Daniel Kraft

Stochastic Variational Approaches to Non-Hermitian Quantum-Mechanical Problems

Masterarbeit

zur Erlangung des akademischen Grades
Master of Science
an der Naturwissenschaftlichen Fakultät
der Karl-Franzens-Universität Graz

Betreuer:

Univ.-Prof. Dr. Willibald Plessas

Institut für Physik

Graz, am 03. November 2013

Abstract

Resonances are very interesting phenomena, which appear in a wide range of classical and quantum systems. The mathematical description of quantum-mechanical resonances is connected with complex eigenvalues of the Hamiltonian of a given system. This fact introduces various difficulties, especially since resonance eigenfunctions are not normalisable. This work aims at developing techniques for the treatment of quantum-mechanical resonances with variational methods. For this purpose particular similarity transformations of the Hamiltonian, such as the Zel'dovich and complex-scaling transformations, are employed. A precise mathematical description of the transformation operators and extended function spaces is given. In addition a generalisation of the Rayleigh-Ritz variational principle is derived, which can be applied to non-Hermitian eigenvalue problems. Several basis-selection criteria are proposed for adapting the stochastic variational method to the non-Hermitian case. Their efficiencies are demonstrated for a one-dimensional problem with a potential considered before in the literature. It is found that the stochastic variational method can practically be extended to quantum-mechanical resonances and complex eigenvalues. It usually outperforms alternative methods for treating non-Hermitian Hamiltonians.

Kurzzusammenfassung

Sowohl in klassischen als auch in Quantensystemen sind Resonanzen sehr interessante Phänomene. Ihre mathematische Beschreibung ist eng mit komplexen Eigenwerten des Hamilton-Operators eines Systems verknüpft, was zu einigen Schwierigkeiten in der Handhabung führt, da die zugehörigen Eigenfunktionen nicht normierbar sind. Das Ziel dieser Arbeit ist es zu beschreiben, wie Variationsmethoden zur Behandlung von quantenmechanischen Resonanzen verwendet werden können. Dafür werden Ähnlichkeitstransformationen wie die Zel'dovich- und die Complex-Scaling-Transformation angewandt, und es wird eine mathematisch präzise Beschreibung der involvierten Operatoren und erweiterten Funktionenräume gegeben. Außerdem wird eine Erweiterung des Rayleigh-Ritz-Prinzips auf nicht-hermitische Eigenwertprobleme vorgestellt, und die stochastische Variationsmethode auf den nicht-hermitischen Fall erweitert. Dafür stellen wir verschiedene mögliche Auswahlkriterien für Basisfunktionen vor und testen unsere Methoden an einem eindimensionalen Problem, das auch schon in der Literatur untersucht worden ist. Unsere Varianten der stochastischen Variationsmethode können die Resonanzpositionen erfolgreich und effizienter als andere Methoden zur Behandlung von nicht-hermitischen Problemen berechnen.

Contents

1 Introduction

In the vast majority of textbooks and academic literature, quantum mechanics is formulated within a mathematical framework that is based on Hilbert spaces for quantum states and operators for observables. These operators are assumed to be *Hermitian*, which ensures that eigenvalues are real. This is considered a very fundamental ingredient to the mathematical description of quantum mechanics as it is conventionally presented (see, for instance, [1]). Most practical methods for solving quantum-mechanical problems are adapted to such a framework. This is particularly true for variational techniques treating bound states. They are based on the fundamental Rayleigh-Ritz theorem, which states that the ground-state energy of a system can be approximated increasingly better if the energy expectation value is minimised over an ever larger subspace of the full Hilbert space. This can be done very well in practice and yields good results. It is a method that is also applicable for high-dimensional problems, when discretisation methods break down because of their exponentially rising computational costs. In particular, the *stochastic variational method*, as it is described, for instance, in [2], is very flexible and well-suited for a wide range of problems, especially because it allows to handle *non-linear* variational parameters.

Quantum resonances are striking phenomena that appear in a wide range of systems of different scales; from molecules and atoms to nuclear and particle physics. They play important roles in the theoretical as well as experimental treatment of systems occurring there. Resonances can be described mathematically by considering complex poles of the scattering matrix, which correspond to *complex* eigenvalues of the Hamiltonian of a given system. This introduces technical difficulties, however: The corresponding resonance wave functions do not satisfy appropriate decay properties for $r \to \infty$ (in fact, they grow exponentially in this limit) and are thus not square-integrable, which prevents the direct application of most methods used commonly in quantum mechanics (including the variational techniques mentioned above). These methods are based on a Hilbert-space setting and assume eigenfunctions in spaces like $L^2(\mathbb{R}^n)$. The way out is to apply suitable similarity transformations (such as *complex scaling*), what changes the problem of finding the resonance positions into finding eigenvalues of a transformed operator on a Hilbert space. This operator is, however, *non-Hermitian*. The basic ideas of these strategies can be found in [3]. Their description is quite often not mathematically well-founded, though.

Unfortunately, after such transformations have been performed, the variational method is still not directly applicable: It is based on the principle of *minimisation*, but with complex values, this is clearly no longer meaningful. Thus a generalised principle has to be derived. It can be based on *stationarity* instead and can be formulated also in the non-Hermitian setting. The stochastic variational method works by generating candidate basis functions based on randomly chosen values for the non-linear parameters, and then selecting the choice that is best according to the minimisation criterion. This has to be adapted to work with complex values and stationarity instead of minimisation, too. Thus, one has to find a new selection criterion that allows to choose good values for the non-linear parameters in the extended situation. In the following, our principal aim is to find a well-founded mathematical formulation of these methods, and to demonstrate that the stochastic variational method can indeed be generalised and applied to the situation of resonances.

2 Complex Energies in Quantum Theory

In Section 2.1 we give a short overview of the principles and notions of quantum mechanics. Our discussion follows standard text books on this topic, such as, for instance, [4]. In the further sections of this chapter, we will introduce scattering theory and show how *complex values* for energies arise in its mathematical description.

