated first (Theorem 2). That the objects introduced are then always well defined can be concluded from the theorem just proven.

3.2.3 Expansion in terms of the bare Green's function and the bare interaction

We need two more Lemmas in preparation of the proof of our first Theorem.

Lemma 2. *The contraction of the full GF with itself is the sum of all connected diagrams with one open end for interaction lines:*

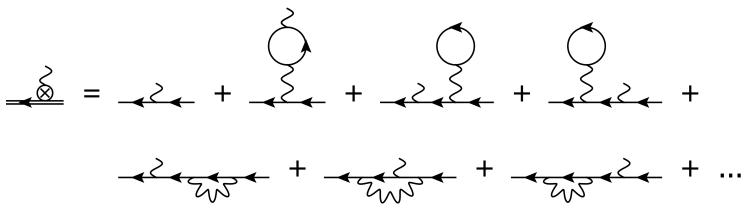


(3.35)

Proof. Consider a connected diagram with one open end for an interaction line (a diagram from the RHS of Eq. (3.35)). At the open end one GF is going in and one is going out. If we replace this vertex by two, each connected to one of these two GFs we obtain a connected diagram with two open ends. Hence this diagram clearly appears on the LHS of Eq. (3.35).

Finally we observe that connected diagrams lead to connected diagrams when contracted and in the case considered different diagrams can never lead to the same diagram. Hence we obtain each diagram on the RHS exactly once. □

Lemma 3. *When acting with the vertex insertion operator on the full GF one obtains the sum of all connected diagrams with the according open ends.*



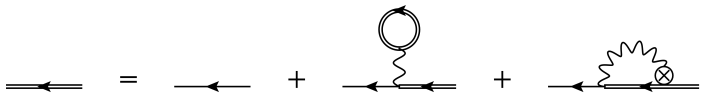
(3.36)

Proof. Clearly the diagrams remain connected when acting with the vertex insertion operator. We need two more steps.

Step 1 We show that every connected diagram appears. Consider a connected diagram with three open ends as on the RHS of Eq. (3.36). One can find the diagram that leads to it when acting with the vertex insertion operator by removing the open end for interaction lines and replacing the two GFs that were linked to it by one. This diagram is clearly connected and hence included in the expansion of the GF.

Step 2 We show that no diagram appears twice. As described in the first step one can construct the single diagram that leads to a given diagram. Also one diagram cannot lead to the same diagram twice, as there are no equivalent GF lines (This is only true since we are considering connected diagrams which have open ends). □

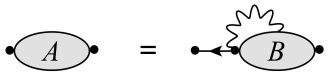
We are now prepared to prove Theorem 2. For convenience we repeat that the statement of this Theorem is that the GF, i.e. the sum of all connected diagrams, satisfies



(3.37)

Proof. Of Theorem 2
The existence and uniqueness can be concluded from Theorem 3. What is left to do is to show that, the full GF is this solution.

Step 1 We show that every diagram appearing on the LHS also appears on the RHS at least once. So consider a GF diagram A (i.e. a connected one with the appropriate open ends). If it is only a single GF line, we are done, otherwise we can write the diagram as

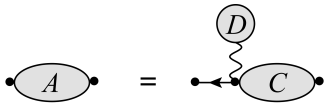


(3.38)

Now we distinguish two cases:

Case 1 Diagram B is connected. Then by Lemma 3 this diagram is included in the third term of the RHS of Eq. (3.37).

Case 2 Diagram B is not connected. Since A is connected this is only possible if A can be written as



(3.39)

where C and D are connected diagrams (We pulled D out of B, leaving C behind. B is hence the combination of C and D). By Lemma 2 this appears in the second term of Eq. (3.37).

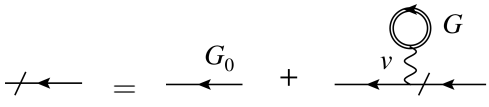
Step 2 We show that on the RHS no diagram appears twice. The first term is only one single diagram, which is clearly not included in the other two terms. Those also do not lead to identical diagrams, as the second term becomes disconnected by cutting the interaction line next to the first GF line, while the other remains connected.

Step 3 Finally it is obvious that on the RHS there is no disconnected diagram. We conclude that each diagrams of the LHS appears exactly once on the RHS. □

It is interesting to note that there is no algegraphic equation that determines G in terms of G_0, which can be seen from the fact that the number of diagrams grows factorially, while algegraphic equations would produce at most exponentially many diagrams.

3.2.4 The Hartree Green's function

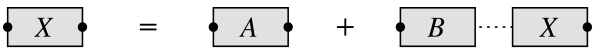
Now we would like to pull part of the "complication" into the links, i.e. we would like to find a way to express G in terms of so-called dressed quantities and reduce the number of diagrams. Towards this end we define the Hartree GF as the one satisfying the equation we get when dropping the last term in Eq. (3.11)

G_{H} = 

. (3.40)

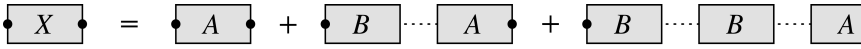
One could also have replaced the GF in the loop by the bare or the Hartree GF itself, but this way of doing it turns out to be more useful. The solution to this equation is provided by

Lemma 4. *Let A and B be linear combinations of diagrams, where B does not include the diagram with one vertex only. An algegraphic equation of the form*



, (3.41)

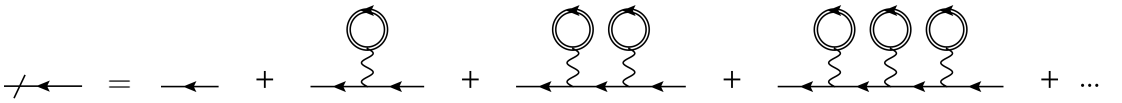
(Dyson equation) where A and X may have more open ends than indicated has the unique solution



(3.42)

Proof. That this is a solution is easily verified. That this is the only solution follows from Theorem 3. □

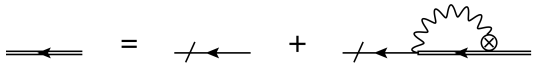
So the Hartree GF can be written as



(3.43)

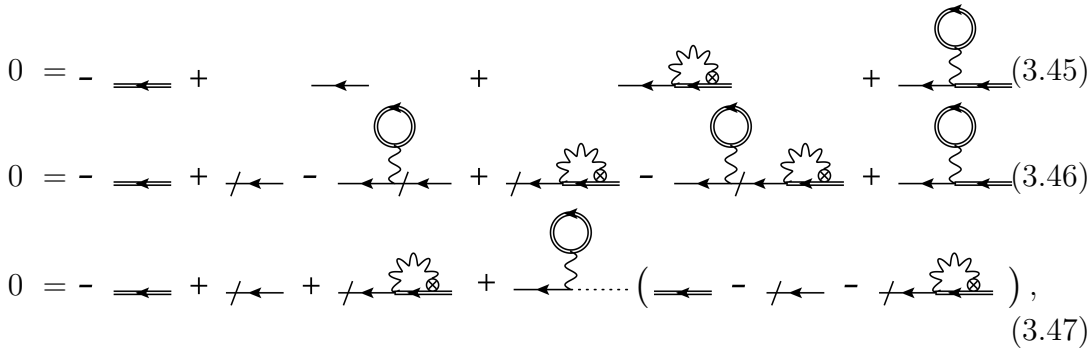
We can now get rid of the second term in Eq. (3.11), at the cost of replacing G_0 by G_{H}.

Lemma 5. *The GF satisfies*



(3.44)

Proof. Using Eq. (3.40) in Eq. (3.11) leads to



(3.45)

(3.46)

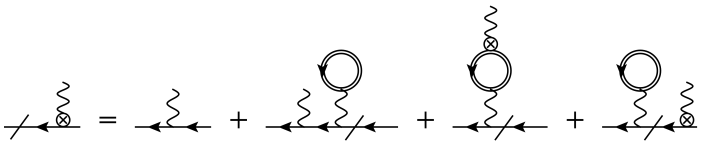
(3.47)

of which the solution to Eq. (3.44) clearly is a solution. □

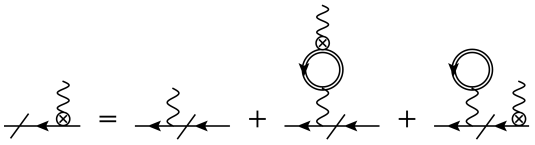
3.2.5 The modified vertex insertion operator and the screened interaction

Note that the vertex insertion operator still requires us to express G_{H} in terms of G_0 before it can be applied. To analyze this problem, we evaluate the action of the vertex insertion operator on a single Hartree line. In anticipation of the result we

link this to an interaction line. Acting on both sides of Eq. (3.40) provides

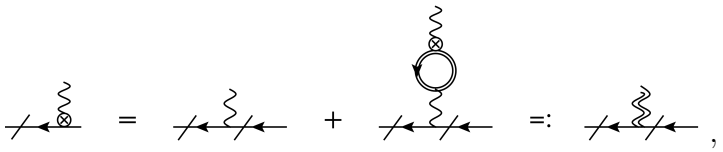


(3.48)



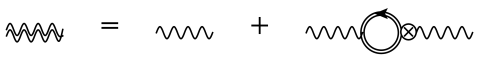
(3.49)

This equation has the structure needed in Lemma 4, where A is the sum of the first two terms on the RHS. Using Eq. (3.43) in the solution, we obtain



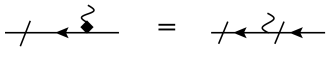
(3.50)

where we defined the so-called screened interaction $W :=$  as



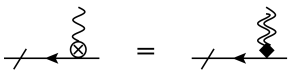
(3.51)

This result motivates the definition of a **modified vertex insertion operator** that acts on Hartree GF in an analogous way to how the ordinary vertex insertion operator acts on bare GFs



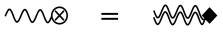
(3.52)

We also want it to act just in the same way, i.e. it shall satisfy a product rule and map bare interaction lines to zero (see Def. 11 (p. 22)). Then Eq. (3.50) can be written as



(3.53)

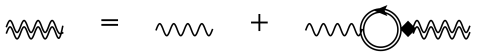
As both operators satisfy a product rule this property can be extended to any diagram. Hence the operator equation



(3.54)

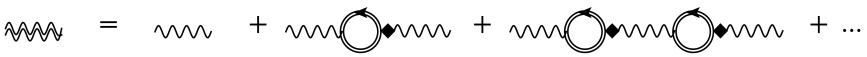
holds, i.e. the result of linking a bare interaction line to the expansion obtained from acting with the ordinary vertex insertion operator is the same as the result of linking

a screened interaction line to what is obtained from acting with the modified vertex insertion operator. We will further discuss the range of validity of this identity in a moment. Using Eq. (3.54) in Eq. (3.51) leads to



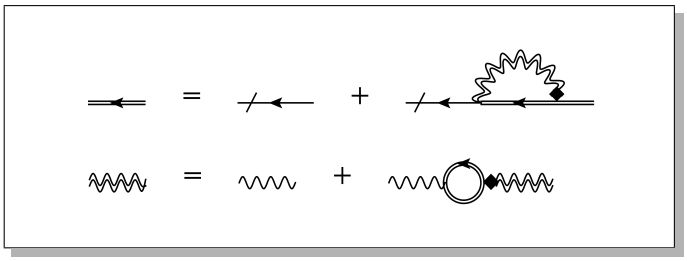
(3.55)

This equation has the structure of a Dyson equation, as needed in Lemma 4 (p. 30). Hence the solution can be written as



(3.56)

We use Eq. (3.54) also on Eq. (3.44) to find



(3.57)

(3.58)

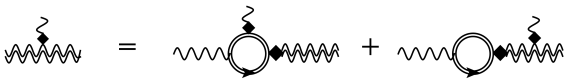
By virtue of Theorem 3 this system of grapherential equations generates expansions $G[G_H, v]$ and $W[G_H, v]$ by iteration as the unique solution of this system. By this we mean that the GF and the screened interaction W are obtained as an expansion in graphs built from GF lines labeled G_{H} and the bare interaction. Remarkably it was possible to eliminate any reference to G_0. As we derived these equations, they of course are correct in the sense that inserting the expansion of G_{H} in terms of G_0 and v provides the expansion $G[G_0, v]$ of Def. 9 (p. 19).

Note that we applied the modified vertex insertion operator on G before we actually knew that it can be expanded in terms of G_{H}. However, this can be justified a posteriori, as we could have taken Eqs. (3.57), (3.58) as definitions and then show that those are equal to the objects we defined in Eq. (3.51) and Def. 9 (p. 19).

If we now consider Eq. (3.40) as an equation to determine G_0 in terms of G_{H} and G, we find that it can be solved by virtue of Lemma 4. If we then replace G by its expansion in terms of G_{H} and v, we obtain an expansion of the bare GF in terms of G_{H} and v. This shows that it is possible to write any expansion in terms of G_0 also as an expansion in terms of G_{H}, and hence the modified vertex insertion operator can be applied to any diagram and the identity Eq. (3.54) is always valid (i.e. we need no longer worry about whether or not we can act with the modified vertex insertion operator on a diagram, as it is always possible).

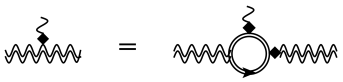
3.2.6 Expansion in terms of the screened interaction

As only W appears in the equation for G, it is natural to ask whether one can also drop the reference to v. In order to do so, one of course needs to work out how the vertex insertion operator acts on W lines. We act with the modified vertex insertion operator on Eq. (3.55) to obtain

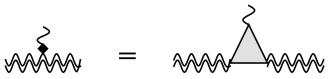


(3.59)

Note that the notation does not allow us to distinguish in which order the vertex insertion operators act. However, this is no problem, as the way it is defined allows us to conclude that both orders lead to the same result. This would not be true if both the ordinary and the modified vertex insertion operator appeared at the same time. Eq. (3.59) has the form of a Dyson equation, as needed in Lemma 4 (p. 30). We immediately use Eq. (3.56) to write the solution as

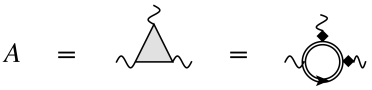


(3.60)



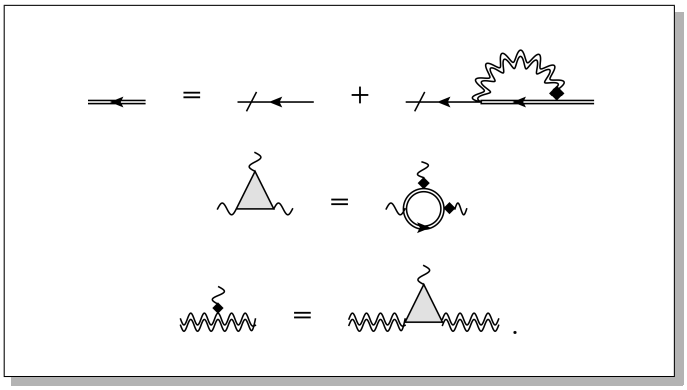
(3.61)

where we introduced a new object



(3.62)

With that we can summarize the action of the modified vertex insertion operator on a diagram given in terms of G_{H} and W lines as follows: One picks out one of the edges of the graph. If it is a Hartree GF it is replaced by a vertex and two Hartree GFs, as before. If it is an interaction line W, one replaces it by two such interaction lines, joined by the object A. Finally one has to sum over each possible edge one can pick out. We summarize the equations we derived:



(3.63)

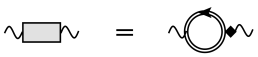
(3.64)

(3.65)

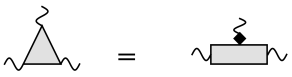
This system of equations allows us to obtain $G[G_H, W]$ and $A[G_H, W]$ by iterating. In this iteration process one needs to use the current expansion of A in the way illustrated by Eq. (3.65) whenever the modified vertex insertion operator acts. This action does of course increase the number of internal vertices of the obtained diagram relative to the diagram the operator acts on. Hence the conditions of Theorem 3 are meet[1].

3.2.7 Expansion in terms of the interacting Green's function and the screened interaction

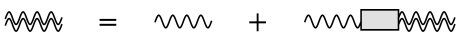
It is worthwhile to introduce a symbol for the already used polarizability

P =  . (3.66)

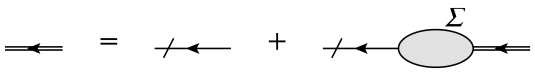
In terms of this, A can be written as

 (3.67)

and the Dyson equation for the screened interaction becomes

 (3.68)

Remarkably G satisfies a Dyson equation as well, reading

 . (3.69)

This equation can be understood as a definition for the selfenergy (SE) Σ, as the following Lemma shows

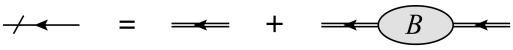
Lemma 6. *The the expansion Σ that solves Eq. (3.69) exits and is unique.*

[1] More accurately, in order to apply the theorem, one needs to regards these equations as a system of the first two equations only, where the operators D_1 and D_2 appearing in the definition of a nice system of grapherential equations are acting with the vertex insertion operator, using the *current* set of diagrams in A in the way indicated by the third equation.

Proof. As we discussed after Eq. (3.58), we can expand the GF in terms of G_{H} and v. In this expansion the diagram with one G_{H}-line only has a prefactor of one. Hence we can write G as

, (3.70)

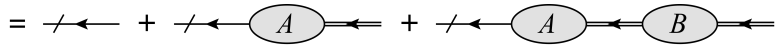
with an appropriate expansion A. Now comes the crucial step: We can understand this as a nice grapherential equation for G_{H}. This provides us with an expansion of G_{H} in terms of G an v. For this expansion the prefactor of the diagram with one G-line also has a prefactor of one. This allows us to write G_{H} as



, (3.71)

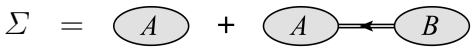
where B is another expansion. Now we have all parts we need to construct the selfenergy. Starting from Eq. (3.70) we get

(3.72)



, (3.73)

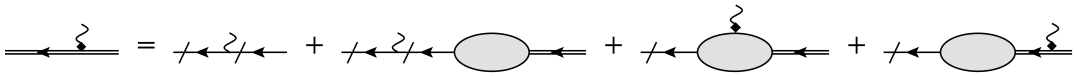
where we used Eq. (3.71). With this we see that



(3.74)

is a solution to Eq. (3.69) □

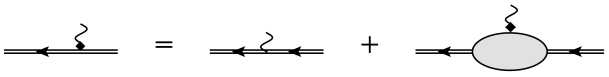
We can now work out the action of the modified vertex insertion operator on G using Eq. (3.69)



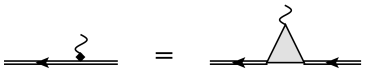
(3.75)

. (3.76)

Again we obtained a Dyson equation. In the resulting solution (see Lemma 4) we use the solution of Eq. (3.69), which is given by Lemma 4. The result is

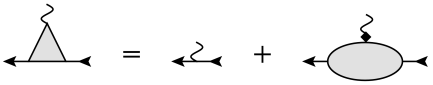


(3.77)



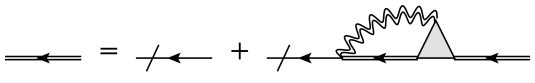
, (3.78)

where we defined the socalled dressed vertex

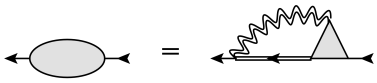
Γ =  . (3.79)

Note that we distinguish the symbols for A and Γ by the open ends only. The first term on the RHS of Eq. (3.79) denotes the graph that consists of a single vertex (endowed with three labels) and no links.