2.1 Conventional Setting

It is commonly known that in quantum mechanics the state of a physical system can be described mathematically by a (normalised) vector in some underlying Hilbert space $\mathcal{H}$, see also [1]. In the so-called *Schrödinger picture*, time evolution is described by assuming that this vector depends on time as a real parameter and that the evolution of the system is governed by the *time-dependent Schrödinger equation*

$$i\hbar \frac{d}{dt}\Psi(t) = \hat{H}\Psi(t). \tag{1}$$

$\hat{H}$ is called the Hamiltonian of the system and can also be understood as the generator of the time-evolution semi-group of (1). At the same time it corresponds to the total energy of the system as an observable, and consequently it is naturally assumed to be a Hermitian operator. Here and in the following, we restrict our considerations for simplicity to time-independent Hamiltonians. The total energy is usually decomposed into a kinetic and an interaction part. For the simplest case of a single particle in an external field, this yields the canonical form of $\hat{H}$ represented as

$$\hat{H} = -\frac{\hbar^2}{2m}\Delta + \hat{V}$$

in position space, where the kinetic-energy part is the Laplacian. If the potential is assumed to be local, $\hat{V}$ becomes a multiplication operator such that

$$\left(\hat{V}\psi\right)(x) = V(x)\psi(x)$$

for every $\psi \in \mathcal{H}$ and a potential *function* V. Thus the Schrödinger equation (1) reads explicitly as

$$i\hbar\frac{\partial}{\partial t}\Psi(t,x) = -\frac{\hbar^2}{2m}\Delta\Psi(t,x) + V(x)\Psi(t,x). \tag{2}$$

By performing the transformations

$$x \mapsto \xi = \frac{x}{a}, \quad t \mapsto \tau = \frac{t}{d}$$

and dividing the whole equation by some ϵ we can make it dimensionless if a has the dimension of a length, d of a time and ϵ of an energy. These transformations result in

$$i\frac{\hbar}{\epsilon d}\frac{\partial}{\partial \tau}\Psi(\tau,\xi) = -\frac{\hbar^2}{2m\epsilon a^2}\Delta\Psi(\tau,\xi) + \tilde{V}(\xi)\Psi(\tau,\xi), \tag{3}$$

where $\tilde{V}(\xi) = \frac{V(x(\xi))}{\epsilon} = \frac{V(a\xi)}{\epsilon}$. In other words, if we find a solution $\Psi(\tau, \xi)$ of (3) then $\Psi(\tau(t), \xi(x))$ also solves (2) and vice-versa. In addition to (3) being dimensionless, the parameters a, d and ϵ can furthermore be chosen in such a way that the constants disappear and the remaining equation (after renaming the variables back) is the *dimensionless Schrödinger equation*

$$i\frac{\partial}{\partial t}\Psi(t, x) = -\frac{1}{2}\Delta\Psi(t, x) + V(x)\Psi(t, x). \tag{4}$$

The Hamiltonian then takes the form

$$\hat{H} = -\frac{1}{2}\Delta + \hat{V}. \tag{5}$$

This convention will be used below to simplify the expressions and to clearly point out the essential structures in contrast to issues related only to a particular system of units.

The equation (4) can be solved formally using the time-evolution operator $e^{-i\hat{H}t}$. According to Stone's theorem (Theorem 3.1 on page 288 of [1]), this operator is unitary if $\hat{H}$ is self-adjoint. Thus, it also preserves the normalisation condition of $\Psi(t)$ over time. However, in order to actually *evaluate* the exponential function of $\hat{H}$, one can use a spectral decomposition of the Hamiltonian. In particular, ansatz functions of the form

$$\Psi(t, x) = e^{-iEt}\psi(x) \tag{6}$$

are solutions to (4) whenever E is in the spectrum of $\hat{H}$ and ψ the corresponding stationary state, i. e.,

$$\hat{H}\psi(x) = E\psi(x). \tag{7}$$

Of course, also *superpositions* of solutions of the form (6) are solutions, since (4) is a linear equation. The relation (7), which is just the eigenvalue equation for $\hat{H}$, is called the *time-independent Schrödinger equation*. By (6) it can be seen that each solution ψ of it corresponds to a stationary solution[1] of (4). If we consider in particular the *discrete spectrum* of $\hat{H}$ consisting of energies E_n and eigenfunctions ψ_n, those states are called *bound states* and very often of significant physical importance. A famous example is that of atomic spectra, where the values of E_n can readily be measured in experiments and the agreement with those experiments finally gave quantum theory its justification.

These bound-state energies E_n are *real quantities* since they are the eigenvalues of a self-adjoint or at least Hermitian operator $\hat{H}$, see also Theorem 5.1 on page 227 of [1]. In contrast to this conventional framework of quantum mechanics, we want to outline in the sections below that it is sometimes interesting to consider (7) also for *complex eigenvalues*. This is in particular useful to describe *resonance phenomena*, but the price to pay is that we then have to deal with a Hamiltonian $\hat{H}$ that is no longer Hermitian (it may still be Hermitian on the usual Hilbert spaces such as $L^2(\mathbb{R}^n)$, but eigenfunctions ψ_n corresponding to non-real $E_n \in \mathbb{C}$ are not contained in them in this case). It is even possible to turn bound states with real energies "continuously" into resonance states corresponding to complex eigenvalues of (7) by only changing the depth of the potential of a system, as we will see in Example 2.

---

[1] "Stationary" means in this case that the only time dependence is in the complex phase e^{-iEt}, which stems from the wave nature of quantum mechanics. It does not, however, influence the probability distribution of finding a particle at a particular position.

2.2 Poles of the S-Matrix

A very interesting phenomenon in scattering processes is that of *resonances*. Phenomenologically, sharp peaks appear in the cross section at distinct scattering energies, while the cross section may otherwise change rather slowly and smoothly with the energy. See, for instance, Figure 13.1 in [5] for experimental results showing these peaks or Figure 2 for the results of the simplified model described below. Scattering theory allows a rather nice and simple mathematical description of this situation, of which we want to outline the basic ideas below. A more complete description can be found in chapter 13 and the preceding chapters of [5]. First recall that the partial-wave cross section σ_l can be expressed via the phase shift δ_l by the relation

$$\sigma_l(k) = \frac{4\pi(2l + 1)}{k^2} \sin^2 \delta_l(k). \tag{8}$$

Furthermore, the phase shift itself can be understood as the (inverse) argument of the Jost function

$$\mathcal{A}(k) = \left|\mathcal{A}(k)\right| e^{-i\delta_l(k)}. \tag{9}$$

The partial-wave S-matrix can be expressed via the Jost function by

$$s_l(k) = \frac{\mathcal{A}(-k)}{\mathcal{A}(k)}. \tag{10}$$

Under certain assumptions on the potential (for more details, see section 12-c of [5]), these functions can all be continued analytically for $k \in \mathbb{C}$. Now assume that the Jost function has a zero $\mathcal{A}(\overline{k}) = 0$ with $\mathrm{Im}\left(\overline{k}\right) < 0$, or, equivalently by (10), that the S-matrix has a pole at $\overline{k} = k_r - ik_i$, where $k_r, k_i \in \mathbb{R}$ are the real and (inverse) imaginary parts of $\overline{k}$ and $k_i > 0$. Then by $\mathcal{A}(\overline{k}) = 0$, the first-order Taylor expansion of $\mathcal{A}$ around $\overline{k}$ gives the approximation

$$\mathcal{A}(k) \approx \frac{d\mathcal{A}(\overline{k})}{dk} \cdot (k - \overline{k}).$$

With (9) this yields

$$\delta_l(k) \approx -\arg\left(\frac{d\mathcal{A}(\overline{k})}{dk}\right) - \arg(k - \overline{k}) = \delta_{\mathrm{bg}} + \delta_{\mathrm{res}}(k). \tag{11}$$