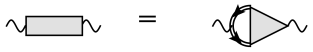
We obtained the remarkable result that it is possible to interpret the action of the modified vertex insertion operator on the full GF again as a vertex insertion operator that replaces one dressed GF by two and an appropriately defined vertex function Γ. Now we use this result in Eq. (3.69) leading to

 . (3.80)

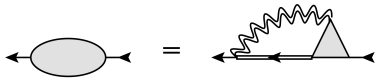
comparison with Eq. (3.69) allows us to conclude

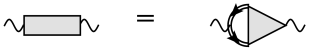
 . (3.81)

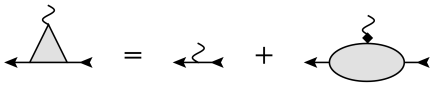
Finally we use Eq. (3.78) in Eq. (3.66) to obtain

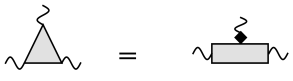
 . (3.82)

We summarize the so-called Hedin equations

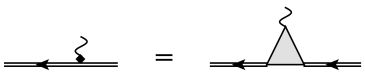
Σ =  (3.83)

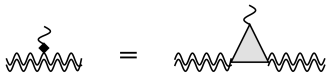
P =  (3.84)

Γ =  (3.85)

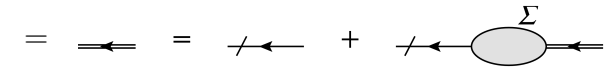
A =  (3.86)

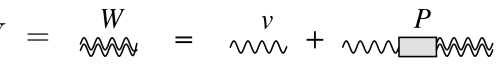
the rules on how the vertex insertion operator acts on different links

 (3.87)

 (3.88)

and the Dyson equations

G =  (3.89)

W =  . (3.90)

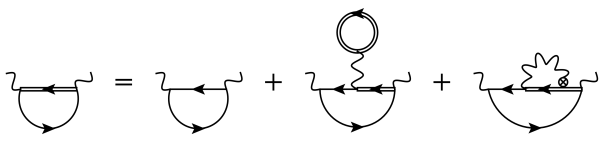
These equations allow us to construct expansions $\Sigma[G,W], P[G,W], \Gamma[G,W]$ and $A[G,W]$ by iteration. The first few iterations have already been carried out by Hedin [30], however, without commenting on whether or not this procedure converges. In a recent book by G. Stefanucci and R. van Leeuwen [22] it is stated that "[the authors] are not aware of any proof that such an iterative scheme converges, nor that it generates all possible skeleton diagrams for the self-energy". Now we fill this gap.

Theorem 4. *Iterating Eq. (3.83) - Eq. (3.86), using Eqs. (3.87), (3.88) for the evaluation of the modified vertex insertion operator, generates (as a limiting expansion) the unique set of expansions* $(\Sigma[G,W], P[G,W], \Gamma[G,W], A[G,W])$ *that solves this set of equations. This solution is correct in the sense that inserting the expansions* $G[G_0,v]$, $G_{\mathrm{H}}[G_0,v]$ *and* $W[G_0,v]$, *defined in Eqs. (3.40), (3.51) and Def. 9 (p. 19) into these expansions* $\Sigma[G,W]$ *and* $P[G,W]$ *and in Eqs. (3.89), (3.90) leads to valid equations.*

Proof. That the unique solution is obtained by iteration is guaranteed by Theorem 3. The correctness of this unique solutions follows from the fact that we derived the equations. □

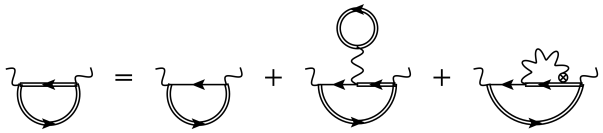
3.2.8 Counterexamples

Every single equation we introduced in the last sections had a unique solution. This might lead to the impression that that is the only possible outcome. One might think: Given that there was no approximation done, what else other than the correct result should come out? To illustrate that things are not that simple, we consider a little example. One may ask: what happens, if we start from Eq. (3.11) and link it to another non-interacting GF? We obtain the new equation

 , (3.91)

which is no longer a nice grapherential equation in the sense of Def. 12 (p. 23). As we derived it from an equation that is valid for G the full GF is still a solution to

it, but is it unique? In this case the answer turns out to be yes, while if we linked with the interacting GF to obtain



, (3.92)

the trivial solution $G = 0$ would appear as an extra solution. Furthermore, if we subtract Eq. (3.92) from Eq. (3.91), we obtain an equation that has two non-trivial solutions, namely the full GF and the bare GF, as is easily verified. It is indeed not difficult to construct equations with any desired finite set of extra solutions (e.g. if we decorated the bare GF in Eq. (3.91) with, say, a Hartree insertion, then the extra solution after taking the difference of the two equations would be this decorated GF). We hope this illustrates not only that the question on uniqueness is not a trivial one, but also that the conditions needed for our Theorems on uniqueness cannot simply be dropped, i.e. the example illustrates that it is not sufficient to demand that there is a diagram in which the unknown object appears only once. Instead we demand that it appears without any decorations (see the definition of a nice grapherential equation, Def. 12 (p. 23)).

3.3 Many body theory

Up to now we have only talked about equations for graphs. We will now establish the link to the many body theory of solid state physics. We stress again that the theory presented in the previous paragraphs is mathematically completely strict. Once we evaluate Feynman Diagrams, however, it is no longer clear whether or not the emerging series do converge. There are many physical examples where it has been shown [28, 23, 24, 25, 26, 29] or argued [27] that a perturbative expansion does not converge. Even though divergence seems to be the rule rather than the exception, this does not at all mean that the perturbative treatment of quantum mechanical problems is useless. Usually the first couple of terms improve a calculated result, even when the series does not converge [28, 44, 25]. Often a system can be described by expanding the underlying wave function in a finite basis set. Once this approximation is done the eigenvalues and eigenstates of the Hamilton are analytic functions of a parameter that appears in the Hamiltonian [45] and consequently perturbative expansions do then converge at least in some finite range of the interaction parameter. Note, however that this remarkable result is no longer valid if there are two or more parameters that are varied simultaneously, as the example of the conical intersection shows [46, 47].

We will now briefly review the parts of MBPT that are relevant for us. We will closely follow Ref. [22] in notation and presentation.

3.3.1 Equations of motion

We want to describe a quantum mechanical systems of electrons in a solid or molecule. In the non-relativistic limit and the Born Oppenheimer approximation. Such a problem can be described by a Hamiltonian operator of the form

$$H(t) = \int \mathrm{d}x\ \psi^\dagger(x)\ h(x,t)\ \psi(x) + \int \mathrm{d}x\mathrm{d}y\ v(x,y)\ \psi^\dagger(y)\psi^\dagger(x)\psi(x)\psi(y), \quad (3.93)$$

where $\psi^\dagger$ and ψ are electron creation and annihilation operators, respectively, $x = (r,\sigma)$ represents a real space coordinate and a spin index, h is the kinetic energy operator plus a possibly time-dependent external potential and v is a two body interaction (e.g. the Coulomb interaction).

There exist several Green's function formalisms which can include a finite temperature and/or a time-dependent perturbation or not. However, "all [these] Green's function formalisms usually treated as independent naturally follow from a single one" [22]. We will need the one body GF and the two body GF, defined as [21, 22]

$$G(1,2) = -\mathrm{i}\ \frac{\mathrm{Tr}\left[\mathrm{T}\ \mathrm{e}^{-\mathrm{i}\int \mathrm{d}\tau\, H(\tau)}\ \psi(1)\psi^\dagger(2)\right]}{\mathrm{Tr}\left[\mathrm{T}\ \mathrm{e}^{-\mathrm{i}\int \mathrm{d}\tau\, H(\tau)}\right]} \quad (3.94)$$

$$G^{(2)}(1,2,3,4) = -\frac{\mathrm{Tr}\left[\mathrm{T}\ \mathrm{e}^{-\mathrm{i}\int \mathrm{d}\tau\, H(\tau)}\ \psi(1)\psi(2)\psi^\dagger(4)\psi^\dagger(3)\right]}{\mathrm{Tr}\left[\mathrm{T}\ \mathrm{e}^{-\mathrm{i}\int \mathrm{d}\tau\, H(\tau)}\right]}, \quad (3.95)$$

where $1 = (\tau_1, x_1) = (\tau_1, r_1, \sigma_1)$ includes a space and spin coordinate and a contour time argument τ_1 (running from zero to infinity and back on the real axis and then from zero to $-\mathrm{i}\beta$ on the imaginary axis), T is the contour-time ordering symbol and $\psi^\dagger$ and ψ are the fermion creation and annihilation operators (the time argument does not refer to the Heisenberg picture, but is artificially introduced to specify how the time-ordering symbol acts on them). The Hamilton $H(\tau)$ is defined as the one given in Eq. (3.93), whenever τ lies on one of the branches on the real axis. On the remaining branch it is a fixed Hamiltonian, that specifies the statistical average in which the system is assumed to be initially prepared.

Depending on the further assumptions or simplifications one makes, one can drop parts of the contour, leading to the finite temperature Matsubara formalism [17],

the zero-temperature formalism [17], the (original) Keldysh formalism [48] or the Konstantinov-Perel' formalism [22].

Using the Heisenberg equation of motion for the field operators, it is straight forward to derive the so-called equation of motion for the one body Green's function [17]

$$\left(\mathrm{i}\frac{\mathrm{d}}{\mathrm{d}\tau_{\mathit{1}}} - h(\mathit{1})\right) G(\mathit{1},\mathit{2}) = \delta(\mathit{1}-\mathit{2}) - \mathrm{i}\int \mathrm{d}\mathit{3}\; v(\mathit{1},\mathit{3})\; G^{(2)}(\mathit{1},\mathit{3},\mathit{2},\mathit{3}^{+})\;. \tag{3.96}$$

For a non-interacting system this equation defines a mathematical GF

$$\left(\mathrm{i}\frac{\mathrm{d}}{\mathrm{d}\tau_{\mathit{1}}} - h(\mathit{1})\right) G_0(\mathit{1},\mathit{2}) = \delta(\mathit{1}-\mathit{2})\;. \tag{3.97}$$

The analogous equation for the Two-Body GF involves also the Three-Body GF and so on. This is the famous Martin-Schwinger hierarchy [22]. The appeal of these equations lies in the fact that very simple approximations on a high level of Green's function, can already provide very accurate approximations for lower level Green's functions.

We want to proceed on a different path, that remains exact in principle. Instead of expressing the two-body GF by higher and higher order GFs, one can express it by a functional derivative of the one-body GF with respect to to an artificial (contour-) time-dependent potential $S(x,\tau)$ that one introduces by adding

$$\int \mathrm{d}x\; S(x,\tau)\psi^{\dagger}(x)\psi(x) \tag{3.98}$$

to the Hamiltonian. Note that S may take different values on different branches of the contour, even when they belong to the same real-time argument. The action of the functional derivative operator can now be evaluated easily, as all operators on which the time ordering symbol acts can be treated as if they would (anti-) commute. We find [22]

$$\frac{\mathrm{d}G(\mathit{1},\mathit{2})}{\mathrm{d}S(\mathit{3})} = -G^{(2)}(\mathit{1},\mathit{3},\mathit{2},\mathit{3}^{+}) + G(\mathit{1},\mathit{2})\; G(\mathit{3},\mathit{3}^{+}). \tag{3.99}$$

Note that in the non-interacting case this simplifies to

$$\frac{\mathrm{d}G_0(\mathit{1},\mathit{2})}{\mathrm{d}S(\mathit{3})} = -G_0(\mathit{1},\mathit{3})\; G_0(\mathit{3},\mathit{2}). \tag{3.100}$$

If we now insert Eq. (3.99) into Eq. (3.96), multiply with G_0 from the left, do an integration by parts and use that G_0 also satisfies an equation analogous to Eq. (3.97) for its second argument, we obtain

$$G(1,2) = G_0(1,2) - \mathrm{i}\int \mathrm{d}3\mathrm{d}4\ G_0(1,3)\ v(3,4)\ G(4,4^+)\ G(3,2) \tag{3.101}$$

$$+\,\mathrm{i}\int \mathrm{d}3\mathrm{d}4\ G_0(1,3)\ v(3,4)\,\frac{\mathrm{d}G(3,2)}{\mathrm{d}S(4)}. \tag{3.102}$$

In the next section we will see that this equation is the one corresponding to the Graph equation we used to define the GF in terms of graphs. We stress again that this derivation is valid in all GF formalisms mentioned above. The only thing that changes is the contour along which one integrates when evaluating the integrals.

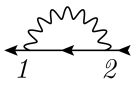
3.3.2 Evaluation of Feynman diagrams

We now link the previous section and the section about graph theory. This link is established by the Feynman rules, which can be found in many textbooks [17].

Surprisingly few authors highlight that Feynman Diagrams are nothing else than a visual representation of tensor contractions [49, 50]. In this work every link should be understood as a rank two tensor (as introduced in the previous section) and the bare vertex as a rank three tensor, namely

$$\gamma(1,2,3) = \delta(1,2)\delta(1,3)\ . \tag{3.103}$$

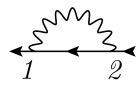
One can then build up every diagram by combinations and contractions of these building blocks. The combination (contraction) of two (one or two) graphs corresponds to the outer product (the contraction) of the tensors corresponding to the graphs. Consider for example the diagram



. (3.104)

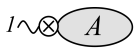
It consists of two bare vertices, a bare GF and an interaction line. The tensors corresponding to the links were already named $G_0(1,2)$ and $v(1,2)$ in the previous section. There is some choice on which tensor indices are considered covariant and which ones are considered contravariant. The only restriction is that the indices of the bare vertex must match the indices of the interaction line and the incoming and outgoing GFs. We could for example take all indices of the vertex to be contravariant. Then all indices of the links have to be covariant. As this choice does not matter, there is no need to distinguish upper and lower indices. We will come

back to this point when we introduce a basis in which we expand the tensors. The four tensors in Eq. (3.104) with a total of ten indices are contracted four times. As each contraction eliminates two indices, there are two external indices left, which are labeled. For the moment the tensor indices live in a continuous space, namely the spin-space-contour-time. Contractions can be expressed as integrations over this space:



$$= \mathrm{i} \int \gamma(1,3,5)\, G_0(3,4)\, v(5,6)\, \gamma(4,2,6)\, \mathrm{d}3456\ . \tag{3.105}$$

As in the previous chapter this notation is understood as a four fold integration, each over the real space and the time contour and also includes a spin summation. Finally, each diagram is associated with a prefactor of $\mathrm{i}^k(-1)^l$, where k is the number of interaction lines and l is the number of GF-loops (i.e. the number of closed paths built from the GF lines alone). We mention that there is another numerical prefactor, related to the numbers of equivalent edges in a diagram [17]. However, it does not appear in the present work, as we are only dealing with connected diagrams with at least one open end. These diagrams have the special property that there are no equivalent edges and hence that this prefactor is always simply one. With that it should be clear how to translate Feynman diagrams into tensor contractions and vice versa. Using Eq. (3.100) we can also see that the vertex insertion operator 11 is related to the functional derivative with respect to S as



$$= -\frac{\mathrm{d}A}{\mathrm{d}S(1)}, \tag{3.106}$$

where A is understood as a diagram in terms of G_0 and v and may have any number of open ends.

With that one can see that the final equation we derived in the previous section Eq. (3.102) is equivalent to the definition of the GF we gave in terms of diagrams Eq. (3.11). Hence Theorem 2 actually *proves* the Feynman rules above. Note in particular that the rule for the prefactor cannot be chosen in any other consistent way if we want Eq. (3.102) to be equivalent to Eq. (3.11). As we already mentioned there is no guarantee that this series for G does converge. That means that even though our proof is valid as long as one is only speaking of diagrams, one cannot always evaluate the diagrams and sum up the series.

We can now use the rules described in this paragraph to write the grapherential equations of the previous section in terms of functions instead of graphs. It was already mentioned that Eq. (3.11) is equivalent to Eq. (3.102). Rewriting Eqs. (3.83)-

(3.90) yields

$$\Sigma(1,2) = \mathrm{i}\int G(1,4)W(1^+,3)\Gamma(4,2,3)\mathrm{d}3\mathrm{d}4 \tag{3.107}$$

$$\Pi(1,2) = -\mathrm{i}\int G(2,3)G(4,2^+)\Gamma(3,4,1)\mathrm{d}3\mathrm{d}4 \tag{3.108}$$

$$\Gamma(1,2,3) = \Gamma_0(1,2,3) + \frac{\delta\Sigma(1,2)}{\delta V(3)} \tag{3.109}$$

$$A(1,2,3) = \frac{\delta\Pi(1,2)}{\delta V(3)}, \tag{3.110}$$

with the functional derivative acting as

$$\frac{\delta G(1,2)}{\delta V(3)} = \int G(1,4)G(5,2)\Gamma(4,5,3)\mathrm{d}4\,\mathrm{d}5 \tag{3.111}$$

$$\frac{\delta W(1,2)}{\delta V(3)} = \int W(1,4)W(5,2)A(4,5,3)\mathrm{d}4\,\mathrm{d}5. \tag{3.112}$$

and the Dyson equations

$$G(1,2) = G_{\mathrm{H}}(1,2) + \int G_{\mathrm{H}}(1,3)\Sigma(3,4)G(4,2)d(3)d(4) \tag{3.113}$$

$$W(1,2) = v(1,2) + \int v(1,3)\Pi(3,4)W(4,2)d(3)d(4). \tag{3.114}$$

Eq. (3.52) becomes

$$\frac{\delta G_{\mathrm{H}}(1,2)}{\delta V(3)} = \int G_{\mathrm{H}}(1,3)G_{\mathrm{H}}(3,2)\ . \tag{3.115}$$

Apart from inconsequential sign conventions one can choose when writing down these equations, the mapping between Feynman diagrams and integrals over functions is one-to-one. Accordingly, the theorems developed do equally well apply to these functional integro-differential equations, i.e. symbolically iterating these equations generates formal series expansions of the appearing objects in terms of G and W, and by Theorem 3 one obtains a unique formal expansion. We will come back to this in the next chapter.

For later reference, we note that, using the chain rule, we can rewrite Eqs. (3.109),

(3.110). If we use Eqs. (3.111), (3.112), we obtain

$$\Gamma(1,2,3) = \Gamma_0(1,2,3) + \int \Gamma(4',5',3)G(4,4')G(5',5)\frac{\delta\Sigma(1,2)}{\delta G(4,5)}\mathrm{d}4\,\mathrm{d}4'\mathrm{d}5\,\mathrm{d}5' \tag{3.116}$$

$$+ \int A(4',5',3)W(4,4')W(5',5)\frac{\delta\Sigma(1,2)}{\delta W(4,5)}\mathrm{d}4\,\mathrm{d}4'\mathrm{d}5\,\mathrm{d}5' \tag{3.117}$$

$$A(1,2,3) = \int \Gamma(4',5',3)G(4,4')G(5',5)\frac{\delta\Pi(1,2)}{\delta G(4,5)}\mathrm{d}4\,\mathrm{d}4'\mathrm{d}5\,\mathrm{d}5' \tag{3.118}$$

$$+ \int A(4',5',3)W(4,4')W(5',5)\frac{\delta\Pi(1,2)}{\delta W(4,5)}\mathrm{d}4\,\mathrm{d}4'\mathrm{d}5\,\mathrm{d}5'. \tag{3.119}$$

If we instead used Eq. (3.115) the equation for the vertex would read [51, 32]

$$\Gamma(1,2,3) = \Gamma_0(1,2,3) + \int G_{\mathrm{H}}(4,3)G_{\mathrm{H}}(3,5)\frac{\delta\Sigma(1,2)}{\delta G_{\mathrm{H}}(4,5)}\mathrm{d}4\,\mathrm{d}5, \tag{3.120}$$

where v is kept fixed in the functional derivative. Note that in the expression above, I used the symbol $\frac{\delta}{\delta G}$ for the partial functional derivative with respect to G, while the (maybe a bit unconventional) notation $\frac{\mathrm{d}}{\mathrm{d}V}$ refers to the total functional derivative.

This relationship has first been derived in ref. [51] in order to provide a set of equations which allows to construct the expansions $G[G_H, v]$, $W[G_H, v]$, $\Sigma[G_H, v]$, $\Pi[G_H, v]$ and $\Gamma[G_H, v]$ by iteration. Note that we derived a different version of Hedin's equations, which leads to expansions in terms of G and W when iterated. However, the theorems proven cover both sets of equations and allow to conclude that in both cases the obtained expansion is unique and independent on the starting expansions.

3.3.3 Martin-Schwinger-Hierarchy

Instead of closing the Martin-Schwinger-Hierarchy in the way done in Eq. (3.102), one can express the one-body-GF in terms of the two-body-GF, which in turn is expressed by the three-body-GF and so on. Finally one can close the hierarchy, by employing a relation analogous to Eq. (3.99). This system of equation can than be rewritten in a fashion analogous to Eq. (3.102). For the obtained system of equations one can again ask whether there is a unique solution, and how one can construct it. Theorem 3 is general enough to be applicable to this situation as well. Hence

for any level of the Martin-Schwinger-Hierarchy, iterating the equations leads to the unique diagrammatic expansions of the appearing objects that solve these equations. Further more this remains valid for a large class of modified equations that one might want to introduce, by introducing for example objects like the reducible particle-hole-propagator etc., whenever the (easily checked) conditions of Theorem 3 are satisfied.