The constant offset δ_{bg} is called the *background phase shift*, and $\delta_{\mathrm{res}}(k)$ the *resonant part* of δ_l. The latter encapsulates the (approximate) k-dependence of δ_l around $\overline{k}$, which is thus given by $-\arg(k - \overline{k})$. This rather simple expression corresponds to the angle between the line connecting $\overline{k}$ to k and the real axis (see Figure 1). In particular, if $\overline{k}$ lies close to the real axis, it is implied that $\delta_l(k)$ changes very rapidly when k moves past $\mathrm{Re}\left(\overline{k}\right)$ along the real axis, and this leads to a sharp peak in the cross section according to (8). The expected cross section following this approximation can be seen for different resonance positions $\overline{k}$ and the case of $\delta_{\mathrm{bg}} = 0$ in Figure 2. The "asymmetry" with respect to left and right of the centre at $k = 1$ is due to the factor $\frac{1}{k^2}$ in (8), but it can be seen nicely that this becomes unimportant and a sharp peak develops if $\overline{k}$ lies close to the real axis. Other possible cases for δ_{bg} are shown schematically in Figure 13.3 in [5], and
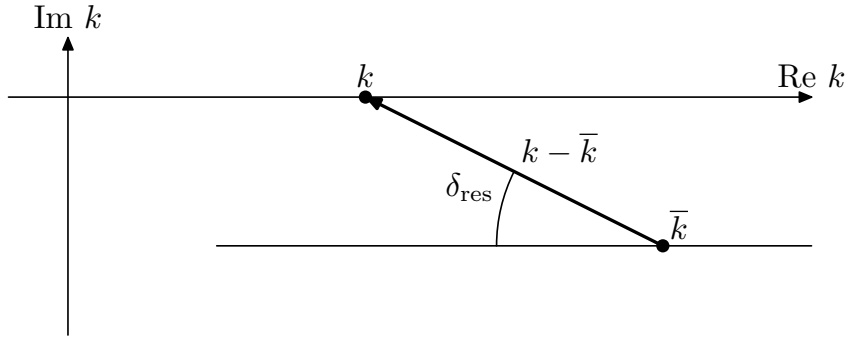
Figure 1: The geometric situation for $\delta_{\mathrm{res}}(k)$ according to the approximation (11) in the complex k-plane. See also Figure 13.2 in [5].

also the example pictured in Figure 13.7 there seems worthwhile to mention here. To conclude this section, we want to add a few further remarks without in-depth discussion, since they are not important for our further goals:

- Of course, using the relation

$$E = \frac{k^2}{2}, \tag{12}$$

  the cross section can also be expressed as a function of E. Following this, (11) can be reformulated with some trigonometry in view of the geometric situation depicted in Figure 1. This way, one can derive the famous *Breit-Wigner formula*:

$$\sigma_l(E) \sim \sin^2 \delta_l(E) = \frac{\left(\frac{\Gamma}{2}\right)^2}{(E - E_r)^2 + \left(\frac{\Gamma}{2}\right)^2}$$

  Here, according to the usual convention, the notation $\overline{E} = E_r - i\frac{\Gamma}{2}$ is used for the position of the resonance pole in the complex E-plane. See also (13.4) in [5] and the surrounding text.

- Resonances can also be interpreted as meta-stable states that can temporarily "capture" scattering projectiles until they decay again after some time. The phase shift and consequently the effect on the cross section can then be explained in terms of a *time-delay*, see, for instance, the sections 13-c and 13-d of [5]. Moiseyev calls these phenomena "shape-type resonances" and points out in section 2.1 of [3] that the decay can also be interpreted as the captured particle tunnelling through the capturing potential after some time.

2.3   Complex Eigenvalues

Above in Section 2.2 we have discussed why it is interesting to find poles of the S-matrix or, equivalently, zeros of the Jost function f. In order to do so, it is convenient to also recall the *regular solution* ϕ_{lk} from scattering theory, which solves the radial Schrödinger equation of a (spherically symmetric) scattering problem for some eigenvalue $E = \frac{k^2}{2}$,
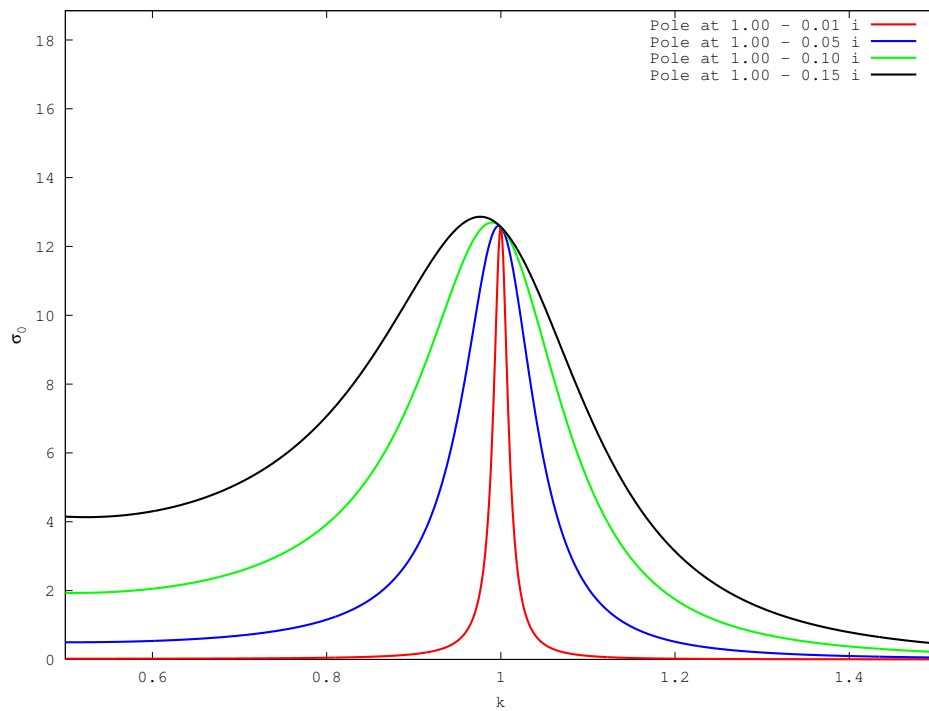
Figure 2: Plot of the expected cross section in the $l = 0$ partial wave according to the approximation (11) for resonance poles of different distances to the real axis for the case of $\delta_{\mathrm{bg}} = 0$.

and is among all solutions uniquely characterised by requiring its asymptotic behaviour for $r \to \infty$ to match

$$\phi_{lk}(r) \to \frac{i}{2} \left(\mathcal{A}(k) \hat{h}_l^-(kr) - \mathcal{A}(k)^* \hat{h}_l^+(kr) \right). \tag{13}$$