Chapter 4

Hedin's equations in discrete space and time

In this chapter, as in the previous, we consider the same equations, however in discretized space and time in a finite space-time-volume. We hope of course that any quantity of interest can be recovered in the continuum limit. However, we have to be aware that we cannot describe certain properties one might be interested in. Phase transitions for instance do not appear in such a discrete finite system, even if we leave the time variable continuous [52, 45]. There is also no continuous spectrum and the total energy of the system is limited.

One can also have some basis set in mind, in which all functions are expanded. This does not change the form of any of the equations that we can express diagrammatically as tensor contractions are basis-independent [50]. Of course the tensors themselves are basis-dependent and even the bare vertex takes different values in a different basis. In energy-momentum space it is just one delta function (if we are dealing with a translationally invariant system), while it is the product of two delta functions in real space and time.

4.1 Hedin's equations in discrete space and time

4.1.1 One Point Model

It is instructive to study the solutions to Hedin's equations with the simplification of taking only one point in space, spin and time, i.e. $N = 1$. This framework has recently been investigated by several authors [31, 32, 33, 34]. One simply drops the tensor indexes appearing in the equations and regards each quantity as a single complex number. Hence one is left with four algebraic equations and one ordinary differential equation. Even though this no longer corresponds to any physical system, we can learn a lot about general properties of the perturbation series and the

structure of the solutions to the equations.

The equations now read

$$G = G_{\mathrm{H}} + G_{\mathrm{H}} \Sigma G \tag{4.1}$$

$$W = v + v \Pi W \tag{4.2}$$

$$\Sigma = \Gamma G W \lambda \tag{4.3}$$

$$\Pi = l\ \Gamma G G \lambda \tag{4.4}$$

$$\Gamma = \lambda + \lambda G^2 \frac{d\Sigma}{dG}, \tag{4.5}$$

where $l = +1$ corresponds to bosons and $l = -1$ to fermions and $' = \frac{d}{dG_{\mathrm{H}}}$. One can now eliminate all but one dependent variable, say Σ. The resulting differential equation is

$$\Sigma(1 - G^2 \frac{d\Sigma}{dG}) = Gv + l\Sigma G^2 v. \tag{4.6}$$

Before we continue to solve this equation, we illustrate the behavior of the solutions on a simpler example that we can solve explicitly but still has the same features. Consider the differential equation

$$f = x^2 f' - 2x^3 + x^2. \tag{4.7}$$

As in Eqs. (4.5), (4.6), we have multiplied the derivative of the function with the square of the independent variable. The last two terms were added to make the solutions not too simple. One easily verifies that the solutions are

$$f = A\, \mathrm{e}^{-\frac{1}{x}} + x^2, \tag{4.8}$$

where A is a constant of integration. These solutions are plotted in Fig. 4.1. The point $x = 0$ obviously needs to be treated carefully. If we are looking for solutions of the differential equation that include that point, then there is only one choice for the constant of integration in the region $x < 0$, while any choice of integration constant provides a solution in the region $x > 0$. So if $x = 0$ is included in the region in which we want to solve the equation, the solutions are

$$f = \begin{cases} x^2, & x \leq 0 \\ A\, \mathrm{e}^{-\frac{1}{x}} + x^2, & x > 0\ . \end{cases} \tag{4.9}$$

Note that all these functions are infinitely often continuously differentiable, even at $x = 0$. In this sense they are perfectly smooth functions, just as x^2 is. The

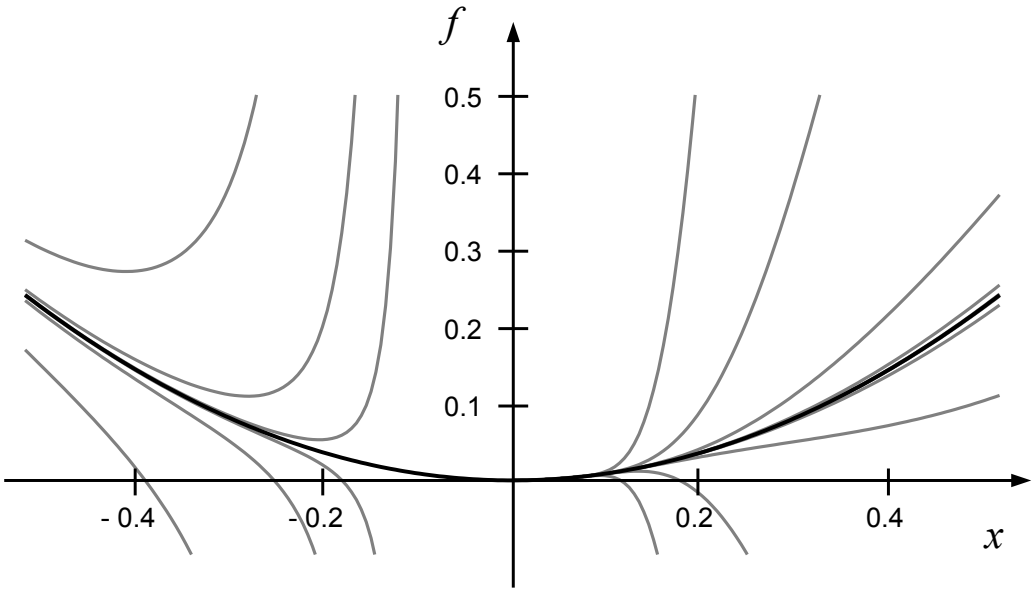


Figure 4.1: Behavior of the Solutions of Eq. (4.7). Coming from negative values of x, there is only one non-diverging solution (black parabola). Coming from positive values of x all solutions tend to the same value. Also all derivatives tend to the same value as x approaches zero.

distinctive feature of the latter function is that it is the only solution that is also analytic. It is interesting to note that if one wanted to solve the equation as an initial value problem, starting with some given value for the function $f(x_0) = f_0$ at some negative x_0, then there is only one value for which the equation can be solved past $x = 0$. Even then, the continuation of the function past $x = 0$ is not unique, as all solutions in Eq. (4.9) are valid continuations, not only the one for $A = 0$.

Let us come back to Eq. (4.6). If we assume that Σ can be expanded in a power series, we can obtain a recurrence relation for the expansion coefficients [32, 33, 53]. When solving "well behaved" differential equations one can usually choose one of the expansion coefficients and the others follow by the recurrence relation. This is not the case for this differential equation (due to the G^2 in front of the derivative term). There is *only one* solution for the recurrence relation and consequently this solution is the only *analytic* solution to the differential equation, just as in the example we solved in the previous paragraph. We will later generalize this result to the case, where N is arbitrary and the equations are partial differential equations.

This unique analytic solution is surprisingly simple:

$$\Sigma = Gv \tag{4.10}$$

and hence $\Sigma + V_H = 0$. Consequently G is in fact the non-interacting Green's function $G = G_0$ (also excluding the Hartree potential).

What can we learn from that? Since the expansion coefficients are equal to the number of the various diagrams [32] it tells us that the number of (self-energy plus hartree)-diagrams with even and odd loop numbers cancel. In other words the number of diagrams with odd number of fermion loops equals the number with even fermion loops. This holds for both expansions $\Sigma[G_0, v]$ and $\Sigma[G, v]$. It is interesting to note that the Dyson equation

$$G = G_{\mathrm{H}} + G_{\mathrm{H}} \Sigma G \tag{4.11}$$

does not have a unique solution for G, even when we provide the exact functional $\Sigma[G] = Gv$. One of the two solutions to this quadratic equation tends to infinity for $v \to 0$.

Remarkably, the situation is qualitatively different for bosons, i.e. $l = +1$. Here the solution to the recurrence relation mentioned above yields expansion coefficients that grow factorially [32] and hence there is no analytic solution to the equations.

Approach without self energy - solved for one point

We now analyze Eq. (3.102) for the case of $N = 1$. In contrast to the situation in the previous paragraph, the analysis of analyticity is more difficult. However, it is possible to solve the equation explicitly and then address this question directly. This is unfortunately not as easily transferable to the general case with arbitrary N. The equation reads

$$G \; = G_0 \; + \; l \; G_0 \; \underbrace{v \; G}_{V_{\mathrm{H}}} \; G \; + G_0 \; \phi \; G \; + \; G_0 \; v \; G', \tag{4.12}$$

There is of course again a family of solutions. It turns out that in this family there is exactly one solution that is analytic with respect to v, while all others are not. It seems reasonable to identify this unique solution with the physical one, since this is known to be analytic, at least in the continuous time and discrete, finite space case [45].

As we will see in a moment the unique analytic solution to Eq. (4.12) is simply the non-interacting Green's function [54] (and hence the result is consistent with the one of the previous paragraph)

$$G = (1 - G_0 \phi)^{-1} G_0. \tag{4.13}$$

We now calculate the full solution of Eq. (4.12) (this has also been done in Ref. [54]). First we simplify the equation by a transformation of variables

$$g = \sqrt{v}G, \qquad g_0 = \sqrt{v}G_0, \qquad x = \frac{\phi}{\sqrt{v}} - \frac{1}{g_0}, \qquad \frac{d}{dx} = \sqrt{v}\frac{d}{d\phi}, \tag{4.14}$$

transforming Eq. (4.12) into

$$0 = g' + lg^2 + xg + 1. \tag{4.15}$$

Note that the three parameters G_0, v and ϕ are reduced to one parameter x. This Riccati ODE can be solved with the ansatz

$$g = l^{-1}\frac{u'}{u} \tag{4.16}$$

leading to the linear equation

$$0 = u'' + xu' + lu. \tag{4.17}$$

We now take $l = -1$ and find the solution set to be

$$u(x) = Ax + B\left(e^{-\frac{x^2}{2}} + \sqrt{\frac{\pi}{2}}\ x\ \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)\right), \tag{4.18}$$

where A and B are constants of integration. We insert this into Eq. (4.16) an get

$$g = -\frac{A + B\sqrt{\frac{\pi}{2}}\operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)}{Ax + B\left(e^{-\frac{x^2}{2}} + \sqrt{\frac{\pi}{2}}\ x\ \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)\right)}. \tag{4.19}$$

Now we can identify the set of analytic solutions for G. So let us assume G is analytic with respect to v at least near $(G_0, v, \phi) = (0, 0, 0)$. Then $g = \sqrt{v}G$ will change its sign when rotating v in the complex plane once ($v \to e^{2\pi i}v$). The same is true for $x = \frac{\phi - {G_0}^{-1}}{\sqrt{v}}$. Hence a necessary condition for G to be analytic is that the right hand side of Eq. (4.19) is an *odd* function of x (we actually use that the right hand side is a meromorphic function of x and hence 1) it cannot introduce the minus sign in any different way and 2) it is odd everywhere once we know it is odd somewhere). This however leaves only two choices: either A or B must be zero (otherwise the function has a finite value at $x = 0$ being a contradiction to being odd). The latter choice lead to the already provided solution of the non-interacting Green's function, while the other has a singularity at $v = 0$.

Of cause I also ran into dead ends. I spend almost a year on the attempt to solve two or three point models, using Borel summation and other integral transformations until I finally realized, that the problem I was trying to solve becommes trivial, if one only chooses the way the space is discretized in a clever way. Without going into the details, the result was that this made the solution of the N-point model very simple. Unfortunately, the next step of taking the $N \to \infty$ limit did not provide any new insights, as the only thing one could get would be that the solution is given by many body perturbation theory (if it converges). Ultimately the problem with this approach is, that there is no obvious way to introduce physically justified approximations. And without that, there is litte chance, that the problem can be cast in a way that is easier to solve numerically than any already known approach.

4.1.2 Unique analytic solution of pde

We now consider not only one point in space and time, but rather a finite lattice of space time points. Then Hedin's equations are a set of partial differential equations (pde). Some things we have learned in the discussion of the one-point-model can be generalized to this case as well.

Theorem 5. *Consider Hedin's equations in finite and discrete space and time. Assume that the selfenergy $\Sigma[G_{\mathrm{H}}]$ is not singular at $G_{\mathrm{H}} = 0$. Then Hedin's equations have no more than one solution $G[G_{\mathrm{H}}]$ that is analytic with respect to G_{H} near $G_{\mathrm{H}} = 0$.*

The idea of the proof is the same as for Theorem 3.

Proof. For technical reasons we eliminate the equation for the polarization by plugging it into the equation for the screened interaction. The remaining equations are

$$G = G_{\mathrm{H}} + G_{\mathrm{H}} \Sigma G \tag{4.20}$$

$$W = v + v(\Gamma G G) W \tag{4.21}$$

$$\Sigma = l \Gamma G W \tag{4.22}$$

$$\Gamma = \Gamma_0 + G_{\mathrm{H}} G_{\mathrm{H}} \frac{\partial \Sigma}{\partial G_{\mathrm{H}}}, \tag{4.23}$$

where we did not write out the tensor indexes and the associated contractions. Now we assume $G[G_{\mathrm{H}}]$ is analytic.

Step 1: We show that G can be written as $G = \tilde{G} G_{\mathrm{H}}$, where $\tilde{G}$ is an analytic function of G_{H} and that also $F = (\tilde{G}, W, \Sigma, \Gamma)$ is analytic.

Eq. (4.20) tells us that $\Sigma = G^{-1} - {G_{\mathrm{H}}}^{-1}$ and hence the selfenergy is a meromorphic function of G_{H}. We also assumed it to have no singularity at $G_{\mathrm{H}} = 0$, hence it is even an analytic function in a neighborhood of $G_{\mathrm{H}} = 0$. We can now also conclude that $\tilde{G} \equiv (1 - G_{\mathrm{H}}\Sigma)^{-1}$ is also analytic and hence $G = \tilde{G} G_{\mathrm{H}}$ can be written as a product of an analytic function and G_{H}. Once we know that the selfenergy is analytic we can conclude the same for the vertex and the screened interaction just by their defining equations Eqs. (4.21), (4.23).

Step 2: We show that Hedin's equations imply a unique series expansion of F in terms of G_{H}.

By step 1 we can write

$$F = \sum_{n=0}^{\infty} \sum_{k_1 \ldots k_n} F_{k_1 \ldots k_n} (G_{\mathrm{H}})_{k_1} \ldots (G_{\mathrm{H}})_{k_n}, \tag{4.24}$$

where each k_i is an index pair $k_i = (1,1), (1,2), \ldots, (N_{\mathtt{bas}}, N_{\mathtt{bas}})$. Now we compare the coefficients of each power of G_{H}. The constant part provides $F[0] = (1, v, 0, \Gamma_0)$. The term linear in G_{H} then provides $F = (1, v, 0, \Gamma_0) + (0, 0, G_{\mathrm{H}} G_{\mathrm{H}} v, 0) + O({G_{\mathrm{H}}}^2)$. In general the coefficients $F_{k_1, \ldots k_n}$ are uniquely determined by the coefficients F, F_{k_1}, $F_{k_1 k_2}$, $\ldots$, $F_{k_1 \ldots k_{n-1}}$. The reason is that each (non-constant) term on the RHS contains at least one factor of G_{H} and the derivative term has a prefactor with two G_{H}'s. □

4.2 GW-type approximations

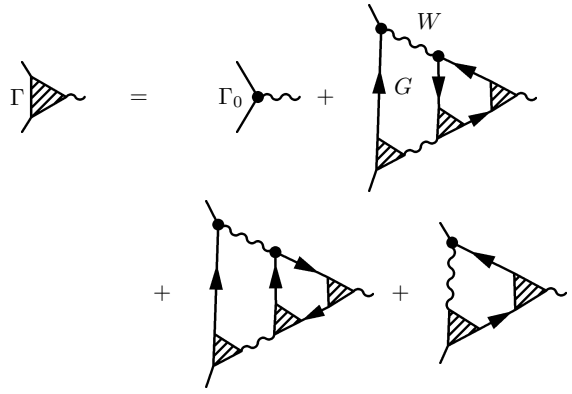
In this section we introduce approximations to the equations that allow us to solve them numerically. The focus is still on the general behavior of the solutions. A natural approximation to express the interacting GF in terms of the non interacting one would be to use MBPT, i.e. Eq. (3.2) and drop all but the first few terms. This way one obtains an explicit expression for G in terms of G_0. Hence it is immediately clear that there is only one solution. However, this can also be understood as an expansion of the GF in terms of the Coulomb interaction. At least in a solid such an expansion is doomed to fail due to the long ranged nature of the interaction. The most severe problem with this is that the charge of the nuclei needs to be compensated by the charge of the electrons. Otherwise the energy density of the system is

formally infinite and hence G_0 is not well defined. To deal with that, one can expand the GF in terms of the Hartree GF and the Coulomb interaction, which is a well defined expansion in which the Hartree diagram no longer appears (see Eq. (3.57)). Even then the quantities one expands in are not well adapted to the problem at hand.At the same time, one can easily compute GFs that describe the system much better than the Hartree GF. Methods for that include the Hartree-Fock approximation [8] and Kohn-Sham Density Functional Theory [3, 4, 5]. The fact that these methods alone provide already remarkably good descriptions in many systems can only be understood, if one views the described system as a system of quasi-particles, i.e. single effective particles that describe the behavior of electrons in combination with a polarization cloud surrounding them [55, 56]. These quasi-particles interact less strongly, as the polarization cloud effectively screens the interaction.

In this light it appears natural to try to use an expansion in terms of a screened interaction. As discussed before one can expand the selfenergy in terms of the screened interaction W and the full GF G. Or equivalently one can use a truncated expansion of the vertex in terms of G and W. The selfenergy and the polarizability can then be expressed by the vertex. The resulting equations in conjunction with the Dyson equations provide a closed system, i.e. a system of equations with as many unknowns as equations. The simplest approximation of the described type is obtained, if one drops all but the constant term in the expansion of the vertex in terms of G and W. This is the famous GW-approximation [30, 18], which is used in almost all practical calculations that are based on Hedin's equations. With that the self-energy takes on the simple form $\Sigma(1,2) = i\lambda^2 G(1,2)W(1,2)$. We will now describe a new algorithm for solving Hedin's equations which includes corrections far beyond this GW approximation.

The computational storage requirements for this first order perturbation theory for the vertex scales as N^3 and the corresponding number of operations scales as N^4, where N is the number of basis functions as detailed in the next paragraph. Algorithm 1 (see below) for calculating $\Sigma[G,W]$ and $\Pi[G,W]$ goes much beyond GW without worsening this scaling. Note that all the relations in the algorithm are exact apart from the two which have the derivatives of Γ removed – these would require N^4 storage. Solving Algorithm 1 together with the Dyson equations we refer to as the 'Starfish' algorithm. It is interesting to work out which diagrams this algorithm corresponds to: finding the the self-consistent solution to Starfish is

equivalent to solving



(4.25)

self-consistently with the equations for Σ and Π (Eqs. (3.107) and (3.108)).

Algorithm 1 Hedin equations solver for $\Sigma[G,W]$ and $\Pi[G,W]$. Here the shorthand $G'(1,2;3) = \delta G(1,2)/\delta V(3)$, etc., is used and the function arguments are omitted[1].

Require: G and W
  Set $\Gamma = \Gamma_0$
  repeat
    $G' = \int G\,G\,\Gamma$
    $\Pi' = -i\lambda \int \left(G'\,G\,\Gamma + G\,G'\,\Gamma + \cancel{G\,G\,\Gamma'} \right)$
    $W' = \int W\,W\,\Pi'$
    $\Sigma' = i\lambda \int \left(G'\,W\,\Gamma + G\,W'\,\Gamma + \cancel{G\,W\,\Gamma'} \right)$
    $\Gamma = \Gamma_0 + \Sigma'$
  until Γ converged
  $\Sigma = i\lambda \int G\,W\,\Gamma$
  $\Pi = -i\lambda \int G\,G\,\Gamma$

4.2.1 Number and stability of solutions

Discretization of space-time is required for the purpose of closely examining solutions to these equations. There is some ambiguity as to how this should be done, but

[1] This algorithm is due to S. Sharma

at this level, we merely assume that there are N space-time points in total, and that limits in the time variable such as $G(4, 2^+)$ are taken to mean $G(4, 2)$. One can equally well think of any different basis like for instance an orbital-frequency basis. By discretizing one loses the physical meaning of these equations, but we will assume that the true physical solution can be recovered in the continuum limit when $N \to \infty$. In discrete space all eigenstates and eigenvalues of the Hamiltonian are *analytic* functions of λ on the whole real axis [45]. Consequently phenomena like phase transitions can only be described in the $N \to \infty$ limit.