Here, $\hat{h}_l^\pm$ denote the Riccati-Hankel functions. Note that for $k \in \mathbb{R}$, ϕ_{lk} is real. The regular solution is discussed in section 11-f of [5], and the relation (13) can be found in (11.48) there. Under certain regularity assumptions on the potential, one can again perform analytic continuation and let $k \in \mathbb{C}$. Of course, this also results in the eigenvalue $E = \frac{k^2}{2}$ being *complex*. Since we are interested in momenta such that $\mathcal{A}(k) = 0$, i. e., where a resonance pole of s_l occurs, in light of (13) we have to look for k such that

$$\phi_{lk}(r) = C \hat{h}_l^+(kr) \to C e^{ikr} \tag{14}$$

as $r \to \infty$, where $C \in \mathbb{C}$ is a generic, complex constant. In other words, we are looking for *eigenfunctions* of the Schrödinger equation that satisfy *outgoing boundary conditions* according to (14). If we have one, the corresponding complex eigenvalue E (or its associated momentum k) provides us with a resonance pole; on the other hand, if we have a resonance pole with $\mathcal{A}(k) = 0$, we know that the corresponding regular solution is an eigenfunction satisfying outgoing boundary conditions. Thus, it is interesting to consider not only real but also *complex eigenvalues* of the Hamiltonian (5), in particular with outgoing boundary conditions. This relation between eigenvalues and poles of the S-matrix is also discussed extensively in section 12-d of [5] for bound states and in chapter 4 of [3] for the general case of complex values.

Now, when we have discussed *complex* eigenvalues, one may wonder how an operator like $\hat{H}$ in (5), which by common knowledge is *Hermitian*, can actually have non-real eigenvalues. The answer to this question lies in the fact that *also the underlying space is "part" of the operator* and its Hermiticity. In other words, $\hat{H}$ of the form (5) may be Hermitian (it is for a real potential) for $\psi \in L^2(\mathbb{R}^n)$, but it may still also possess eigenfunctions for complex eigenvalues as long as they are from another space! And indeed, the resonance wave functions we are interested in are not integrable over the whole of $\mathbb{R}^n$ because they do not have the necessary fall-off: If $k \in \mathbb{C}$ is a resonance pole with $k = k_r - ik_i$, where $k_r, k_i \in \mathbb{R}$ and $k_i > 0$ as before, then the asymptotic behaviour of the regular solution is, according to (14),

$$\phi_{lk}(r) \to C e^{ikr} = C e^{ik_r r} \cdot e^{k_i r} \to \infty \tag{15}$$

as $r \to \infty$. Thus a negative imaginary part of k makes the function explode exponentially rather than decay far away from the scattering centre, which certainly means that in this case $\phi_{lk} \notin L^2(\mathbb{R}^n)$. This creates problems when we actually want to calculate the eigenvalues (which will be discussed in detail in Chapter 3 below).

As a simple demonstration, consider the Hamiltonian (5) in one dimension and with a real potential, with $L^2(\mathbb{R})$ being the underlying Hilbert space. Then $\hat{V}$ is trivially Hermitian on that space because of our requirement that $V(x)$ must be real for every $x \in \mathbb{R}$. The kinetic-energy part can be handled with integration by parts,[2] which gives

---

[2] See also (4.3) in [3] for basically the same example calculation.

for some interval $I = (a, b) \subset \mathbb{R}$:

$$
\begin{aligned}
\left\langle \phi \middle| -\frac{1}{2}\Delta\psi \right\rangle_{L^2(I)} &= -\frac{1}{2} \int_I \phi(x)^* \frac{d^2}{dx^2} \psi(x) \, dx \\
&= -\frac{1}{2} \int_I \left(\frac{d^2}{dx^2} \phi(x) \right)^* \psi(x) \, dx - \frac{1}{2} \left. \phi(x)^* \frac{d}{dx} \psi(x) \right|_a^b + \frac{1}{2} \left. \left(\frac{d}{dx} \phi(x) \right)^* \psi(x) \right|_a^b \\
&= \left\langle -\frac{1}{2}\Delta\phi \middle| \psi \right\rangle_{L^2(I)} + \frac{1}{2} \left. \left(\left(\frac{d}{dx}\phi(x) \right)^* \psi(x) - \phi(x)^* \frac{d}{dx}\psi(x) \right) \right|_a^b
\end{aligned}
$$

Thus the operator is Hermitian *only if* we take also the right boundary conditions into account, such that they make the additional terms vanish! For the space $L^2(\mathbb{R})$, we have that because there integrability implies automatically that $\phi(x), \psi(x) \to 0$ for $x \to \pm\infty$.

2.4   Feshbach Reduction and Optical Potentials

Another approach to resonances and the need for non-Hermitian operators is the *Feshbach reduction*, which leads to genuinely *complex* (and thus non-Hermitian) effective "potentials" termed *optical potentials*. This procedure was invented by [6], and in the context of non-Hermitian quantum mechanics it is also described in section 2.2 of [3]. For an application that explicitly highlights the complex nature of the optical potential, see, for instance, [7]. Below we describe the basic idea behind this framework, although our work in the following will mainly be based on the interpretation of resonances as given already in the sections above.

For outlining the Feshbach reduction, we consider a problem with *multiple coupled channels*. For simplicity, let us assume that we only have two channels, and that we thus consider a product space made up of two subspaces corresponding to each one of these channels. In this situation, the total state can be described by two vectors, one from each channel, and we will use the notation $\Psi = (\psi_1, \psi_2)^\top$. Then the Schrödinger equation for this system can be written as

$$
\begin{pmatrix} \hat{H}_1 & \hat{V}_{12} \\ \hat{V}_{21} & \hat{H}_2 \end{pmatrix} \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} = E \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}. \tag{16}
$$

Here, the $\hat{H}_i$ denote the free Hamiltonians in the subspaces, and $\hat{V}_{12}$ as well as $\hat{V}_{21}$ denote the interactions between the channels. Note that $\hat{V}_{12} = \hat{V}_{21}^\dagger$ because the compound Hamiltonian on the left-hand side of (16) should be Hermitian. If we are in fact only interested in ψ_1 but do not want to completely neglect the second channel (because it also influences the first one), we can use the second line in (16) to eliminate ψ_2 from the first line. The equation that defines ψ_2 is

$$
(\hat{H}_2 - E)\psi_2 = -\hat{V}_{21}\psi_1.
$$

The operator $\hat{H}_2 - E$ is not invertible for $E \in \mathbb{R}$ above the threshold since the continuous spectrum of $\hat{H}_2$ lies there. Nevertheless, we can formally solve the equation for ψ_2, if we introduce an infinitesimal offset $i0$ to specify the correct side of the branch cut. This trick yields

$$
\psi_2 = (E - \hat{H}_2 + i0)^{-1} \hat{V}_{21}\psi_1,
$$

thus reducing (16) to

$$\left(\hat{H}_1 + \hat{V}_{21}^\dagger(E - \hat{H}_2 + i0)^{-1}\hat{V}_{21}\right)\psi_1 = E\psi_1. \tag{17}$$

The "effective potential" appearing here is the *optical potential*

$$\hat{V}_{\text{opt}}(E) = \hat{V}_{21}^\dagger(E - \hat{H}_2 + i0)^{-1}\hat{V}_{21}.$$

It depends on E, thus (17) must be solved *self-consistently* and is not an ordinary eigenvalue equation. $\hat{V}_{\text{opt}}$ is complex and generally also non-local. It gives rise to an effective $\hat{H}$ for the first channel only, which is then non-Hermitian. Compare the above also to (3) in [7] or page 364 of [6].