The equations to be solved now form a closed system of *polynomial equations* enabling us to prove general theorems about such a system. To do so we define a concise notation. Let

$$F \equiv (G, W, \Pi, \Sigma, \Gamma) \in \mathbb{C}^n,$$

be the vector of all dependent (or unknown) variables. Here $n = 4N^2 + N^3$ is the number of unknowns. We regard the matrices G_0 and v (and the bare vertex) as known and fixed. Equations (3.107), (3.108), (3.113), (3.114) and a vertex equation like Eq. (4.25) can now be written in a compact form as $F = \mathfrak{g}(F)$ or $\mathfrak{h}(F) = 0$ with $\mathfrak{h}(F) \equiv \mathfrak{g}(F) - F$, where $\mathfrak{g} = (\mathfrak{g}_1, \ldots, \mathfrak{g}_n)$ is a set of polynomials in several variables.

It is important to point out that most of the following considerations do not depend on the precise form of the vertex equation. For example the trivial vertex equation $\Gamma = \Gamma_0$ corresponding to the GW approximation is also allowed here. In this case one can eliminate the vertex from the equations and redefine $F, \mathfrak{g}$ and $\mathfrak{h}$ in order to include only the smaller set of quantities and equations. Similarly, if one were to fix W in the Starfish algorithm then F would be (G, Σ, Γ) and the equations for W and Π could be eliminated.

Since we are interested in the dependence of these equations and their solutions on the coupling strength λ, the equations to be solved are

$$F = \mathfrak{g}^\lambda(F) \qquad \text{or} \qquad \mathfrak{h}^\lambda(F) = 0. \tag{4.26}$$

For the non-interacting case the solution is unique. We call this F^0.

We now want to determine the number of solutions to Eq. (4.26) in the interacting case. An upper bound is provided by Bézout's theorem [57, 58] which states that the maximum number of solutions to a system of polynomial equations (if finite) is equal to the product of the total degree of each equation. Thus the GW approximation with fixed W has at most 2^{N^2} solutions, and the Starfish algorithm, also for fixed W, has at most $7^{N^3} 2^{2N^2}$ solutions. A similar observation for the number of solutions of the equations underlying Hartree Fock calculations was made in Ref [7]. There also

a *lower* bound for the number of solutions is provided, which grows exponentially with the number of electrons in the system. It has also been shown numerically that in HF [8, 59, 60, 61], DFT [13, 14] and many other methods based on self-consistent equations [62, 63, 64, 65, 9, 10, 11] one can obtain more than one solution.

Buchberger's algorithm is a method of systematically determining the exact number of roots by decomposing the equations into a Gröbner basis (see Section 2.1). This procedure is, however, computationally very demanding and can be performed only for small (and therefore non-physical) N. For example for $N = 2$ the GW approximation with fixed W turns out to have precisely 6 solutions for generic G_0 and W. Likewise, for $N = 1$ the Starfish algorithm yields 3 solutions. From these considerations, it seems quite surprising that self-consistent GW works at all for realistic values of N. We will now provide two theorems that may explain this apparent success.

Theorem 6. *For all choices of Eq.(4.26) mentioned above (GW, Starfish and the versions with fixed W), the solutions have the following properties*

(i) Eq.(4.26) has one solution F^λ_{phys} *that tends to the non-interacting one* F^0 *in the non-interacting limit:*

$$F^\lambda_{\text{phys}} \underset{\lambda\to 0}{\longrightarrow} F^0.$$

(ii) All other solutions tend to infinity:

$$\inf\left\{\|F\| \;:\; \mathfrak{h}^\lambda(F) = 0,\; F \neq F^\lambda_{\text{phys}}\right\} \underset{\lambda\to 0}{\longrightarrow} \infty,$$

where $\|\cdot\|$ *is, say, the Euclidean vector norm.*

(iii) F^λ_{phys} *is an analytic function with respect to* λ *in a vicinity of* $\lambda = 0$.

This behavior suggests that at least in some low coupling regime F_{phys} is indeed the physical solution while all other are far away from the right result.

Proof. In the non-interacting case the Jacobian $\partial\mathfrak{h}/\partial F$ is a triangular matrix having -1's on the diagonal. Hence its determinant is not zero. In this situation the implicit function theorem [41] implies that one can solve for F as a function of λ in a neighborhood $U = U' \times U'' \subset \mathbb{C}^{n+1}$ of $(0, F^0) \in U$. It also implies that this function is analytic and that there is no other solution in U. This proves (i) and (iii).

Now for any compact set Ω where $\Omega \subset \mathbb{C}^n$ one can restrict λ such that $\mathfrak{h}^\lambda$ and $\mathfrak{h}^0$ are point-wise closer to each other in $\Omega \setminus U''$ than $\inf_{F\in\Omega\setminus U''}\|\mathfrak{h}^0(F)\| \neq 0$. This

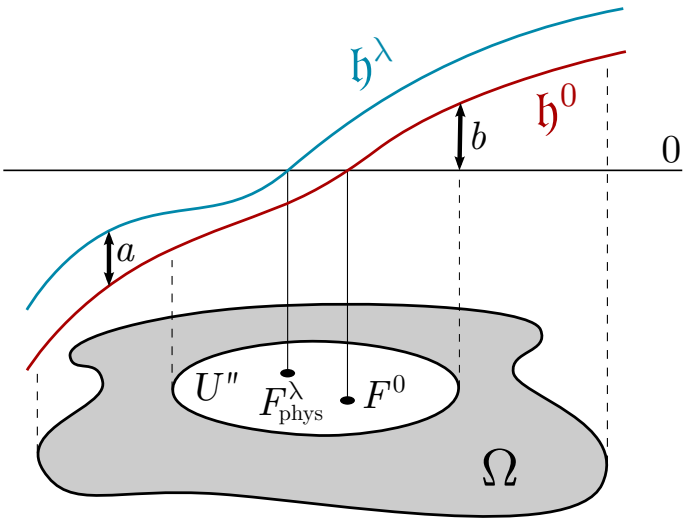


Figure 4.2: Schematic of the reasoning used in Theorem 6. Polynomial functions $\mathfrak{h}^0$ and $\mathfrak{h}^\lambda$ are drawn in red and blue on the domain $\Omega \subset \mathbb{C}^n$. The arrow at a represents the largest point-wise separation between $\mathfrak{h}^\lambda$ and $\mathfrak{h}^0$ on $\Omega \setminus U''$. The arrow at b represents $\inf_{F\in\Omega\setminus U''}\|\mathfrak{h}^0(F)\|$.

is schematically illustrated in Fig. 4.2. Then $\mathfrak{h}^\lambda$ has no zero in $\Omega \setminus U''$ and by the previous paragraph only F^λ_{phys} in U''. Since this can be done for any Ω this proves (ii). □

In practice, one solves Eq. (4.26) iteratively. That is, one starts with some initial guess F_0, inserts it into the right hand side of $F = \mathfrak{g}(F)$ and obtains a new guess. Iterating this procedure defines a sequence

$$F_{i+1} = \mathfrak{g}(F_i). \tag{4.27}$$

A natural starting point for this is the non-interacting solution $F_0 = F^0$, but this is by no means necessary. The hope is that this sequence converges to a fixed point, i.e. a solution to the equations. *A priori* it is unknown if the calculation will actually converge, and if a fixed point obtained this way actually corresponds to the *physical* solution. So one may wonder, why *GW* and similar schemes do, in fact, work in many situations. We now show that, for weak coupling convergence to the unique solution is guaranteed.

Theorem 7. *For small λ the physical solution F_{phys} is an attractive fixed point of Eq. (4.27). The size of the attracting region tends to infinity as $\lambda \to 0$.*

Here *size* can for instance be understood as the diameter of the largest ball that is contained in the attracting region.

Proof. It is sufficient to show that for any starting point we can restrict λ such that convergence to F_{phys} is guaranteed. For $a, b, c \in \mathbb{R}$ we define the following transformation:

$$\begin{aligned} G_0{}' &= aG_0 & v' &= bv & \Gamma_0' &= c\Gamma_0 \\ G' &= aG & W' &= bW & \Gamma' &= c\Gamma \\ \Sigma' &= a^{-1}\Sigma & \Pi' &= b^{-1}\Pi & \lambda' &= c\lambda \end{aligned} \tag{4.28}$$

with $a^2bc^2 = 1$. (We will use these transformations in this section only, to avoid confusion of the meaning of the primes with the derivatives as used earlier). This transformation leaves Eq. (4.27) invariant:

$$F_{i+1}' = \mathfrak{g}'(F_i').$$

We now apply the transformation with, say, $a = \lambda^{1/4}, b = \lambda^{1/2}, c = \lambda^{-1/2}$. This way $\mathfrak{g}'$ depends on λ explicitly and implicitly through a, b and c. Observe that *all* coefficients appearing in $\mathfrak{g}'$ tend to zero as $\lambda \to 0$. This is not true for the transformed starting values since the self-energy becomes larger due to the transformation. This is repaired by the first iteration: The quantities $F_1' = \mathfrak{g}'(F_0')$ tend to zero as $\lambda \to 0$. In this situation Banach's fixed point theorem can be applied for the map $\mathfrak{g}'$ defined in an appropriate neighborhood of zero that contains $F_1{}'$. We can conclude that for small λ the transformed quantities tend to a fixed point. Since the transformation can be inverted, this remains true for the original quantities. By Theorem 6 for small λ the solution F_{phys} is the solution nearest to the non-interacting one. Hence the fixed point obtained is indeed F_{phys}. □

Both theorems apply not only to the mentioned examples but to a large class of algorithms: it has already been demonstrated that they apply to different vertex equations. Another example would be to choose the solved form of the Dyson equations, e.g. $G = (1 - G_0\Sigma)^{-1}G_0$. This does not change the proof of Theorem 1 at all. For theorem 2 one would need the additional assumption that in the first step the matrix inverse exists. Then they will also exist in subsequent steps provided λ is small enough and due to the same arguments as before F_i would converge to the physical solution.

4.2.2 Numerical investigation

Numerical checks of the above theorems were performed for both self-consistent *GW* and the Starfish algorithm. General methods to obtain solutions to polynomial equations are described in ref. [39]. Some important mathematical properties of the

behavior of the solutions under change of parameters in the equations are covered there as well, while still remaining much more readable than most of the literature on algebraic geometry. As already mentioned, obtaining all solutions is in practice only possible for very small N. Hence for GW we set $N = 2$ and for Starfish we use $N = 1$. Plotted in Fig. 4.3 are *all* the solutions for these algorithms as

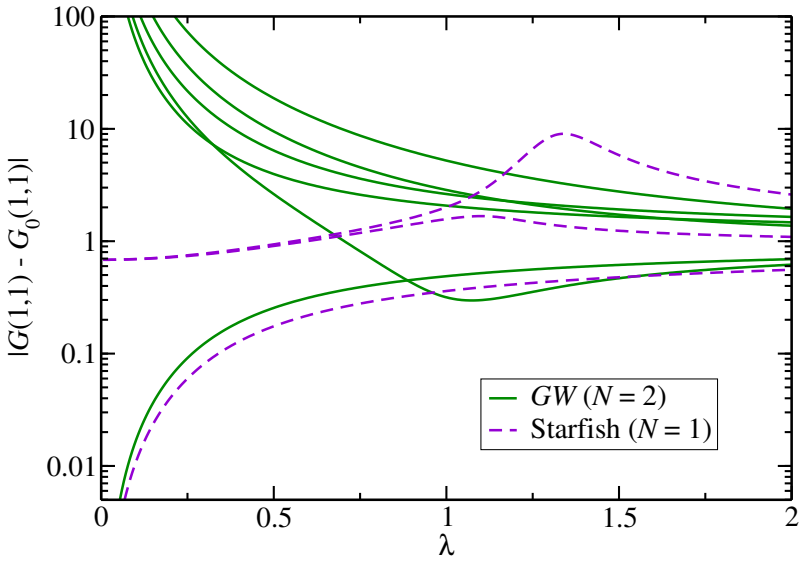


Figure 4.3: Plot of the distance of a matrix element of the Green's function to the non-interacting one versus the coupling strength λ for all possible solutions of GW with $N = 2$ and Starfish with $N = 1$, for random G_0 and W. W is kept fixed in both cases. Always only one solution tends to the non-interacting one for the weak coupling limit.

a function of λ. The numerical input, in this case G_0 and W, were chosen to be random complex numbers, and W was kept fixed (which is common practice for real GW calculations). As mentioned earlier, there are 6 solutions in the GW case. Of these 5 tend to infinity and the remaining solution tends to G_0 as $\lambda \to 0$. This is a visualization of Theorem 6. For Starfish 2 of the 3 solutions tend to a constant. This may seem to be in violation of the theorem but in this case the vertex Γ (and therefore F) diverges. We can also examine the domains of convergence for both of these algorithms. For $N = 1$ we fix $v = 1$ and $\lambda = 1$ and plot the region of convergence of G_0 for the fully self-consistent GW and Starfish algorithm (for which W is also computed self-consistently) in Fig. 4.4. It can be observed that

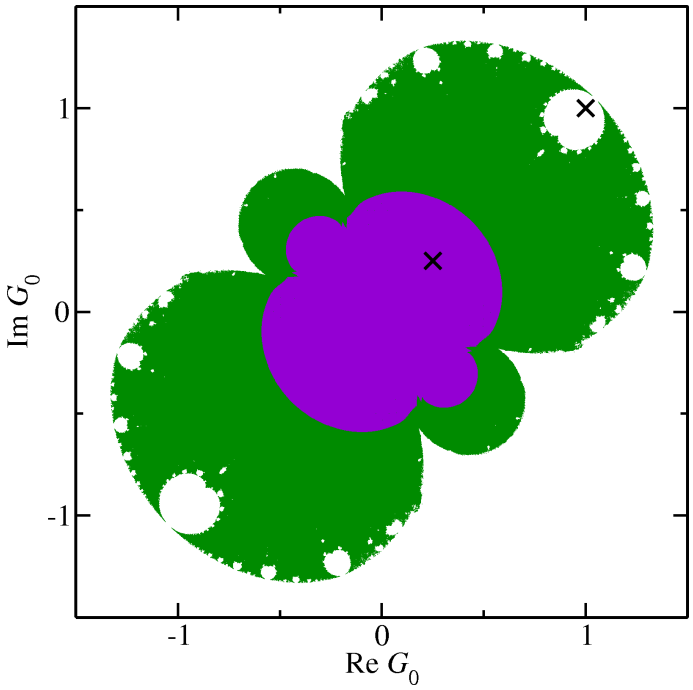


Figure 4.4: Domain of convergence of GW (green) and Starfish (purple) with $N = 1$ for input values of G_0 in the complex plane, when using the non-interacting solution as starting point[1]. Here $v = 1$ and $\lambda = 1$. The crosses mark the chosen G_0 for investigating the starting point dependence while fixing G_0, see Fig. 4.5.

the region of stability shrinks for the higher-order method. Also noteworthy is that the region has a fractal boundary (this may be unsurprising since for case $G_0W = 1$ the domain is the Mandelbrot set). Perhaps even more interesting is the region of *starting points* for which the algorithms converge. These are plotted in Fig. 4.5 for the same v and λ but this time with $G_0 = 1 + i$ for GW and $G_0 = 1/4 + i/4$ for Starfish, and with a variable starting point for G. These points are indicated by crosses in Fig. 4.4. Once again the region of convergence is smaller for Starfish, but in both cases *only one* solution is found, irrespective of the starting point. This is a numerical confirmation of Theorem 7. Note that for GW a situation was picked, where the non-interacting starting point does *not* lead to convergence. Hence this can be considered a large coupling situation. But still there seems to be only one stable fixed point. The boundary of the region is also fractal (this corresponds to the Julia set).

[1]This calculation was done by K. Dewhurst.

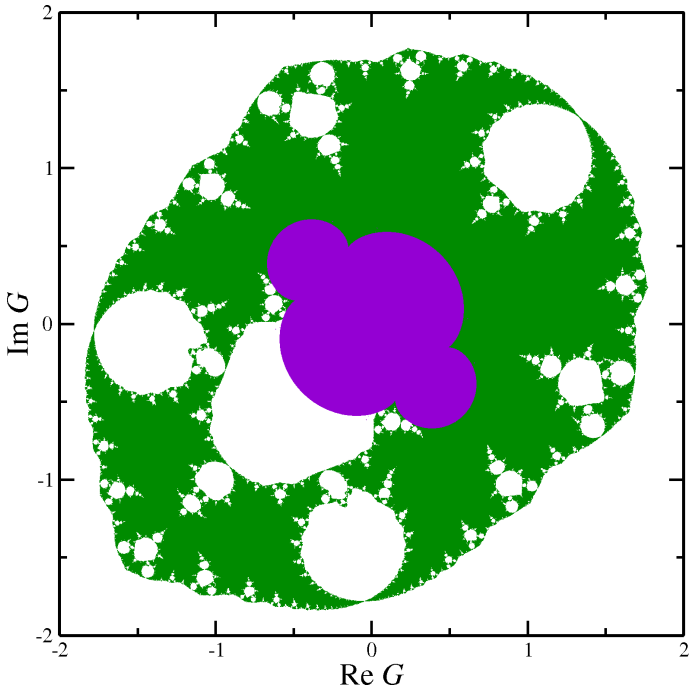


Figure 4.5: Domain of convergence of GW (green) and Starfish (purple) with $N = 1$ for different starting points of the fixed point cycle[1]. The values of G_0 are fixed to $1 + i$ for GW and $1/4 + i/4$ for Starfish, as indicated by the crosses in Fig. 4.4.

4.2.3 Mixing

In most realistic calculations of the self-consistent GW approximation, the previous and current quantities are mixed with one another using various mixing schemes [66]. This can both ensure convergence to the fixed point as well as accelerate it. In our notation, this is equivalent to modifying Eq. (4.27) as follows:

$$F_{i+1} = \beta\mathfrak{g}(F_i) + (1 - \beta)(F_i), \tag{4.29}$$

where, in general, β is a matrix. The modified map $\tilde{\mathfrak{g}}(F) = \beta\mathfrak{g}(F) + (1 - \beta)F$ yields $\tilde{\mathfrak{h}}(F) = \beta(\mathfrak{g}(F) - F) = \beta\mathfrak{h}$. Thus if β is an invertible matrix then the fixed points of $\mathfrak{g}$ are preserved. We also note that a fixed point is asymptotically stable if all the eigenvalues of its Jacobian $\tilde{J} = \partial\tilde{\mathfrak{h}}/\partial F = \beta J$ have negative real parts. If any eigenvalue has a positive real part, then the fixed point is unstable. One may therefore choose the matrix β such that the eigenvalues of $\tilde{J}$ have arbitrary real

[1]This calculation was done by K. Dewhurst.

parts. In this way particular fixed points can be made either stable or unstable as desired.

To illustrate the utility of mixing we recalculate the stability regions in Fig. 4.5 for the *GW* approximation by mixing the Green's function alone

$$G_{i+1} = \beta G_i + (1 - \beta) G_i,$$

with β taken to be a positive real number. The effect of this is plotted in Fig. 4.6.

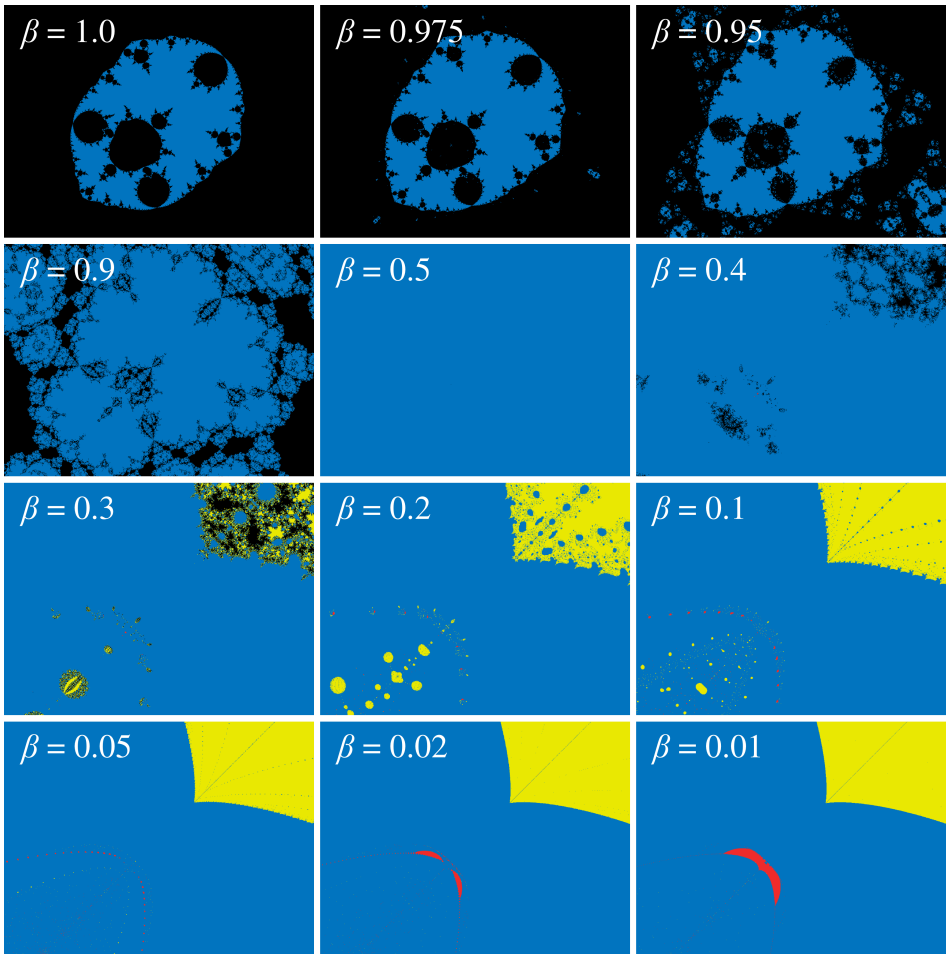


Figure 4.6: Domain of convergence of *GW* as a function of the mixing parameter[1]. Three different solutions are plotted with different colors: blue, yellow and red. Black indicates non-convergence.

[1]This calculation was done by K. Dewhurst.

Two interesting aspects are observed: first, the domain of convergence increases dramatically with decreasing mixing parameter β; and second, *different solutions* emerge as a function of β (these are indicated by three different colors in the plot). Note that for $\beta = 1$ only one solution was found. These observations may have consequences for realistic GW calculations, namely that adjusting the mixing parameter can fundamentally change the obtained solution. In our simple example, we chose β to be a positive real number but it could also have been negative, complex, a diagonal matrix with non-zero entries or a general invertible matrix.

We have argued that truncating Hedin's equations to some order yields systems of polynomial equations which have a large number of solutions. As an example of this, the Starfish algorithm was introduced which includes vertex corrections beyond GW and consequently has even more fixed point solutions. Two theorems were presented that shed some light on the general behavior of these fixed points. In particular we have shown that there is exactly one solution that tends to the non-interacting case for small coupling, while all others are divergent in this limit [67, 68, 69]. Numerical tests of self-consistent GW and the Starfish algorithm for small N demonstrated that the system also converges uniquely to one fixed point even for fairly large coupling. Furthermore, the region of stability may be fractal in nature, indicating that finding simple necessary and sufficient conditions for ensuring convergence of GW calculations *a priori*, may be impossible. Lastly, we found that mixing current and previous solutions with a certain mixing parameter not only increases the radius of convergence but also allows different solutions to be obtained. This observation may be of use in realistic GW calculations to assist in finding these other solutions and investigate their properties.

Chapter 5

The Berry phase in the exact factorization

Notation As in the previous chapters we try to be ISO-80000 conformal, i.e. we use italic letters for variables and upright letters for mathematical an physical constants and fixed symbols, including names for the reasons already detailed in Chapter 2. From here on, we also use bold letters to denote 2- and 3-vectors. Apart from that, we decided not to introduce any notation to emphasize vectors, matrices or operators, as this would only clutter formulas. This is also intended to highlight the difference between 3-vectors (bold letters) and higher dimensional vectors (not bold). To ease the reading, we use Latin letters $(a, b, c, \dots)$ as subscripts of coordinates of electrons and Greek letters $(\alpha, \beta, \gamma, \dots)$ for nuclear coordinates. We use Latin letters, starting from $i, j, k, \dots$, if an index runs through both of these sets of coordinates.

5.1 Choice of coordinates

We investigate the wave function $\psi(R, r)$ that describes a molecule in the non-relativistic description. It depends on $d \cdot N_\mathrm{n}$ nuclear coordinates $R = (\boldsymbol{R}_1, \dots, \boldsymbol{R}_{N_\mathrm{n}})$ and $d \cdot N_\mathrm{e}$ electronic coordinates $r = (\boldsymbol{r}_1, \dots, \boldsymbol{r}_{N_\mathrm{e}})$, where d is the number of spatial dimensions considered (usually $d = 3$). The Hamiltonian for such a system is

$$H = T_\mathrm{n} + T_\mathrm{e} + V, \tag{5.1}$$

where $V(R, r)$ is the Coulomb interaction among all particles and T_n and T_e are the kinetic energy of nuclei and electrons respectively

$$T_\mathrm{n} = -\sum_{\alpha=1}^{d\cdot N_n} \frac{1}{2M_\alpha} \frac{\partial^2}{\partial R_\alpha^2}, \qquad T_\mathrm{e} = -\sum_{a=1}^{d\cdot N_e} \frac{1}{2m_a} \frac{\partial^2}{\partial r_a^2}. \tag{5.2}$$

M_α and m_a are the mass of the nucleus or electron to which this coordinate corresponds. If we want to describe a molecule in an external potential, this can be absorbed into V. If instead V is a potential that depends on relative coordinates only, the center of mass motion can be split off explicitly. A particular convenient choice of coordinates are the Jacobi coordinates [47]

$$\boldsymbol{q}_k = \frac{\sum_{i=1}^{k} m_i \boldsymbol{x}_i}{\sum_{i=1}^{k} m_i} - \boldsymbol{x}_{k+1}, \qquad \boldsymbol{q}_N = \frac{\sum_{i=1}^{N} m_i \boldsymbol{x}_i}{\sum_{i=1}^{N} m_i}, \tag{5.3}$$

where $N = N_e + N_n$ is the total number of particles considered and $\boldsymbol{x}_i$ runs through all electronic and nuclear coordinates with the corresponding masses m_i. The k^{th} new coordinate is the difference vector of the $(k+1)^{\text{th}}$ particle and the center of mass of the first k particles. In these new coordinates the total kinetic energy of the system still has the same form

$$T_\text{n} + T_\text{e} = -\sum_{k=1}^{d\cdot N} \frac{1}{2\mu_k} \frac{\partial^2}{\partial q_k^2}. \tag{5.4}$$

The only difference is that the masses need to be changed according to

$$\mu_k^{-1} = \sum_{i=1}^{k+1} m_i^{-1}, \qquad \mu_N = \sum_{i=1}^{N} m_i, \tag{5.5}$$

where the first formula defines the $(N-1)$ reduced masses and the second one, the total mass. The simple form of the total kinetic energy is the reason that these coordinates are so useful. If we would have made the more obvious choice of taking just the difference of each coordinate to the center of mass, while dropping one of the coordinates in favor of the center of mass, the expressions would have become slightly more complicated.

Note that the coordinate transformation described here is just one choice among many which preserve this simple form of the kinetic energy operator. We just picked the simplest choice of Jacobi coordinates possible for the sake of illustration. There may be coordinates which are better suited for numerical calculations. It is also worth mentioning that one can also get rid of the mass parameters by a simple rescaling of the coordinates according to

$$q_k \to \sqrt{\mu_k} q_k. \tag{5.6}$$

That way they no longer appear in the kinetic energy operator. However, they are merely shifted to the interaction potential V.

After these coordinate transformations, we are left with $(N-1)$ relative coordinates $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_{N-1}$ and one center of mass coordinate $\boldsymbol{q}_N$. As we assume that the interaction potential does not depend on the latter, the Hamiltonian commutes with the total momentum operator $\mathrm{i}\frac{\partial}{\partial q_N}$. It can even be written as the sum of the kinetic energy of the center of mass motion and a remaining part. The latter no longer depends or acts on the center of mass coordinate. The total wave function can hence be decomposed into a product of a wave function that depends on the $(N-1)$ relative coordinates only and a wave function which describes a plane wave with respect to to the center of mass coordinate.

The remarkable result of this exercise is that we cannot only get rid of d (d is the dimension of the space, usually three) coordinates by splitting off the center of mass motion, but also keep the original form of the total kinetic energy operator. We will even go back to the same symbols we initially introduced. We should, however, keep in mind that the meaning of the coordinates and parameters changed. The m_a and M_α are now effective masses. Though one can think of the R_α as relative nuclear coordinates and the r_a as coordinates that are essentially the electronic coordinates relative to the total center of mass, this is by no means necessary. We will in a moment split the system into two sets of coordinates R and r. As we will treat the equations in an exact manner, i.e. we are not making the Born-Oppenheimer approximation or the like, this splitting can be done in any way. One might for instance consider only some of the nuclei as the system described by the coordinates R and consider the others and the electrons through the coordinates r. Another interesting possibility is to regard the electronic coordinates as R and the nuclear coordinates as r. That being said, we will go back to calling the coordinates R and r nuclear and electronic coordinates respectively. The same applies to the mass parameters. We will later discuss the topology of the nodal structure of the wave function. As we are dealing with a linear and invertible coordinate transformation all topological properties of the original wave function are identical to those of the wave function with the center of mass coordinate split off.

There is another symmetry of the molecular Hamiltonian that is usually utilized. The rigid rotation of the molecule does not change its total energy. The operators of the total angular momentum hence commute with the Hamiltonian. This allows elimination of three more coordinates if we consider the three dimensional case. In two dimensions only one more degree of freedom could be removed. Either way, the remaining Hamiltonian does, in contrast to the center of mass motion case, depend on the chosen quantum numbers. This should not come as a surprise, as there are Coriolis and centrifugal forces. What is even worse, the form of the kinetic energy operator becomes much more involved. For that reason it is more convenient to discuss the problem without explicitly eliminating these coordinates.

5.2 Born-Oppenheimer Expansion

We consider eigenfunctions of the total Hamiltonian, which still has the form of Eq. (5.1), but with the slightly changed meaning of the coordinates and parameters described in the previous section

$$H\psi = E\psi, \tag{5.7}$$

with

$$H = T_{\mathrm{n}} + T_{\mathrm{e}} + V. \tag{5.8}$$

We define the electronic Hamiltonian as the operator

$$h(R) = T_{\mathrm{e}} + V(R), \tag{5.9}$$

which can be viewed as an operator that depends parametrically on the nuclear coordinates and acts in the space of electronic wave functions, i.e., wave functions which depend on the electronic coordinates alone. Many authors split up the potential further into a part that acts on the nuclei alone, a part that act on the electrons alone and a part that describes the interaction among them. Also the part that acts on the nuclei alone is often not included in this electronic Hamiltonian but is added later, when the full Hamiltonian is considered. These definitions are completely equivalent, the only difference is that the interaction among the nuclei is already included in the eigenvalues of the electronic system. As this operator depends on R, its eigenfunctions and eigenvalues depend on R as well

$$h(R)|a_R\rangle = \epsilon_a(R)|a_R\rangle. \tag{5.10}$$

The eigenenergies as a function of R are called Born-Oppenheimer surfaces. We will use the bra-ket-notation exclusively for the electronic space, while the R-dependence will be made explicit. For instance the total wave function is viewed as a R-dependent ket $|\psi(R)\rangle$.

The expansion of the full wave function in terms of this basis is called Born-Oppenheimer expansion

$$|\psi(R)\rangle = \sum_a \chi_a(R)|a_R\rangle, \tag{5.11}$$

where the R-dependent coefficients $\chi_a(R)$, known as nuclear wave functions, have been introduced.

Using this expansion in the full Schrödinger equation Eq. (5.7) and multiplying with a bra-vector leads to a Schrödinger like equation for the nuclear wave functions, which can be written as [47]

$$\left[\frac{1}{2}(A(R) - \mathrm{i}\nabla)^2 + \epsilon(R)\right] \chi(R) = E\,\chi(R), \tag{5.12}$$

where $\chi = (\chi_1, \chi_2, \dots)^{\mathrm{T}}$ is the column vector of the nuclear wave functions,

$$\epsilon = \mathrm{diag}(\epsilon_1, \epsilon_2, \dots) \tag{5.13}$$

is the diagonal matrix of Born-Oppenheimer surfaces and

$$A_{ab}(R) = \langle a|\mathrm{i}\nabla|b\rangle \tag{5.14}$$

is the so-called gauge connection, which can be viewed as a vector of matrices or equivalently a matrix of vectors. Note that the square in Eq. (5.12) imposes a matrix multiplication as well as a scalar product of this vector. Specifically Eq. (5.12) stands for

$$\left[\frac{1}{2}((A_\alpha)_{ac} - \mathrm{i}\delta_{ac}\partial_\alpha)((A_\alpha)_{cb} - \mathrm{i}\delta_{cb}\partial_\alpha) + \epsilon_a\delta_{ab}\right] \chi_b = E\,\chi_a. \tag{5.15}$$

The gauge-covariant curvature two-form is defined as [47]

$$F_{\alpha\beta} = \partial_\alpha A_\beta - \partial_\beta A_\alpha - \mathrm{i}[A_\alpha, A_\beta], \tag{5.16}$$

This turns out to be zero wherever the derivatives are well defined [47], as the following calculation shows (remember that Latin letters correspond to the electronic space and Greek letters to the nuclear coordinates):

$$\begin{aligned}
(F_{\alpha\beta})_{cd} &= \partial_\alpha(A_\beta)_{cd} - \mathrm{i}(A_\alpha)_{ce}(A_\beta)_{ed} \;-\; \alpha \leftrightarrow \beta && (5.17)\\
&= \mathrm{i}\partial_\alpha\langle c|\partial_\beta|d\rangle + \mathrm{i}\underbrace{\langle c|\partial_\alpha|e\rangle}_{=-\langle\partial_\alpha c|e\rangle}\langle e|\partial_\beta|d\rangle \;-\; \alpha \leftrightarrow \beta && (5.18)\\
&= \mathrm{i}\langle\partial_\alpha c|\partial_\beta|d\rangle + \mathrm{i}\langle c|\partial_\alpha\partial_\beta|d\rangle - \mathrm{i}\underbrace{\langle\partial_\alpha c|e\rangle\langle e|\partial_\beta|d\rangle}_{=\langle\partial_\alpha c|\partial_\beta|d\rangle} \;-\; \alpha \leftrightarrow \beta && (5.19)\\
&= \mathrm{i}\langle c|\partial_\alpha\partial_\beta|d\rangle \;-\; \alpha \leftrightarrow \beta && (5.20)\\
&= 0, && (5.21)
\end{aligned}$$

where "$\alpha \leftrightarrow \beta$" stands for "the same terms but with α and β interchanged", the first line is just the definition, in the second and third line we used that the electronic states form an orthonormal and complete basis for any fixed set of nuclear coordinates and in the last line we used the symmetry of second derivatives.

It can be shown that this implies A can be gauged away by a unitary transformation in any simply connected region in which the electronic wave functions are continuously differentiable [47, 70]. In the next section, we will analyze the conditions for the electronic wave functions to be continuously differentiable with respect to nuclear coordinates.

5.2.1 Conical intersections

We clearly need to understand how and when it is possible that the electronic wave functions do not depend smoothly on the nuclear coordinates. In order to analyze this, it is necessary (or at least convenient) to consider a truncated Hilbert space for the electronic coordinates only. That way the electronic Hamiltonian $h(R)$ can be represented by a hermitian matrix of the dimension d_e of this space. Fortunately, the behavior of the eigenvectors and eigenvalues of a parameter dependent hermitian matrix is understood very well [45]. A remarkable property is that if the matrix depends only on a single parameter, then all eigenvalues and eigenvectors can be chosen to be analytic functions of it. This is no longer true, if two parameters are considered simultaneously. The crucial point is degeneracy. All of this can be understood by considering the properties of the eigenvectors and eigenvalues in a two dimensional subspace that we consider decoupled from the rest of the space. We write the electronic Hamiltonian as

$$h = \begin{pmatrix} h_{2\times 2} & 0 \\ 0 & \tilde{h} \end{pmatrix}, \qquad h_{2\times 2} = \begin{pmatrix} d+c & a+b\mathrm{i} \\ a-b\mathrm{i} & d-c \end{pmatrix}, \tag{5.22}$$

where a, b, c and d are real parameters, which depend on the nuclear coordinates. This is the most general two dimensional selfadjoint operator. It is real if $b = 0$. We could obviously have parametrized it in a slightly simpler way, but this way the upcoming formulas are a little simpler. The Hamiltonian can be written as

$$h_{2\times 2} = \boldsymbol{B} \cdot \boldsymbol{\sigma} + d, \tag{5.23}$$

where $\boldsymbol{B} = (a, b, c)^\mathrm{T}$ and $\boldsymbol{\sigma}$ is the vector of the three Pauli matrices. So d is merely an R-dependent shift in total energy, while a, b and c can be viewed as the components of a magnetic field. Note that all we have done is to consider the most general two by two Hamiltonian. This just means that any such system is *equivalent* to a spin one-half particle in a magnetic field. This also applies in cases in which no magnetic field is present at all. The eigenvalues are

$$\epsilon_\pm = d \pm \sqrt{a^2 + b^2 + c^2} = d \pm B, \tag{5.24}$$

where B was defined as the modulus of $\boldsymbol{B}$. Now one can see that for fixed a, b and d with a or b non-zero, these are clearly analytic functions with respect to c. This is an illustration of the fact mentioned earlier that the eigenvalues are analytic functions of a single parameter. However, what happens if $a = b = 0$? Then $\epsilon_+ = d + |c|$, which is clearly not analytic near $c = 0$. In this case, we can change the labeling of the eigenvalues to $\epsilon_{1/2} = d \pm c$. That way we get back to two analytic functions. Such a trick does no longer work if we consider two or more parameters at the same time. If we consider the case $d = c = 0$, then the two energy eigenvalues as a function of a and b form a double cone. At the tip of such a cone, the functions are obviously not analytic. This is what is called a conical intersection.

The eigenvectors are most conveniently expressed in terms of spherical coordinates, which we introduce via

$$a = B\cos(\beta)\sin(\alpha) \tag{5.25}$$
$$b = B\sin(\beta)\sin(\alpha) \tag{5.26}$$
$$c = B\cos(\alpha). \tag{5.27}$$

So the introduced coordinates α and β are just the standard spherical coordinates to describe the angle of the $\boldsymbol{B}$-vector. All directions can be covered for $0 \leq \alpha \leq \pi$ and $0 \leq \beta \leq 2\pi$. The eigenfunctions do not depend on d or B. They can be written as

$$|+\rangle = \quad \cos(\alpha/2)\mathrm{e}^{-\mathrm{i}\beta/2}|\uparrow\rangle + \sin(\alpha/2)\mathrm{e}^{\mathrm{i}\beta/2}|\downarrow\rangle \tag{5.28}$$
$$|-\rangle = -\sin(\alpha/2)\mathrm{e}^{-\mathrm{i}\beta/2}|\uparrow\rangle + \cos(\alpha/2)\mathrm{e}^{\mathrm{i}\beta/2}|\downarrow\rangle, \tag{5.29}$$

where $|\uparrow\rangle$ and $|\downarrow\rangle$ are the basis functions spanning the two dimensional space we are considering. Let us now focus on the state with lower energy $|-\rangle$. The way we have chosen the phase of this state it changes sign under a full rotation $\beta \to \beta + 2\pi$. Since the choice of phase is arbitrary, we can switch to a different gauge. That is we can choose the phase of the wave function in a different R or equivalently β dependent way. The function

$$|\phi\rangle = \mathrm{e}^{\mathrm{i}\beta/2}|-\rangle = -\sin(\alpha/2)|\uparrow\rangle + \cos(\alpha/2)\mathrm{e}^{\mathrm{i}\beta}|\downarrow\rangle \tag{5.30}$$

is single valued. The corresponding vector potential can be calculated

$$A_\mu = \langle\phi|\mathrm{i}\partial_\mu|\phi\rangle = -\left(\frac{1}{2} + \frac{1}{2}\cos(\alpha)\right)\frac{\partial\beta}{\partial R_\mu} \tag{5.31}$$

The phase associated with a closed loop for fixed α is

$$\gamma = \oint A_\mu \mathrm{d}R_\mu \tag{5.32}$$

$$= -\oint \left(\frac{1}{2} + \frac{1}{2}\cos(\alpha)\right) \frac{\partial \beta}{\partial R_\mu} \mathrm{d}R_\mu \tag{5.33}$$

$$= -\oint \left(\frac{1}{2} + \frac{1}{2}\cos(\alpha)\right) \mathrm{d}\beta \tag{5.34}$$

$$= -(1 + \cos(\alpha))\, n\, \pi, \tag{5.35}$$

where n is the winding number, how often the two dimensional vector (a, b) winds around the origin.