2.5 Examples for Resonances

Example 1. A very famous example in quantum mechanics is the square-well potential of finite depth. It is less well-known that it represents already an example that gives rise to an infinite number of resonances with complex energies in addition to the finite number of bound states. Here, we demonstrate the behaviour of resonances in the one-dimensional case, i. e., for the potential

$$V(x) = \begin{cases} 0 & |x| > a \\ -V_0 & |x| < a \end{cases}, \tag{18}$$

where $x \in \mathbb{R}$. $V_0, a > 0$ are parameters describing the potential's depth and its width, respectively. This potential is shown in Figure 3. A treatment of the bound states for this problem can be found in exercise 14 of [8], while the resonances are explained, for instance, in section 4.3 of [3]. Following the classic solution method for the eigenvalue problem of this system, we can first construct the general solution for each of the three parts of the domain where V is constant. This yields

$$\psi(x) = \begin{cases} Ce^{ik_0x} + C'e^{-ik_0x} & x < -a \\ Ae^{ikx} + Be^{-ikx} & x \in (-a, a) \\ De^{ik_0x} + D'e^{-ik_0x} & x > a \end{cases}, \tag{19}$$

where $k_0^2 = 2E$ and $k^2 = 2(E + V_0)$, which follows (12). As mentioned on page 10, we want to apply outgoing boundary conditions. Thus we require $C = D' = 0$ to achieve the asymptotic behaviour according to (14). The remaining unknown amplitudes A, B, C' and D must be chosen to achieve continuity[3] of ψ and ψ'. Thus the well-known conditions

$$\begin{array}{rclcrcl} \psi(-a^-) & = & \psi(-a^+), & & \psi(a^-) & = & \psi(a^+), \\ \psi'(-a^-) & = & \psi'(-a^+), & & \psi'(a^-) & = & \psi'(a^+) \end{array}$$

must be satisfied. We adhere to the usual notation of

$$\psi(a^\pm) = \psi(a \pm 0) = \lim_{x \to a^\pm} \psi(x)$$

---

[3]An alternative explanation is that—as will be explained in the proof of Theorem 5 below—we need $\psi \in H^2(\mathbb{R})$ according to elliptic regularity when the potential is in $L^2(\mathbb{R})$. Here, V is in particular discontinuous and thus also no higher regularity can be expected.
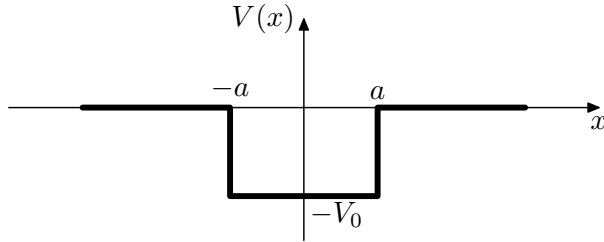
Figure 3: The square-well potential (18) used in Example 1.

to denote the one-sided limits of a function at a (possible) jump position. In our case of (19), we can write these conditions as the linear system

$$
\begin{pmatrix}
-e^{-ika} & -e^{ika} & e^{ik_0a} & 0 \\
-e^{ika} & -e^{-ika} & 0 & e^{ik_0a} \\
-ke^{-ika} & ke^{ika} & -k_0e^{ik_0a} & 0 \\
-ke^{ika} & ke^{-ika} & 0 & k_0e^{ik_0a}
\end{pmatrix}
\begin{pmatrix}
A \\
B \\
C' \\
D
\end{pmatrix}
=
\begin{pmatrix}
0 \\
0 \\
0 \\
0
\end{pmatrix}.
\tag{20}
$$

Since a valid eigenfunction ψ must be normalisable, we need to find a *non-trivial* solution to the homogeneous equation (20). Thus, some $E \in \mathbb{C}$ is an eigenvalue if and only if the matrix on the left-hand side of (20) is singular for the values of k and k_0 that correspond to this E. This is in turn equivalent to requiring that its determinant vanishes. Thereby we get a characterisation of the eigenenergies, and it is possible to locate them numerically simply by trying out possibilities for E on some grid and checking for zeros of the matrix determinant.[4]

For the parameters chosen as $V_0 = 10$ and $a = 5$ in (18), the eigenvalues satisfying the singularity condition are shown in Figure 4. Compare especially the dots corresponding to resonances (those with $\mathrm{Re}\,(E) > 0$ and $\mathrm{Im}\,(E) < 0$ in the right half of the figure) to Figure 4.2 in [3]. For these parameters, there exist 15 bound states and the resonance states start at $n = 16$. Note that Moiseyev uses a potential shifted by a to the right with respect to (18), but this does not affect the eigenvalues. The wave functions for selected eigenvalues of bound states and the first resonance state are shown in Figure 5. It can be seen that as we continue to higher bound-state levels, the number of oscillations in the wave function increases such that there are always exactly n maxima. This is a common pattern in quantum mechanical problems, which can, for instance, also be observed with the square-well potential of infinite depth or the harmonic oscillator. When we cross the imaginary axis to the resonance states, this pattern continues (when the function's infinite rises at the left and right sides are also counted as two maxima). This is a nice indication that resonances and bound states are related, see also Example 2 below. However, the difference between both cases is that the wave function suddenly no longer vanishes for $|x| \to \infty$. Instead, it grows exponentially as it should according to the argument on page 10.

Example 2. Continuing with Example 1, we also show here how a bound state changes into a resonance when the potential is made shallower (in other words, when V_0 is decreased). See also section 13-b of [5] for a theoretical treatment about this behaviour in