We would like to discuss this result for two cases. First consider the case that $\cos(\alpha) = 0$, which means that the Hamiltonian can be chosen real. In this case the phase is an integer multiple of π. It does no longer depend on the precise shape of the integration path, but only on the winding number n. The phase is a topological quantity. Second consider the case of a general complex Hamiltonian. Then the phase is actually the solid angle $\boldsymbol{B}$ encircles by on the integration path. It depends in a continuous way on the geometry of the integration contour. It is a geometric phase.

Dimension of Conical Intersections In the previous section we established the relevance of degeneracies of the electronic Hamiltonian. We would now like to get a better understanding of the dependency of degeneracies on the nuclear coordinates. Towards this end we apply the spectral theorem in order to write the electronic Hamiltonian as

$$h = U^\dagger \, \epsilon \, U, \tag{5.36}$$

where U is a $d_\mathrm{e} \times d_\mathrm{e}$ unitary matrix and $\epsilon = \mathrm{diag}(\epsilon_1, \ldots, \epsilon_{d_\mathrm{e}})$ is a the already introduced diagonal matrix of Born Oppenheimer surfaces (though now it is considered finite dimensional). We can now ask: What is the dimension of the space of electronic Hamiltonians? We consider the case that we are dealing with r distinct eigenvalues, with multiplicities g_i, such that $\sum_{i=1}^{r} g_i = d_\mathrm{e}$. From Eq. (5.36) we can see that we are free to choose r eigenvalues and a $d_\mathrm{e} \times d_\mathrm{e}$ unitary matrix. The former provides r parameters and the latter d_e^2 real parameters. Then the action of a $g_i \times g_i$ dimensional unitary matrix acting in the eigenspace corresponding to the $i^{\mathrm{t}h}$ distinct eigenvalue, will not change the Hamiltonian (i.e. the action of U and $U^\dagger$ will compensate each other). Hence we have to subtract the dimension of these

spaces. We end up with a total dimension of [47]

$$d_{\mathrm{e}}^2 + r - \sum_{i=1}^{r} g_i^2 \tag{5.37}$$

for the space of electronic Hamiltonians with degeneracies g_i. That means that the space of Hamiltonians without any degeneracy is d_{e}^2 (take $r = d_{\mathrm{e}}$ and $g_1 = g_2 = \cdots = g_r = 1$). As expected, this is just the dimension of the space of hermitian matrices. The dimension of the space of Hamiltonians with one degeneracy ($r = d_{\mathrm{e}} - 1$, $g_1 = 2$, $g_2 = g_3 = \cdots = g_r = 1$) is $d_{\mathrm{e}}^2 - 3$ instead. The condition that a degeneracy is present is hence equivalent to three real constraints [47]. For real Hamiltonians the diagonalization can be done with orthogonal matrices. The number of parameters determining a Hamiltonian is then [47]

$$\frac{1}{2} d_{\mathrm{e}}(d_{\mathrm{e}} - 1) + r - \sum_{i=1}^{r} \frac{1}{2} g_i(g_i - 1). \tag{5.38}$$

The dimension of the space of real Hamiltonians with one degeneracy turns hence out to be two less than the dimension of the space of all real Hamiltonians. So for real Hamiltonians, the condition that there is a degeneracy is equivalent to two real constraints [47].

At this point one tends to conclude that this means that the dimension of the space of nuclear coordinates that lead to such a degeneracy in the electronic Hamiltonian is also the dimension of the entire space minus three or minus two for complex or real Hamiltonians respectively [46, 47]. However this conclusion is not true in general. A more detailed analysis of the symmetries of the problem is needed. Consider the case of H_2^+. After removing the center of mass motion as described in a previous section, we are left with the three component relative coordinate of the nuclei and the three coordinates of the electron (relative to the center of the nuclei).

5.3 Exact factorization

Up until now, we considered the Born Oppenheimer expansion

$$|\psi(R)\rangle = \sum_a \chi_a(R)|a_R\rangle, \tag{5.39}$$

where the electronic wave functions $|a_R\rangle$ were eigenfunctions of the electronic Hamiltonian. If the sum is truncated after the first term, this is the Born Oppenheimer

approximation, in which the total wave function is a single product of a nuclear wave function and an electronic wave function. There is an alternative approach, which does not impose any approximation, but still expresses the total wave function as a single product [71]

$$\psi(R,r) = \bar{\chi}(R)\bar{\phi}_R(r), \tag{5.40}$$

where we put the bar over χ and ϕ to distinguish from the wave functions introduced in the previous section. Instead of demanding that the electronic wave function is an eigenstate of the electronic Hamiltonian, we define the nuclear wave function in such a way that it provides the correct nuclear density of the full wave function

$$|\bar{\chi}(R)|^2 = \int \mathrm{d}r \; |\psi(R,r)|^2 \tag{5.41}$$

and then the electronic wave function is defined by Eq. (5.40) and is normalized for any fixed nuclear configuration

$$\int \mathrm{d}r \left|\bar{\phi}\right|^2 = 1. \tag{5.42}$$

This does not yet fix the phase of $\bar{\chi}$ and $\bar{\phi}$, leaving the gauge-like freedom

$$\bar{\chi}(R) \;\rightarrow\; \mathrm{e}^{\mathrm{i}\theta(R)} \; \bar{\chi}(R) \tag{5.43}$$

$$\bar{\phi}_R(r) \;\rightarrow\; \mathrm{e}^{-\mathrm{i}\theta(R)} \; \bar{\phi}_R(r), \tag{5.44}$$

which leaves the total wave function ψ unchanged. It is not hard to show that the nuclear wave function satisfies a Shrödinger-like equation [72]

$$\left[\frac{1}{2}(\bar{A}(R) - \mathrm{i}\nabla)^2 + \bar{\epsilon}(R)\right] \bar{\chi}(R) = E\,\bar{\chi}(R), \tag{5.45}$$

which has the exact same form as the equation we obtained using the Born Oppenheimer expansion, but with $\bar{A}$ defined as

$$\bar{A}_k(R) = \mathrm{i}\int \mathrm{d}r \; \bar{\phi}^*_R(r)\frac{\partial}{\partial_k}\bar{\phi}_R(r) \tag{5.46}$$

and a function $\bar{\epsilon}$, which can also be expressed in terms of $\bar{\phi}$ (see Ref. [71]). The $\bar{A}$-field transforms under the gauge transformation above as

$$\bar{A}_k(R) \;\rightarrow\; \bar{A}_k(R) - \frac{\partial}{\partial_k}\theta(R) \tag{5.47}$$

Though Eq. (5.45) looks like a Schrödinger equation for the nuclei in an effective external field ϵ and an external magnetic field described by $\bar{A}$, the situation is more complicated than that. It is not too surprising that $\epsilon(R)$ can depend on all nuclear coordinates simultaneously, even though the Coulomb interaction depended only on the pairwise distance. The $\bar{A}_k$-term does, however, also depend on all nuclear coordinates simultaneously, meaning that this equation cannot simply be interpreted as (effectively) interacting nuclei in one effective electromagnetic field. Instead one could say that the effective magnetic field felt by one nucleus depends on the positions of all other nuclei. Note that also the effective interaction between the nuclei is very different from its original. When describing a bound state of a molecule, this interaction must clearly be attracting at least partly, in contrast to the original interaction between the nuclei, which was repelling. This is very intuitive, though, as this effective attraction can be viewed as the electrons gluing the nuclei together.

We define the Berry phase associated to a gauge-field $\bar{A}$ and a closed path C through the space of nuclear coordinates R as

$$\bar{\gamma} = \int_C \bar{A}(R) \cdot \mathrm{d}R, \tag{5.48}$$

where the dot emphasizes the inner product of $\bar{A}$ and $\mathrm{d}R$. If we define the Berry curvature as the exterior derivative of $\bar{A}$

$$\bar{\Omega}_{ab}(R) = \frac{\partial}{\partial_a}\bar{A}_b(R) - \frac{\partial}{\partial_b}\bar{A}_a(R). \tag{5.49}$$

then Stokes theorem links such an integral to an integral over a surface S

$$\bar{\gamma} = \int_S \bar{\Omega}_{ab}(R)\ \mathrm{d}R^a \mathrm{d}R^b, \tag{5.50}$$

where C is the boundary of S. Both the Berry phase and the Berry curvature are gauge-independent.

Let us end this section with a remark that appears to be made too rarely in the literature: Instead of defining the Berry curvature via a derivative, as in Eq. (5.49), we could define it as the function which satisfies Eq. (5.50), where the Berry phase is retrieved from Eq. (5.48). This way of defining the Berry curvature is not only more natural, in the sense that the definition directly tells us what the object means, but it is also more general, as it allows us to define the Berry Curvature even when the derivatives of $\bar{A}$ do not exist. In this case the Berry phase is, of course, a distribution.

The alternative to the way of making the definition of the Berry phase more general is to simply deal with the limitation of the given definition. This approach

is probably easier to grasp, as we can always consider the objects we are dealing with as ordinary functions. The downside is that we cannot use Stokes theorem, as the space where our quantities are well defined might not be simply connected.

In order to avoid this whole issue, we will use the Berry connection most of the time. It is generally enough to consider it on paths for which it is well defined, as the exceptional points are sufficiently rare, as we will see later.

5.3.1 Triviality for current-free wave functions

We now introduce the nuclear density and current density

$$n(R) = \int \mathrm{d}r \ |\psi(R,r)|^2 = |\bar{\chi}(R)|^2 \tag{5.51}$$

$$j_k(R) = \mathrm{Im}\left[\int \mathrm{d}r \ \psi^*(R,r)\frac{\partial}{\partial_k}\psi(R,r)\right]. \tag{5.52}$$

Using the definition of $\bar{A}$ and Eq. (5.40) in the latter equation reveals that $\bar{\chi}$ not only reproduces the correct nuclear density, but also the nuclear current density one would expect for a gauge theory, with the given gauge connection

$$j_k(R) = -n(R)\bar{A}_k(R) + \mathrm{Im}\left[\bar{\chi}^*(R)\frac{\partial}{\partial_k}\bar{\chi}(R)\right]. \tag{5.53}$$

A similar equation can be derived by using the Born Oppenheimer expansion in Eq. (5.52). Both ways we obtain the same nuclear current, which implies an interesting relationship between the exact factorization and the non-abelian gauge connection

$$n\bar{A}_k - \mathrm{Im}\left[\bar{\chi}^*\frac{\partial}{\partial_k}\bar{\chi}\right] \quad = \quad \chi_a^*(A_k)_{ab}\chi_b - \mathrm{Im}\left[\chi_a^*\frac{\partial}{\partial_k}\chi_a\right]. \tag{5.54}$$

If in both descriptions the gauge in which the nuclear wave functions are real is chosen then the gauge connection of the exact factorization is the average of the matrix elements of the original gauge connection, with the nuclear wave functions as weights (note that they can be negative as well, though, implying the possibility of cancellations).

Let us now consider a system in which the nuclear current density is zero and choose a gauge in which $\bar{\chi}$ is real and non-negative:

$$\bar{\chi}(R) = \sqrt{\int \mathrm{d}r \ |\psi(R,r)|^2} \tag{5.55}$$

$$\bar{\phi}_R(r) = \frac{\psi(R,r)}{\bar{\chi}(R)} \ . \tag{5.56}$$

In this special case both the LHS and the second term on the RHS of Eq. (5.53) vanish

$$n(R)\bar{A}_k(R) = 0. \tag{5.57}$$

To satisfy this equation one of the factors must be zero. We arrive at the important conclusion

Theorem 8. *Consider a wave function for which the nuclear n-body current density vanishes everywhere. Then in the gauge given in Eq. (5.55), the gauge connection is zero everywhere except possibly where the nuclear density vanishes. The same holds for the (gauge-independent) Berry curvature*

$$\bar{\chi}(R) \neq 0 \;\Rightarrow\; \bar{A}_k(R) = 0. \tag{5.58}$$

$$\bar{\chi}(R) \neq 0 \;\Rightarrow\; \bar{\Omega}_{ab}(R) = 0. \tag{5.59}$$

Proof. To get from Eq. (5.57) to Eq. (5.59) we use that $n(R)$ is continuous and hence whenever it is non-zero for a given R, then it is also non-zero in a neighborhood, implying that not only $\bar{A}(R)$, but also its derivatives are zero. □

This means that the Berry phase is zero for all paths except if they contain nodes of the nuclear wave function. We stress that here "contains nodes" refers to points on the path - *paths winding around a node are covered*. The exclusion of the nodal structure of the nuclear wave function is natural, as the gauge connection and the Berry curvature are not well defined there.

We will see in the next section that nodes appear only wherever the Fermi exclusion principle demands it, i.e. whenever two fermionic nuclei have the same coordinates. Consequently, for any path that contains a node, one can make a small deformation of this path to avoid the node (Any generic deformation would achieve this). This is true for any dimension $d \geq 2$.

Hence, it will turn out that the exceptional set where $\bar{\chi}(R) = 0$ is a removable singularity and the Berry curvature is zero everywhere.

5.3.2 Nodal structure

As we have learned in the previous paragraph, we need to understand the nodal structure of the nuclear wave function in order to learn more about the Berry curvature and the gauge potential. The general properties of the nodal structure of a wave function is an active field of research [73, 74]. However, this refers to wave functions directly derived from a Schrödinger equation, rather than a function obtained by integrating out degrees of freedom after solving a Schrödinger equation.

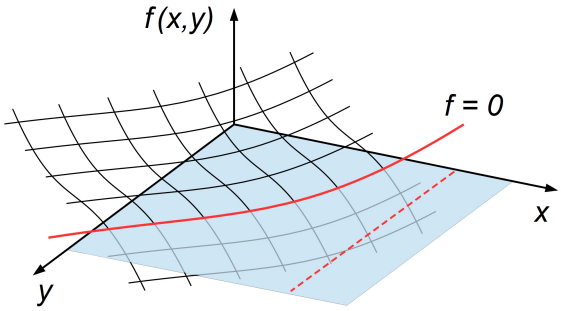


Figure 5.1: A function in two variables and its nodal line (red). A function with a nodal line parallel to the y-axis (like the dashed red line) would lead to a function with a zero, when its modulus is integrated versus y.

Based on literature on the exact factorization and the discussion of the nodal structure [75, 76] one could get the idea that the nuclear wave function of the exact factorization is nodeless in general. Unfortunately this is not the case. Obviously the nuclear wave function has nodes if fermionic nuclei are present.

In order to illustrate what happens when we integrate out the electronic degrees of freedom, consider a function of two real coordinates $f(x, y)$. We define the integrated function

$$g(x) = \int \mathrm{d}y \; |f(x, y)| \tag{5.60}$$

and ask: When does g have a node? For a generic function f, we can expect that the space where f is zero is a one dimensional manifold, i.e. one or several lines in the x-y-plane, see Fig. 5.1. Now the integrand in Eq. (5.60) is non-negative. Hence if f is continuous $g(x)$ is only zero, if f is zero for all y simultaneously at this x. In other words g can only acquire nodes, if f has a nodal line parallel to the y-axis. This however is not to be expected for a generic function f, except if there is more known about it. For instance if it can be written as a single product of one function depending on x and another depending on y then g would share the nodes of the x-dependent function. Another example would be that f is antisymmetric with respect to x. Then g would automatically have a node at $x = 0$.

After this example let us come back to the nuclear wave function $\bar{\chi}(R)$ that corresponds to a given total wave function $\psi(R, r)$ via

$$|\bar{\chi}(R)|^2 = \int \mathrm{d}r \; |\psi(R, r)|^2 \tag{5.61}$$

The situation is very similar to the example just discussed: The integrand is still a non-negative function and hence $\bar{\chi}(R)$ is only zero for a given R if $\psi(R,r)$ is zero for all electronic coordinates r simultaneously. The squares appearing do not change the discussion, as the square of a function is zero whenever the function is zero and vice versa. What is different is that we are now dealing with a high dimensional space, namely $r \in \mathbb{R}^{d\cdot N_\mathrm{e}}$ and $R \in \mathbb{R}^{d\cdot N_\mathrm{n}}$, where as before d is the number of spatial dimension (usually 3) and N_e and N_n are the number of electrons and nuclei, respectively. We are interested in the space

$$\begin{aligned} \Lambda_{\chi=0} &:= \{R \mid \bar{\chi}(R) = 0\} && (5.62)\\ &= \{R \mid \forall r : \psi(R,r) = 0\}. && (5.63) \end{aligned}$$

If we consider time reversal symmetric systems (i.e. we do not allow for an external magnetic field) the wave function can be chosen real and the condition to have a node $\psi(R,r) = 0$ is one constraint (without time reversal symmetry these would be two constraints, one for the real and one for the imaginary part). This reduces the dimensionality of the total space by one, leaving us with a $(d(N_\mathrm{e}+N_\mathrm{n})-1)$-dimensional nodal manifold for the total wave function (we exclude the already discussed case that the wave function is positive everywhere)

$$\begin{aligned} \Lambda_{\psi=0} &:= \{(R,r) \mid \psi(R,r) = 0\} && (5.64)\\ d_{\psi=0} &:= \dim(\Lambda_{\psi=0}) = d(N_\mathrm{e}+N_\mathrm{n}) - 1. && (5.65) \end{aligned}$$

There may be two or more identical fermionic nuclei in the system. These are important for the nodal structure. We define

$$\begin{aligned} \Lambda_{\texttt{Pauli}} := \{R \mid R_i = R_j, &\text{ for some } i \neq j \text{ that correspond}\\ &\text{to identical fermionic nuclei}\}. \end{aligned}$$

By Pauli's exclusion principle the wave function is zero whenever two identical fermionic nuclei have the same coordinates. Hence

$$\Lambda_{\texttt{Pauli}} \subset \Lambda_{\chi=0}. \tag{5.66}$$

Now imagine sitting on a point (R,r) with $R \notin \Lambda_{\texttt{Pauli}}$ but $\psi(R,r) = 0$. This is a point on the nodal manifold of ψ that is not automatically zero by the exclusion principle. We want to know whether or not this point leads to a node in $\bar{\chi}$, i.e. is $R \in \Lambda_{\chi=0}$? For this to be possible the normal vector on the nodal hyper surface must be orthogonal to all directions that change only electronic coordinates. Even if it is, this vector generally will change its direction when we change the electronic

coordinates. If the total wave function can be written as a single product of a nuclear wave function and an electronic wave function that depends on r only, then a node of this nuclear wave function would lead to a nodal hyper surface of the total wave function which is parallel to all electronic coordinate directions simultaneously. However for an interacting system, one would in general expect that this is not the case. Instead the nodal manifold is expected to be curved, in the sense that the normal vector will not stay perpendicular to all electronic coordinate axes upon arbitrary change of electronic coordinates. We can conclude

Lemma 7. *If the total wave function ψ is an eigenstate of a Hamiltonian with a generic interaction between nuclei and electrons then the only nodes of the nuclear wave function are those dictated by the exclusion principle:*

$$\Lambda_{\chi=0} = \Lambda_{\texttt{Pauli}}. \tag{5.67}$$

We put the term "generic interaction" because the arguments that lead to this result are based on the tricky assumption that the wave function is itself in a way generic. Interestingly, a similar observation was already made for the special case of diatomic molecules in Ref. [77]. It was also argued that $\bar{\chi}$ should have no node at all, based on the Born-Huang expansion (see Ref. [75]). The idea is conceptually quite similar to the reasoning presented here, though they did not consider fermionic nuclei, as already pointed out in Ref. [78].

Though we find it plausible that the result is valid in general for the Coulomb interaction, it is not difficult to construct an example of an interacting Hamiltonian which leads to any nodal structure one desires: We can take a wave function with the nodal structure we want to have, e.g. the nodal structure of a wave function which is a single product, like the one of a non-interacting system. We can then construct an interaction potential by inverting the Schrödinger equation

$$V(R,r) = -\frac{(T_\mathrm{n} + T_\mathrm{e} - E)\psi(R,r)}{\psi(R,r)}. \tag{5.68}$$

The wave function with our desired nodal structure is a solution to the Schrödinger equation with this potential. In order to ensure that one obtains a truly interacting system one can multiply the wave function with the desired nodal structure by a strictly positive function of R and r that does not factorize. This guarantees that the obtained potential V describes a truly interacting system, while the nodal structure remains the same.

Note that it is not hard to construct such an interaction potential that is rather regular, i.e. a potential that is both smooth and bounded can be achieved by starting from a system without electron-nuclear interaction and modifying the wave

function only in regions where it is non-zero. This guarantees that the potential is equal to the one of the non-interacting system in the vicinity of the nodes of ψ and is modified only where the expression in Eq. (5.68) has a non-zero denominator and hence is smooth and bounded. The question, whether one can also find two-particle interactions that lead to a given nodal structure is probably much harder.

Let us finally remark that one of course has to treat results that are based on the argument that something is "generic" with caution. If for instance we had overlooked that Pauli's exclusion principle dictates nodes of the nuclear wave function, our conclusion would have been that $\bar{\chi}$ has no nodes at all, which is not correct (this point was not mentioned in Ref. [75]). Symmetries can indeed lead to situations which one would not expect "in a generic situation". A famous example is the hydrogen atom in which the excited states are degenerate due to a not so obvious symmetry of the Coulomb interaction [79]. Hence one might be worried that the reasoning above does not apply to the Coulomb interaction in particular. Though this is logically possible, it is not very plausible. The reason being that the presence of a Berry phase in the usual Born-Oppenheimer sense is stable with respect to slight changes of the shape of the Coulomb interaction or the external field and hence not depending on any symmetry properties of them. Note also that spatial symmetries do not imply any nodes in the nuclear wave function [80].

5.3.3 Time reversal symmetry versus rotational symmetry

After we established that the Berry phase does not appear in the framework of the exact factorization whenever the wave function considered has a vanishing nuclear current, one might ask for situations where this happens. The answer to this is time reversal symmetry. If we think of non-relativistic systems without external magnetic fields, the time reversal symmetry operator commutes with the Hamiltonian of the molecule. Hence we can choose the eigenstates of the Hamiltonian to be eigenstates of this symmetry operator as well. As the current is odd under time reversal symmetry, it must vanish for these states. Hence Theorem 8 actually tells us that for any molecule the eigenstates can be chosen in such a way that no Berry phase appears.

On the other hand we might want to choose our energy eigenstates to be eigenstates of the total angular momentum operator. In this case the nuclear current is not zero except for total angular momentum zero. Hence for this case the theorem is not applicable. However, note that the discussion about the nodal structure of the nuclear wave function is still valid.

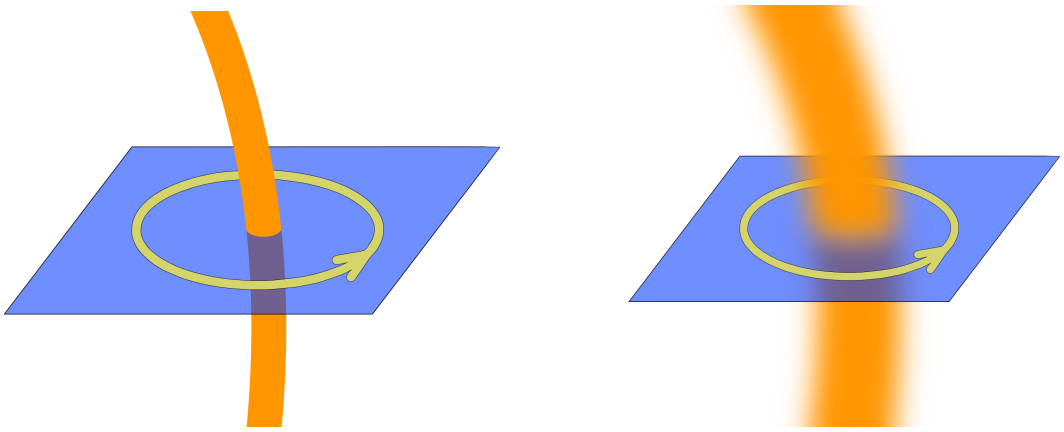


Figure 5.2: When the Berry curvature is distribution valued on a line (left), any path that winds around it will lead to the same Berry phase. The value of the phase does not depend on the details of the path, but only on its topology. In the exact factorization, the Berry curvature is smeared out along a line-like region (right). Deformations of the path lead to a continuous change of the the Berry phase.

5.3.4 States with nuclear current

As we discussed in the previous section, one might want to consider nuclear wave functions that have a non-zero nuclear current. One may then wonder whether or not a Berry phase appears.

In the usual description of conical intersections one talks about the so-called seam space, which is the set of nuclear coordinates for which two potential energy surfaces are degenerate. For the most common conical intersections [46] this space has a dimension of the total space minus two and the Berry phase for a path that winds around such an intersection is an odd multiple of π. In such a case the Berry curvature is zero everywhere except for the seam space, on which it is distribution valued. We will now show that such a situation is not possible in the exact factorization.

Theorem 9. *Under the same conditions as for Lemma 7 and for $d = 3$, there is no topological Berry phase in the exact factorization, i.e. any closed path in coordinate space can be shrunk to a point in such a way that the associated Berry phase changes continuously to zero.*

Proof. Fortunately it has already been shown that the wave function ψ for a bound eigenstate of a molecule is guaranteed to be continuous and have bound first partial derivatives everywhere except for the coordinates where the Coulomb interaction is singular [52]. The latter set of coordinates is again a set of the total dimension of

the space minus 3. The theorem of dominated convergence allows us to conclude that this continuity and the property of a bounded first partial derivative transfers to the nuclear wave function in the positive gauge Eq. (5.55) everywhere except where $\bar{\chi}(R) = 0$. With this it is immediate that also the electronic wave function Eq. (5.56) is continuous (this was already observed in Ref. [81], though they excluded the coordinates where the coulomb potential is singular, rather than the set where the nuclear wave function vanishes) and has bounded first derivatives in any compact set where $\bar{\chi} \neq 0$. That is all we need to conclude that also the Berry connection we defined is bounded and continuous for these sets. Now for $d = 3$ any closed loop that does not run through a point on which $\bar{\chi}(R) = 0$ can be continuously deformed and be shrunk to a point where $\bar{\chi}(R) \neq 0$, without running trough a node at any time (Remember that by Lemma 7, the space where $\bar{\chi}(R) = 0$ is a set of the total dimension of the space minus 3). During this deformation the Berry phase can also only change in a continuous manner, as we established that $\bar{A}$ is continuous and bounded. As we shrink the loop to a point the Berry phase must clearly tend to zero, again because $\bar{A}$ is continuous and bounded. □

This implies that a situation where the Berry phase is an odd multiple of π for loops like in the usual description of conical intersections as mentioned above is not possible. This implies that the Berry phase of the exact factorization is not a topological quantity as in the Born Oppenheimer expansion case. The Berry curvature is no longer distribution valued, but at best a smeared out function, as sketched in Fig. 5.2 [80].

All in all the Berry phase, as it appears in the exact factorization is conceptually very different from the one that is defined in terms of the Born Oppenheimer expansion, despite the similarities in the defining equations. While the latter is defined solely in terms of the electronic Hamiltonian, the former depends on the details of the full Hamiltonian. Namely, the masses of the nuclei and also the nucleus-nucleus interaction potential can change the Berry phase of the exact factorization. Another feature that is not present in the original Berry phase, is the dependency on the choice of the total wave function. Though the Berry phase of the exact factorization can be non-trivial in general, in time reversal symmetric systems, there is always a choice of eigenstates that shows no Berry phase at all. Even if a state without time reversal symmetry is chosen, there is still a fundamental difference between the two kinds of Berry phase. While the original Berry phase can have a topological character, this is not possible for the Berry phase of the exact factorization. Instead of a distribution valued Berry curvature, one is left with a smeared out function.

5.3.5 Structure of the nuclear wave function and the potential energy surface

When two or more of the nuclei under consideration are identical fermions, it seems natural to try to chose a nuclear wave function that is antisymmetric with respect to exchange of those particles. This, however, is more problematic than one might anticipate.

Lemma 8. *In $d = 2$ dimensions, the nuclear wave function $\bar{\chi}$ cannot be chosen real, antisymmetric with respect to exchange of two nuclei and continuous simultaneously.*

Proof. We assume that $\bar{\chi}$ is antisymmetric and real. We have to show that it cannot be continuous at the same time. Let $\boldsymbol{R} \in \mathbb{R}^2$ be the relative coordinate of two nuclei. Then antisymmetry with respect to to exchange of these two nuclei means

$$\bar{\chi}(-\boldsymbol{R}) = -\bar{\chi}(\boldsymbol{R}). \tag{5.69}$$

We consider all other coordinates fixed in such a way that no other pair of nuclei happens to be on top of another. Then, by virtue of Lemma 7 we can pick a radius $|\boldsymbol{R}_0|$ such that

$$|\bar{\chi}(\boldsymbol{R})| > 0 \qquad \text{for all } \boldsymbol{R} \text{ with } |\boldsymbol{R}| = |\boldsymbol{R}_0|. \tag{5.70}$$

We can assume w.l.o.g. that $\bar{\chi}(\boldsymbol{R}_0) > 0$. Eq. (5.69) then implies that $\chi(-\boldsymbol{R}_0) < 0$. This means that on the circle $|\boldsymbol{R}| = |\boldsymbol{R}_0|$, $\bar{\chi}$ is a nodeless function, which takes on positive as well as negative values. In this situation the intermediate value theorem implies that $\bar{\chi}$ cannot be continuous. □

It is probably not a good idea to drop the continuity of the nuclear wave function: This would introduce discontinuities of the electronic wave function as well, as it has to compensate the discontinuities in such a way that the total wave function is continuous. So, if one wants the nuclear wave function to be antisymmetric, choosing it complex is the only choice left. However, in the case that the total wave function is real, the gauge connection is zero only when choosing a real nuclear wave function as well. Hence one can either choose a gauge in which there is no A-field, or choose the nuclear wave function antisymmetric. Both at once is not possible. The situation is even more problematic in the three dimensional case.

Lemma 9. *In $d = 3$ dimensions, the nuclear wave function $\bar{\chi}$ cannot be chosen both antisymmetric with respect to exchange of two nuclei and continuous.*

Proof. As in Lemma 8, we assume antisymmetry

$$\bar{\chi}(-\boldsymbol{R}) = -\bar{\chi}(\boldsymbol{R}) \tag{5.71}$$

and use Lemma 7 to ensure that

$$|\bar{\chi}(\boldsymbol{R})| > 0 \qquad \text{for all } \boldsymbol{R} \text{ with } |\boldsymbol{R}| = |\boldsymbol{R}_0|, \tag{5.72}$$

for some $\boldsymbol{R}_0$. This time $\boldsymbol{R}, \boldsymbol{R}_0 \in \mathbb{R}^3$. We use reductio ad absurdum, i.e. we assume that $\bar{\chi}$ is continuous and seek a contradiction. Towards this end we consider a continuous path $\boldsymbol{R}(s)$ on the sphere with $|\boldsymbol{R}| = |\boldsymbol{R}_0|$, where $s \in [0, 1]$. On this path $\bar{\chi}$ takes complex values. This defines a continuous path $\bar{\chi}(\boldsymbol{R}(s))$ in the complex plane. According to Eq. (5.72) this path does not run through zero. For a closed path a winding number can be defined [82]. This is the number of times the complex number winds around zero in counter-clock-wise sense minus the number how often it winds around zero in clock-wise sense.

We can now consider the path that runs along the equator of the sphere. Let w.l.o.g. R_0 be the starting point of this path. As $\bar{\chi}(-\boldsymbol{R}_0) = -\bar{\chi}(\boldsymbol{R}_0)$, after half a rotation, the phase of $\bar{\chi}$ changed by a total of

$$\pi + 2\pi k, \tag{5.73}$$

with an integer k. The antisymmetry ensures that the remaining half of the circle changes the phase of $\bar{\chi}$ by the same amount, with the same orientation. The total change of phase is hence

$$2(\pi + 2\pi k) \;=\; (2k + 1)2\pi, \tag{5.74}$$

i.e. the winding number is odd.

Now, we consider deformations of the path. As $\bar{\chi}$ is assumed to be continuous and non-zero this cannot change the winding number. If we shrink the path $\boldsymbol{R}(s)$ to a point on the sphere, also the path of $\bar{\chi}(\boldsymbol{R}(s))$ in the complex plane shrinks to a point. This point cannot be zero, as $\bar{\chi}$ is non-zero on the sphere. In this limit the winding number is clearly zero, in contradiction to being odd. □

This Lemma shows that any choice of the nuclear wave function that respects the antisymmetry of the nuclei (like the one suggested in Ref. [78]) necessarily renders it discontinuous. It is hence probably more natural to simply choose the gauge in which the nuclear wave function is positive everywhere. The antisymmetry is then absorbed in the parametric dependence of the electronic wave function on the nuclear coordinates. The nuclear wave function will then have a cusp at coordinates

on which two fermionic nuclei are on top of each other. This might remind one of the cusp condition [52, 83]. However, the cusps discussed here emerge from the exchange symmetry alone. They are present even when considering an interaction potential that is not as singular as the Coulomb interaction.

The existence of cusps in the nuclear wave function imply that the potential energy surface diverges near fermionic nuclei that are getting close to one another. To see this consider a nuclear wave function that can be approximated by a constant times the modulus of $\boldsymbol{R}$ in a region near $\boldsymbol{R} = 0$

$$\bar{\chi}(\boldsymbol{R}) = a|\boldsymbol{R}| + \mathrm{O}(\boldsymbol{R}^2), \tag{5.75}$$

where $\boldsymbol{R}$ is again the relative coordinate of two identical fermionic nuclei. Then for vanishing A-field, the potential energy surface can be obtained by inverting Eq. (5.45), providing

$$\epsilon(\boldsymbol{R}) \;=\; E + \frac{\nabla^2\chi}{2\chi} \;\approx\; \begin{cases} a\,|\boldsymbol{R}|^{-2} & \text{for } d = 3 \\ \frac{a}{2}\,|\boldsymbol{R}|^{-2} & \text{for } d = 2\ , \end{cases} \tag{5.76}$$

where on the R.H.S only the most singular part near $\boldsymbol{R} = 0$ was written out. The result can be obtained by using the Laplace operator in spherical or polar coordinates, respectively. Note that the potential energy surface is gauge-independent [71, 78]. Fortunately, this singularity is in a region of nuclear coordinate space, in which the wave function is usually small anyways due to the nucleus-nucleus repulsion.

It is interesting to study the behavior of the potential energy surface, in the case that the Born Oppenheimer approximation is already very accurate. In such a situation the Born Oppenheimer expansion

$$|\psi(R)\rangle = \sum_a \chi_a(R)|a_R\rangle \tag{5.77}$$

is dominated by a single term

$$\chi_1 \gg \chi_a, \qquad a = 2, 3, 4, \dots\ , \tag{5.78}$$

where the electronic state with index one may or may not be the electronic ground state. To be more accurate, let us assume that

$$\chi_1 = (1 - \lambda^2)^{\frac{1}{2}}\ \tilde{\chi}_1, \tag{5.79}$$

$$\chi_a = \lambda\ \tilde{\chi}_a, \qquad a = 2, 3, 4, \dots\ , \tag{5.80}$$

where λ is a positive real number and $\tilde{\chi}_a$ are some λ-independent nuclear wave functions. We can then consider the leading terms in the $\lambda \to 0$ limit. Note that

this is kind of a short cut to get an idea of what happens if the Born-Oppenheimer approximation is good. To make more accurate statements, one should study the high mass limit more carefully, but we keep this simpler approach here.

The $\lambda \to 0$ limit has particularly interesting consequences in regions where $\chi_1(R)$ has a node. Expressing the nuclear wave function of the exact factorization $\bar{\chi}$ in terms of χ_a and using the relationship to the potential energy surface Eq. (5.45) leads to

$$\epsilon(R) \approx \frac{|\nabla\chi_1(R)|^2}{2|\bar{\chi}(R)|^2} = \frac{|\nabla\chi_1(R)|^2}{2\sum_a |\chi_a(R)|^2}, \tag{5.81}$$

where only the term that diverges in the $\lambda \to 0$ limit was kept. This shows that the potential energy surface is peaked on the hyper-surfaces of nodes of the dominating nuclear wave function of the Born Oppenheimer expansion. The height of these peaks diverges in the Born Oppenheimer limit. This result can be viewed as a generalization and improvement of an observation that was made in Ref. [77]. They found that for the special case of a non-rotating diatomic molecule (i.e. for the case that there is only one dimension in the space of nuclear coordinates left) in the Born Oppenheimer limit, the height of the potential exceeds the total energy. It also shows that a similar analysis done in Ref. [75] comes to the right conclusion for the wrong reason. There this leading term was neglected (which is not a sensible approximation) and it was argued that the potential energy surface has a barrier (which is correct, as we have shown here).

This is bad news, if one is attempting to solve the time-independent or time-dependent equations (which we did not discuss here; see Ref. [72]) for the wave functions of the exact factorization. Firstly, because peaks in the potential energy surface make the numeric solution of the equations more difficult. And secondly and more importantly, one needs to find good approximations to determine the potential energy surface in the first place. There appears to be no way to locate such peaks without prior knowledge of the nodal structure of the Born Oppenheimer nuclear wave functions. One should hence carefully check any approximation based on these equations regarding the behavior in the case that the BO description shows nodes in the nuclear wave function.

Chapter 6

Summary

I studied Hedin's equations on several levels. First I discussed the equations on a symbolic level. For that I introduced the concept of the vertex insertion operator, which is the diagrammatic correspondence of the Martin-Schwinger functional derivative operator. Based on this, a general theory on equations in terms of Feynman diagrams (termed grapherential equations) and the emerging formal expansions was developed. I proved a general existence and uniqueness theorem and applied it to a number of systems of such grapherential equations. The equation of motion of the Green's function can be closed using the Martin-Schwinger functional derivative to re-express the two body Green's function in terms of the one-body Green's function. This equation served as the first application of the developed formalism, which immediately yields existence and uniqueness of the solution. The proof that this solution is the well known perturbative series of all connected Feynman diagrams is surprisingly straight forward within this formalism.

Starting from there, I found other systems of equations that generate expansions in terms of dressed quantities, namely the Hartree Green's function, the screened interaction and eventually also the full interacting Green's function, culminating in Hedin's equations, which generate expansions of the selfenergy and polarization in terms of the interacting Green's function and the dressed Coulomb interaction.

These findings can either be viewed in the language of Feynman diagrams in which I developed the formalism or be translated to formal expansions in functions like the Green's function or the screened interaction. In the latter case the introduced vertex insertion operator is a functional differential operator instead. While the latter view is more commonly used, the first one is more rigorous in the sense that there is no doubt that all the introduced objects are well defined, the series converge and all operations are meaningful.

I also studied Hedin's equations as non-linear partial differential equations and found that restricting the space of functions considered to analytic ones renders the solution unique. For the case of one space-time point only, called the one-point-

model, I discussed the general solution to the emerging set of equations.

The most common approximation to Hedin's equations is the *GW* approximation. I found that on a discretized finite space-time, it is enough to demand that the solution is bounded in the non-interacting limit to render it unique. As I found that all other solutions diverge in this limit, one can conclude that this is also the solution closest to the exact one if the interaction is only small enough. Convergence to this unique solution is no longer guaranteed when iterating the equations. Instead, I found that for any given starting point of the iteration, one can choose the interaction strength small enough to make the process converge to this unique solution. Though this explains, why the approach works in cases in which *GW* is only a small correction to the starting solution, it also tells us that one needs to be cautious in situations where the solution is no longer close to the starting point.

In summary I found for all the cases studied an appropriate space of solutions, which is sufficiently small to render the solution unique. This is an important step to improve our understanding of the equations we are looking at.