---

[4]More elegant conditions for bound-state and resonance energies can be found in (14.8) of [8] and in (4.27) of [3], respectively.
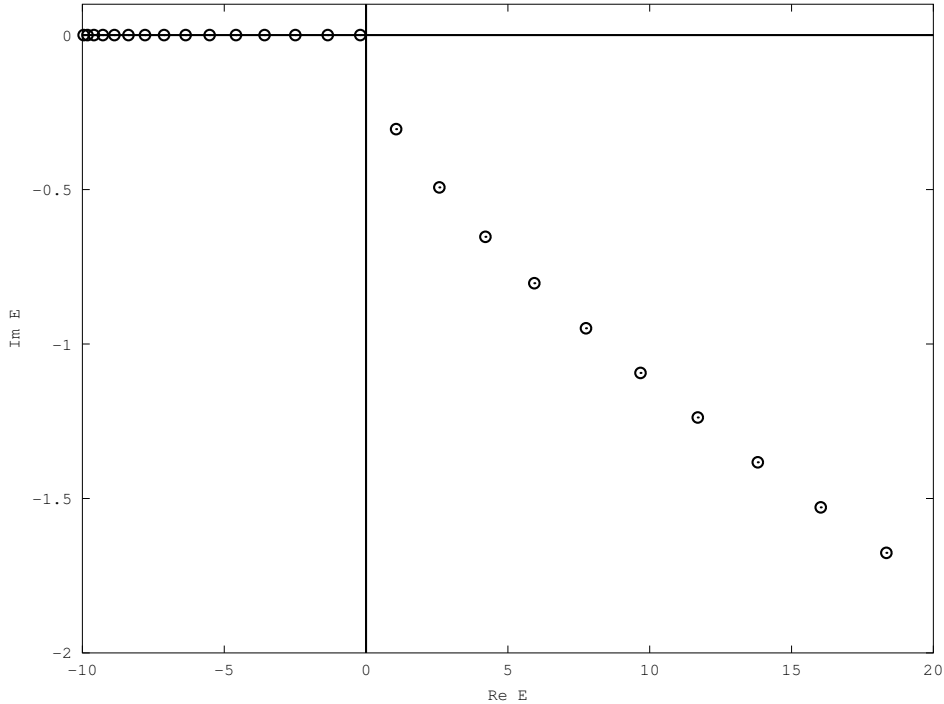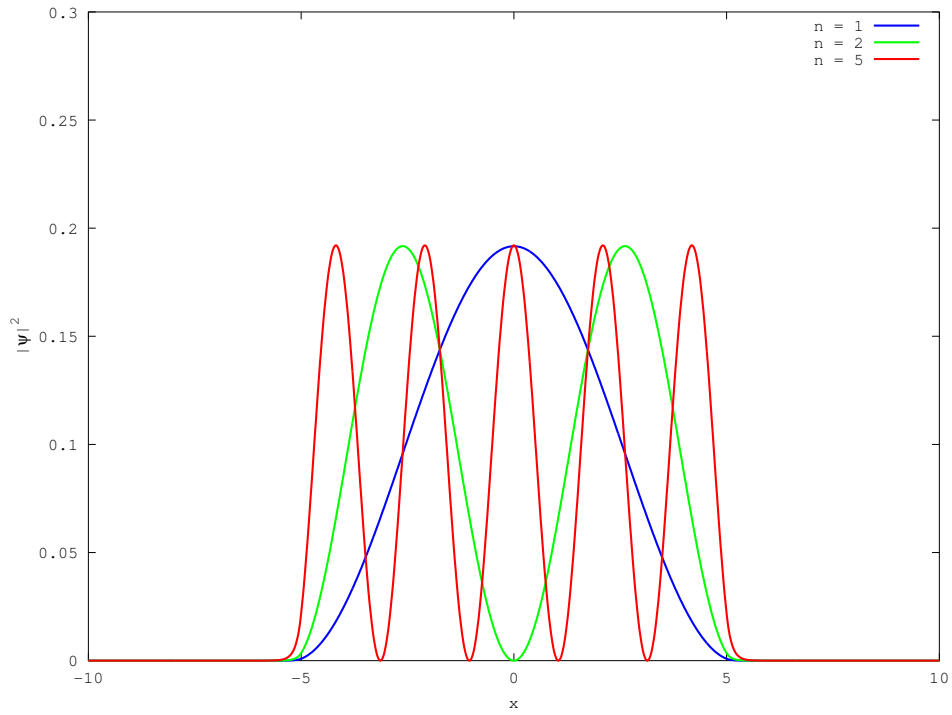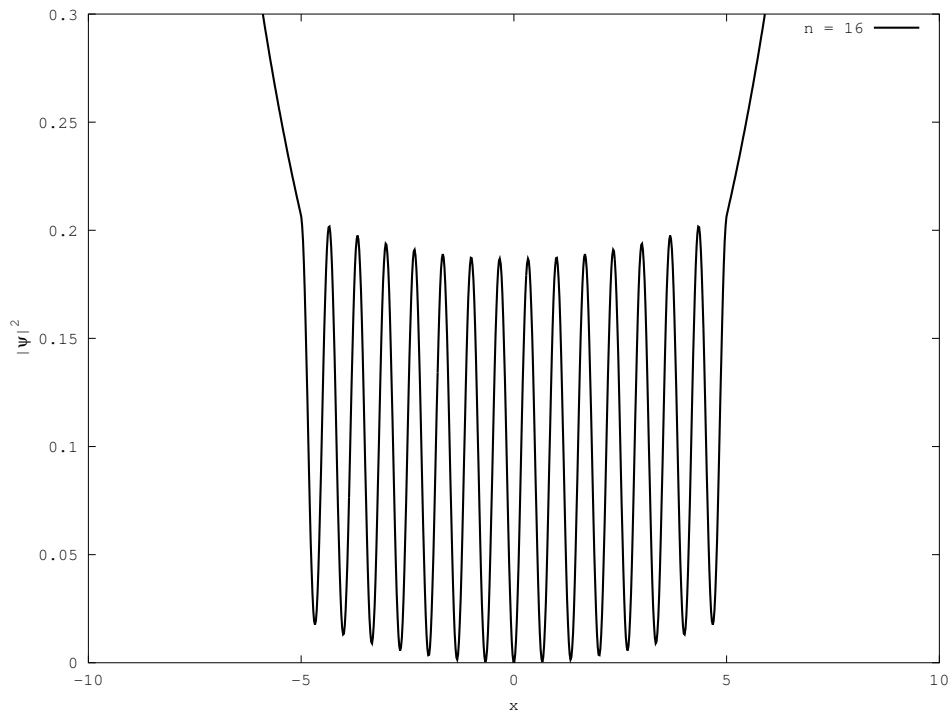
Figure 4: Eigenvalues for the square-well potential of Example 1 with $V_0 = 10$ and $a = 5$.

general. For the numerical demonstration, we use again the square-well potential (18) with $a = 5$, and let V_0 vary from 1 down to 0.1 while tracking the movement of the $n = 5$ eigenvalue. At $V_0 = 1$, it is a bound state with $E_5 \approx$ -0.414, which then turns into a virtual state somewhere above $V_0 \approx 0.4$. The eigenvalue depends continuously on V_0, and thus one can "follow" its movement by changing V_0 at each step only so little that we find precisely one (changed) eigenvalue in some small search window around the old position. If the step in V_0 as well as the size of the search window are controlled correctly, this allows to precisely trace the movement of our selected eigenvalue with arbitrary precision. The resulting path is shown in Figure 6 in both the k- and E-planes, and two critical events along the path have been marked with circles: The red circle is at the origin, which means that at this position the bound state turns into a virtual state. At this moment, the pole changes from the physical to the unphysical sheet of the E-plane. The blue circle marks the point when it leaves the negative imaginary k-axis (or, equivalently, the real E-axis). It may also be interesting to note that when the speed of movement in the E-plane is measured, those two points correspond to minimal and maximal speed, respectively. One may also want to compare this figure to Figure 13.5 in [5], which depicts the expected behaviour during the transition from a bound to a virtual state for an s-wave. This also applies here for our one-dimensional problem. Note that the analysis in [5] is done based on a series expansion around the transition point, which does not allow definite statements about the further development of the eigenvalue beyond the region of convergence of this series. Our result that the eigenvalue eventually leaves the negative imaginary k-axis (as shown in Figure 6) is thus not in contradiction to [5].