Finally I discussed the exact factorization of the electron-nuclear wave function. I found that for states without nuclear currents, no Berry phase or gauge connection appears. This means in particular that for molecules without external magnetic field in the non-relativistic description, one can choose the eigenstates of the total Hamiltonian in such a way that there is no Berry phase at all (and therefore no gauge connection) in the exact factorization. Even for states with nuclear current, I was able to show that no topological Berry phase can appear. This fits in very well with the findings in Ref. [80], where for a specific model system, only a smeared out, rather than a topological Berry curvature appears and a choice of wave function without a Berry phase is possible.

I was able to show that in the exact factorization a continuous nuclear wave function cannot be antisymmetric under exchange of nuclei. Also the potential energy surface of the exact factorization necessarily has divergences, whenever identical fermionic nuclei of the same spin are present. I also argued that the potential energy surface may have peaks that diverge in the Born Oppenheimer limit, indicating that a numerical treatment may be difficult also for finite nuclear masses. Such peaks were already found numerically for certain model systems [80, 84].

Acknowledgements I am indebted to Kay Dewhurst for supervising me during this work. Not only did he always have an open door whenever I had questions and was willing to discuss anything in depth, but he also managed to push me to keep focusing on the relevant and interesting aspects of the theory under investigation, just by asking the right questions and asking them again if I ever lost target, until

they would be answered. Though Kay and Hardy suggested which topics would be interesting to investigate, I never felt pushed in any particular direction. On the contrary I had the freedom to focus on whatever I considered interesting and to go deeper in which ever direction I thought was worth further investigations. I am also very thankful and at the same time impressed how much time and effort Hardy was willing to invest in detailed scientific discussions about the topics I was working in. There is hardly anything more motivating and encouraging than sincere interest of others in obtained results.

The topics I worked on were all really original fundamental research, making them particularly interesting. However, such topics are naturally risky, in the sense that it was not at all clear, weather any results could be found in a particular direction at all. I really appreciate that both Kay and Hardy communicated this very well from the very beginning and still gave me the opportunity to tackle these questions.

I thank Seung Kyu Min, Federica Agostini, Ali Abedi, Axel Schild and Ryan Requist for each of them helping me a lot to get into the topic of Berry phases and the exact factorization. This goes to Seung Kyu Min and Ryan Requist in particular for countless sometimes day-filling constructive discussions about these topics. Similarly I am very grateful to Pina Romaniello, Lucia Reining, Giovanna Lani, Arjan Berger and Adrian Stan for many interesting discussions on the subject of Hedins equations and multiple solutions. The same goes for Robert van Leeuwen, who pointed me on the interesting open question about the convergence of symbolically iterating Hedin's equations which is covered in the first part of this work.

I also thank Martin Hoffmann, Sebastian Frank, Attila Changi, Arkady Davidov, Neha Sardana, Antonio Sanna, Sangeeta Sharma, David Jacob, Tim Baldsiven, Oleg Brovko, Alexander Senichev, Andreas Linscheid, Frank Essenberger, Julia Beck, Sajeela Hunger and everyone from the theory apartment as well as the IMPRS group for making the time in Halle just as amazing and unforgettable as it was. I also thank my family for always supporting me and beeing there for me.

Bibliography

[1] Ross E. Larsen and Benjamin J. Schwartz. Efficient real-space configuration-interaction method for the simulation of multielectron mixed quantum and classical nonadiabatic molecular dynamics in the condensed phase. *The Journal of Chemical Physics*, 119(15):7672–7684, October 2003.

[2] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal. Quantum monte carlo simulations of solids. *Reviews of Modern Physics*, 73(1):33–83, January 2001.

[3] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Physical Review*, 136(3B):B864–B871, November 1964.

[4] W. Kohn and L. J. Sham. Self-Consistent equations including exchange and correlation effects. *Physical Review*, 140(4A):A1133–A1138, November 1965.

[5] E. Gross and W. Kohn. *Time-Dependent Density-Functional Theory*, volume 21 of *Advances in Quantum Chemistry*, pages 255–291. Elsevier, 1990.

[6] Erich Runge and E. K. U. Gross. Density-Functional theory for Time-Dependent systems. *Physical Review Letters*, 52(12):997–1000, March 1984.

[7] Hideo Fukutome. Theory of the unrestricted Hartree-Fock equation and its solutions. i. *Progress of Theoretical Physics*, 45(5):1382–1406, 1971.

[8] Pablo Echenique and J. L. Alonso. A mathematical and computational review of HartreeFock SCF methods in quantum chemistry. *Molecular Physics*, 105(23-24):3057–3098, December 2007.

[9] Richard E. Stanton. Multiple solutions to the Hartree-Fock problem. i. general treatment of Two-Electron Closed-Shell systems. *The Journal of Chemical Physics*, 48(1):257–262, 1968.

[10] P. L. Lions. Solutions of Hartree-Fock equations for coulomb systems. 109(1):33–97, 1987.

[11] Peter Pulay and Rui F. Liu. Methods for finding unrestricted Hartree-Fock solutions and multiple solutions. *J. Phys. Chem.*, 94(14):5548–5551, July 1990.

[12] N. Helbig, I. V. Tokatly, and A. Rubio. Exact Kohn-Sham potential of strongly correlated finite systems. *The Journal of Chemical Physics*, 131(22):224105+, August 2009.

[13] Klaus Capelle and G. Vignale. Nonuniqueness of the potentials of Spin-Density-functional theory. *Physical Review Letters*, 86(24):5546–5549, June 2001.

[14] E. Prodan. Symmetry breaking in the self-consistent KohnSham equations. *Journal of Physics A: Mathematical and General*, 38(25):5647–5657, June 2005.

[15] Lucas O. Wagner, E. M. Stoudenmire, Kieron Burke, and Steven R. White. Guaranteed convergence of the Kohn-Sham equations. *Physical Review Letters*, 111(9), August 2013.

[16] L. D. Landau. The theory of a fermi liquid. *Soviet Physics Jetp-Ussr*, 3.6:920–925, 1957.

[17] A. L. Fetter, J. D. Walecka, and Leo P. Kadanoff. Quantum theory of many particle systems. *Physics Today*, 25(11):54–55, December 2008.

[18] F. Aryasetiawan and O. Gunnarsson. The GW method. *Reports on Progress in Physics*, 61(3):237–312, January 1999.

[19] V. M. Galitskii. The energy spectrum of a non-ideal fermi gas. *SOVIET PHYSICS JETP-USSR*, 7(1):104–112, 1958.

[20] V. M. Galitskii and A. B. Migdal. Application of quantum field theory methods to the many body problem. *Zhur. Eksptl'. i Teoret. Fiz.*, 34, 1958.

[21] Robert van Leeuwen and Gianluca Stefanucci. Equilibrium and nonequilibrium many-body perturbation theory: a unified framework based on the Martin-Schwinger hierarchy. *Journal of Physics: Conference Series*, 427:012001+, March 2013.

[22] Gianluca Stefanucci and Robert van Leeuwen. *Nonequilibrium Many-Body Theory of Quantum Systems: A Modern Introduction*. Cambridge University Press, 1 edition, April 2013.

[23] Predrag Cvitanović. Asymptotic estimates and gauge invariance. *Nuclear Physics B*, 127(1):176–188, August 1977.

[24] Predrag Cvitanović, B. Lautrup, and Robert B. Pearson. Number and weights of feynman diagrams. *Physical Review D*, 18(6):1939–1949, September 1978.

[25] Alex Buchel and James P. Sethna. Elastic theory has zero radius of convergence. *Physical Review Letters*, 77:1520–1523, August 1996.

[26] Robert Damburg, Rafail Propin, Sandro Graffi, Vincenzo Grecchi, Evans Harrell, Jiří Čížek, Josef Paldus, and Harris Silverstone. $1/r$ expansion for h_2^+: Analyticity, summability, asymptotics, and calculation of exponentially small terms. *Physical Review Letters*, 52(13):1112–1115, March 1984.

[27] F. Dyson. Divergence of perturbation theory in quantum electrodynamics. *Physical Review*, 85(4):631–632, February 1952.

[28] Carl M. Bender and Tai T. Wu. Anharmonic oscillator. *Physical Review Online Archive (Prola)*, 184:1231–1260, August 1969.

[29] W. Janke. Large-order perturbation theory of the zeeman effect in hydrogen from a four-dimensional anisotropic anharmonic oscillator. *Physical Review A*, 41(11):6071–6084, June 1990.

[30] Lars Hedin. New method for calculating the One-Particle green's function with application to the Electron-Gas problem. *Physical Review*, 139(3A):A796–A823, August 1965.

[31] Giovanna Lani, Pina Romaniello, and Lucia Reining. Approximations for many-body green's functions: insights from the fundamental equations. *New Journal of Physics*, 14(1):013056+, January 2012.

[32] L. G. Molinari. Hedin's equations and enumeration of feynman diagrams. *Physical Review B*, 71:113102+, March 2005.

[33] L. G. Molinari and N. Manini. Enumeration of many-body skeleton diagrams. *The European Physical Journal B - Condensed Matter and Complex Systems*, 51(3):331–336, June 2006.

[34] Y. Pavlyukh and W. Hübner. Analytic solution of hedin's equations in zero dimensions. *Journal of Mathematical Physics*, 48(5):052109+, 2007.

[35] Clay Mathematics Institute. Existence and smoothness of the Navier-Stokes equation. `http://www.claymath.org/millennium-problems`. [Online; accessed January 2016].

[36] L. Hedin and S. Lundqvist. Effects of Electron-Electron and Electron-Phonon interactions on the One-Electron states of solids. In *Solid State Physics, Volume 23*. Academic Press, New York, San Francisco, London, 1969.

[37] Lars Hedin. On correlation effects in electron spectroscopies and the GW approximation. *Journal of Physics: Condensed Matter*, 11(42):R489+, October 1999.

[38] J. M. Luttinger and J. C. Ward. Ground-State energy of a Many-Fermion system. II. *Physical Review Online Archive (Prola)*, 118(5):1417–1427, June 1960.

[39] A. J. Sommese and C. W. Wampler. *The Numerical Solution of Systems of Polynomials*. World Scientific, 2005.

[40] B. Buchberger. Gröbner bases: An algorithmic method in polynomial ideal theory. In *Multidimensional Systems Theory*. New York: van Nostrand Reinhold, 1982.

[41] K. Fritzsche and H. Grauert. *From Holomorphic Functions to Complex Manifolds*. Springer-Verlag, 2002.

[42] Wikipedia. Implicit function theorem. `https://en.wikipedia.org/wiki/Implicit_function_theorem`. [Online; accessed January 2016].

[43] Walter Metzner. Linked-cluster expansion around the atomic limit of the hubbard model. *Physical Review B*, 43(10):8549–8563, April 1991.

[44] Matthew L. Leininger, Wesley D. Allen, Henry F. Schaefer, and C. David Sherrill. Is Mo/llerPlesset perturbation theory a convergent ab initio method? *The Journal of Chemical Physics*, 112(21):9213–9222, June 2000.

[45] Tosio Kato. *Perturbation theory in a finite-dimensional space*, chapter 2, pages 72–148. Springer US, New York, NY, 1982.

[46] Wolfgang Domcke, David R. Yarkony, and Horst Köppel, editors. *Conical Intersections - Electronic Structure, Dynamics and Spectroscopy*, volume 15. World Scientific Publishing Co. Pte. Ltd., 2004.

[47] Arno Bohm, Ali Mostafazadeh, Hiroyasu Koizumi, Qian Niu, and Joseph Zwanziger. *The Geometric Phase in Quantum Systems*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003.

[48] A. P. Jauho. Introduction to the keldysh nonequilibrium green function technique, October 2006.

[49] P. Cvitanovic. *Group Theory - Birdtracks, Lie's and Exceptional Groups*. April 2011.

[50] Rainer Oloff. *Geometrie der Raumzeit*. Vieweg+Teubner Verlag, Wiesbaden, 2004.

[51] Arno Schindlmayr and R. W. Godby. Systematic vertex corrections through iterative solution of hedin's equations beyond the GW approximation. *Physical Review Letters*, 80:1702–1705, February 1998.

[52] Tosio Kato. On the eigenfunctions of many-particle systems in quantum mechanics. *Comm. Pure Appl. Math.*, 10(2):151–177, January 1957.

[53] A. Pelster and K. Glaum. Many-body vacuum diagrams and their recursive graphical construction. *phys. stat. sol. (b)*, 237(1):72–81, May 2003.

[54] Adrian Stan, Pina Romaniello, Santiago Rigamonti, Lucia Reining, and J. A. Berger. Unphysical and physical solutions in many-body theories: from weak to strong correlation. *New Journal of Physics*, 17(9):093045+, September 2015.

[55] Yasutami Takada. Exact self-energy of the many-body problem from conserving approximations. *Physical Review B*, 52(17):12708–12719, November 1995.

[56] Richard D. Mattuck and Physics. *A Guide to Feynman Diagrams in the Many-Body Problem: Second Edition (Dover Books on Physics)*. Dover Publications, 2 edition, June 1992.

[57] W. Fulton and R. MacPherson. Intersecting cycles on an algebraic variety. In *Real and Complex Singularities*. Sijthoff and Noordhoff, 1977.

[58] W. Fulton. *Intersection Theory, no. 2 in Ergebnisse der Math.* Springer-Verlag, 1984.

[59] ElliottH Lieb and Barry Simon. The Hartree-Fock theory for coulomb systems. 53(3):185–194, 1977.

[60] E. Prodan and P. Nordlander. Hartree approximation II: The thermodynamic limit. *Journal of Mathematical Physics*, 42(8):3407–3423, 2001.

[61] Marcel Griesemer and Fabian Hantsch. Unique solutions to HartreeFock equations for closed shell atoms. *Archive for Rational Mechanics and Analysis*, 203(3):883–900, September 2012.

[62] L. Foldy and W. Tobocman. Application of formal scattering theory to Many-Body problems. *Physical Review*, 105(3):1099–1100, February 1957.

[63] Marco Casadei, Xinguo Ren, Patrick Rinke, Angel Rubio, and Matthias Scheffler. Density-functional theory for f electron systems: the α-γ phase transition in cerium, September 2012.

[64] Alex J. W. Thom and Martin H. Gordon. Locating multiple Self-Consistent field solutions: An approach inspired by metadynamics. *Physical Review Letters*, 101:193001+, November 2008.

[65] Karol Kowalski and Karol Jankowski. Towards complete solutions to systems of nonlinear equations of Many-Electron theories. *Physical Review Letters*, 81(6):1195–1198, August 1998.

[66] P. Koval, D. Foerster, and D. Sánchez-Portal. Fully self-consistent *gw* and quasiparticle self-consistent *gw* for molecules. *Phys. Rev. B*, 89:155417, Apr 2014.

[67] Mats-Erik Pistol. N-representability of two-electron densities and density matrices and the application to the few-body problem. *Chemical Physics Letters*, 400(4-6):548–552, December 2004.

[68] David A. Mazziotti. Two-Electron reduced density matrix as the basic variable in Many-Electron quantum chemistry and physics. *Chem. Rev.*, 112(1):244–262, August 2011.

[69] David A. Mazziotti. Structure of fermionic density matrices: Complete N-Representability conditions. *Physical Review Letters*, 108(26):263002+, June 2012.

[70] Aref'eva. Non-Abelian stokes formula. 43(1):353–356, 1980.

[71] Geoffrey Hunter. Conditional probability amplitudes in wave mechanics. *Int. J. Quantum Chem.*, 9(2):237–242, March 1975.

[72] Ali Abedi, Neepa T. Maitra, and E. K. U. Gross. Exact factorization of the Time-Dependent Electron-Nuclear wave function. *Physical Review Letters*, 105:123002+, September 2010.

[73] Arne Lüchow and Tony C. Scott. Nodal structure of schrödinger wavefunction: general results and specific models. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 40(5):851–867, March 2007.

[74] Lubos Mitas. Structure of fermion nodes and nodal cells. *Physical Review Letters*, 96(24), June 2006.

[75] Geoffrey Hunter. Nodeless wave functions and spiky potentials. *Int. J. Quantum Chem.*, 19(5):755–761, May 1981.

[76] R. Lefebvre. Factorized molecular wave functions: Analysis of the nuclear factor. *The Journal of Chemical Physics*, 142(21):214105+, June 2015.

[77] J. Czub and L. Wolniewicz. On the non-adiabatic potentials in diatomic molecules. *Molecular Physics*, 36(5):1301–1308, November 1978.

[78] Nikitas I. Gidopoulos and E. K. U. Gross. Electronic non-adiabatic states: towards a density functional theory beyond the BornOppenheimer approximation. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 372(2011):20130059+, March 2014.

[79] W. Pauli. Über das wasserstoffspektrum vom standpunkt der neuen quantenmechanik. *Zeitschrift für Physik A Hadrons and Nuclei*, 36(5):336–363, May 1926.

[80] Ryan Requist, Falk Tandetzky, and E. K. U. Gross. Molecular geometric phase from the exact electron-nuclear factorization. *Phys. Rev. A*, 93:042108, Apr 2016.

[81] Thierry Jecko, Brian T. Sutcliffe, and R. Guy Woolley. On factorization of molecular wavefunctions. *Journal of Physics A: Mathematical and Theoretical*, 48(44):445201+, November 2015.

[82] Wikipedia. Winding number. `https://en.wikipedia.org/wiki/Winding_number`. [Online; accessed January 2016].

[83] Russell T. Pack and W. Byers Brown. Cusp conditions for molecular wavefunctions. *The Journal of Chemical Physics*, 45(2):556–559, July 1966.

[84] Seung K. Min, Ali Abedi, Kwang S. Kim, and E. K. U. Gross. Is the molecular berry phase an artifact of the Born-Oppenheimer approximation? *Physical Review Letters*, 113(26), December 2014.

Eigene Veröffentlichungen

Ryan Requist, Falk Tandetzky, and E. K. U. Gross. Molecular geometric phase from the exact electron-nuclear factorization. Phys. Rev. A, 93:042108, Apr 2016.

J. A. Berger, Pina Romaniello, Falk Tandetzky, Bernardo S. Mendoza, Christian Brouder, and Lucia Reining. Solution to the many-body problem in one point. New Journal of Physics, 16(11):113025+, November 2014.

F. Essenberger, A. Sanna, A. Linscheid, F. Tandetzky, G. Profeta, P. Cudazzo, and E. K. U. Gross. Superconducting pairing mediated by spin fluctuations from first principles. Physical Review B, 90(21), December 2014.

Björn Oetzel, Frank Ortmann, Lars Matthes, Falk Tandetzky, Friedhelm Bechstedt, and Karsten Hannewald. Large bandwidths in synthetic one-dimensional stacks of biological molecules. Physical Review B, 86(19):195407+, November 2012.

F. Tandetzky, J. K. Dewhurst, S. Sharma, and E. K. U. Gross. Multiplicity of solutions to GW-type approximations. Physical Review B, 92(11), May 2012.

Lebenslauf

Geboren am 26.10.1986 in Schleiz.

Promotion	
2011 - 2016	Promotionsstipendiat am Max Planck Institut für Mikrostrukturphysik, Halle
Hochschulstudium	
2006 - 2010	Physik an der Friedrich Schiller Universitt Jena Abschluss: Diplom Physiker
Schulbildung	
2001 - 2005	Carl Zeiss Gymnasium Jena
1997 - 2001	Christian Gottlib Reichard Gymnasium Lobenstein
1992 - 1997	Grundschule Ebersdorf

Eidesstattliche Erklärung

Ich erkläre an Eides statt, gemäß §5 der Promotionsordnung der Naturwissenschaftlichen Fakultät II der Martin-Luther-Universität Halle-Wittenberg vom 13 Juni 2012, dass ich die vorliegende Arbeit

Effective interactions in the quantum theory of molecular and condensed matter physics

selbständig und ohne fremde Hilfe verfasst, sowie keine als die von mir angegebenen Quellen und Hilfsmittel benutzt und die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. Weiterhin erkläre ich, dass ich bisher keine vergeblichen Promotionsversuche unternommen habe und die Arbeit in gleicher oder ähnlicher Form weder einer anderen Prüfungsbehörde vorgelegt, noch veröffentlicht wurde.

Halle (Saale), den 30. Dezember 2017

Falk Tandetzky

Liste der Gutachter

Prof. Dr. E. Gross, Max-Planck-Institut für Mikrostrukturphysik, Halle

Prof. Dr. M. Marques, Martin Luther Universität Halle

Prof. Dr. G. Stefanucci, Universität Tor Vergata, Rom

Verteidigungsdatum: 6.12.2018