14

(a) Bound states.

(b) The first resonance state.

Figure 5: Wave functions for Example 1 corresponding to selected eigenvalues.

15
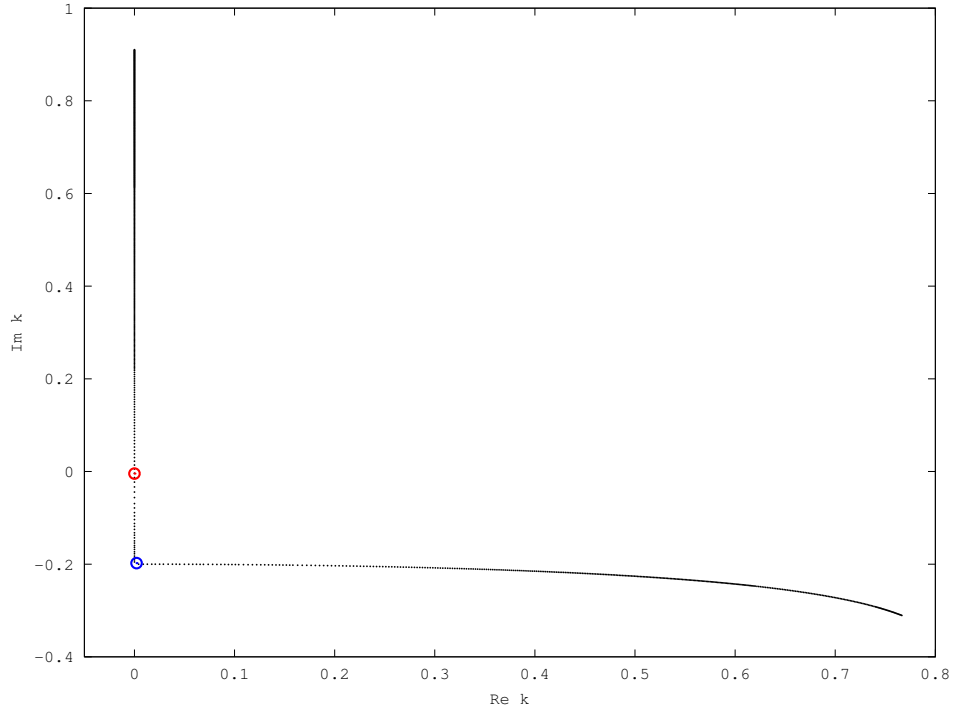
(a) Path in the k-plane.
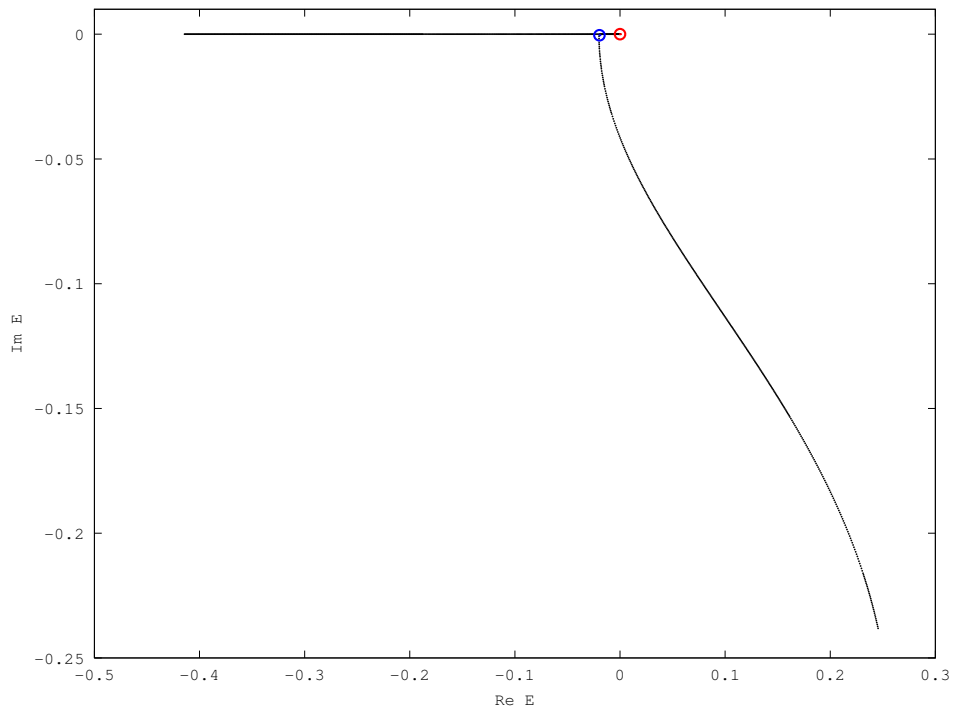
(b) Path in the E-plane.

Figure 6: Movement of the $n = 5$ eigenvalue for the square-well potential (18) when the depth V_0 is changed as explained in Example 2. The same path is shown both in the k- and E-planes, and the positions where the bound state becomes a virtual state as well as when it leaves the imaginary k-axis have been marked.

Example 3. As another example of a real potential developing resonance states (complex poles), consider

$$V(x) = \left(\frac{1}{2}x^2 - J\right)e^{-\lambda x^2}. \tag{21}$$

The parameters are chosen as $J = 0.8$ and $\lambda = 0.1$. This particular potential is well-suited because of multiple reasons: First, V as given in (21) is a nice and smooth (in fact, analytic) function. That allows us to apply complex scaling to it (see Section 3.3 below). Second, this particular problem has already been studied before and independent numerical results for the eigenvalues are available, for instance in [9].[5]

The potential admits a single bound state, and Figure 7a shows how the potential looks like with the bound-state energy level marked by the red line. In Figure 7b, the eigenvalues are shown together with the results of [9], available for resonances with odd index n. In general, good agreement is found in the comparison of our results to the ones of [9]; only for the state with $n = 17$ there is a noticeable deviation, whose cause is unclear. The results here were found with a finite-difference discretisation of the Hamiltonian together with complex scaling, similarly to Example 5 below.

---

[5]Note, however, that in this paper the potential has an additional offset J added to it in comparison to (21). This results in a corresponding shift of the eigenvalues, which, of course, has to be taken into account when comparing the results.
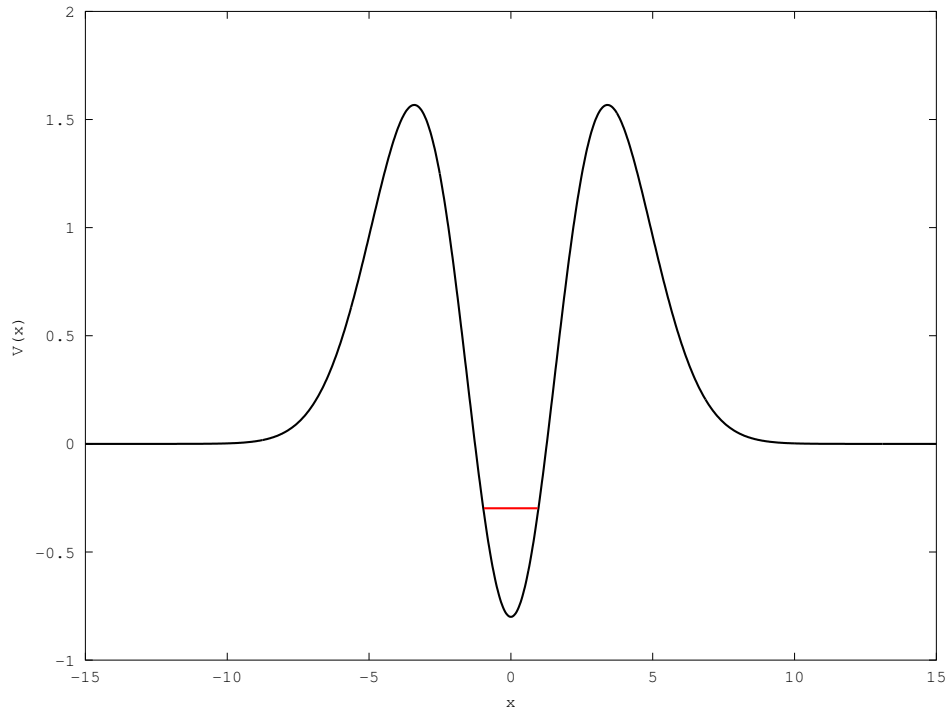
(a) The potential (21) developing a single bound state.
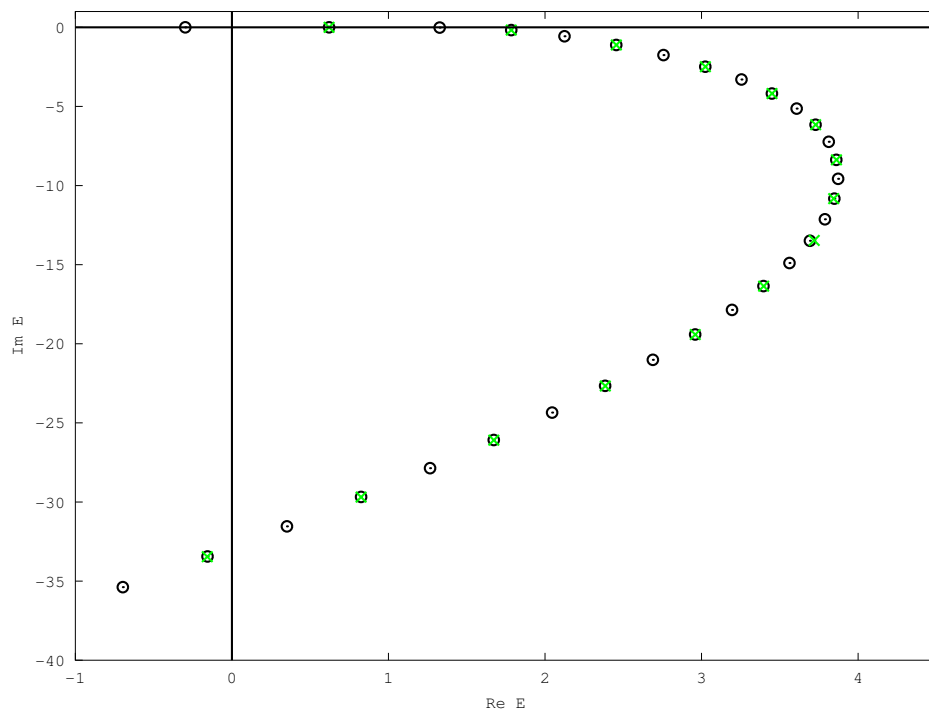
(b) Our eigenvalue results together with those of [9] (green crosses).

Figure 7: The potential (21) on top and the corresponding eigenvalues below. Green crosses mark the eigenvalues given in [9], which are available only for resonances with odd index.

3   Similarity Transformations of the Eigenfunctions

As mentioned already in Section 2.1 above, the usual setting in Hermitian quantum mechanics is to consider the eigenvalue problem of the Hamiltonian operator $\hat{H}$ on some underlying Hilbert space. In particular, the Hilbert space is usually based on $L^2(\mathbb{R}^n)$, so that the Laplacian and thus also $\hat{H}$ are *closed* but not bounded. In contrast, in the following we consider the setting as

$$\hat{H} : H^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n),$$

which is now, due to a restricted domain, a *bounded* linear operator *between* the two Hilbert spaces. It holds that $H^2(\mathbb{R}^n) \subset L^2(\mathbb{R}^n)$ with compact embedding, but not equality between the spaces. Nevertheless this is enough to define the eigenvalue problem in a meaningful way. In fact, if we consider Theorem 5 below, it will become clear that this is no restriction at all for solving the eigenvalue problem, since eigenfunctions in $L^2(\mathbb{R}^n)$ must automatically lie also in all higher Sobolev spaces (including, in particular, $H^2(\mathbb{R}^n)$). The notation H^2 and in general H^s is used in accordance to the literature about PDEs to denote the Sobolev spaces of (equivalence classes of) functions with square-integrable weak derivatives up to order s (see, for instance, chapter 5 of [10]).[6] Where not explicitly specified, norms and inner products are always meant in the space $L^2(\mathbb{R}^n)$ in the following.

In the case of resonance eigenfunctions, however, as seen above in Section 2.3, it can happen that the eigenfunctions we are interested in do not lie in such a Hilbert space. They may be smooth enough, but not integrable over the whole space if they lack proper decay properties. Nevertheless, it is clear that an operator like $\hat{H}$ of (5) (with a local potential represented by a multiplication operator $\hat{V}$) can be defined for functions that are in a space like $H^2(K)$, even if K is not the full coordinate space $\mathbb{R}^n$ but only a compact subset of it. This is possible because decay properties do not play a role for those Hamiltonians. In fact, if one considers an operator that is Hermitian on a Hilbert space but also has complex resonance eigenvalues, it is clear that the corresponding eigenfunctions *must be outside* of the Hilbert space. Unfortunately, this situation introduces a technical difficulty: We no longer operate on Euclidean, normed or even metric spaces, but have to content ourselves with only a topological vector space. In this chapter, we will discuss techniques that can be used to transform the situation back to an operator on a Hilbert space, such that also its eigenfunctions corresponding to complex resonance eigenvalues are within the new Hilbert space. Of course, in order not to violate the "no-free-lunch" theorem (as well as the obvious property that Hermitian operators on a Hilbert space have only real eigenvalues), one has to pay a price; namely, the transformed operator will no longer be Hermitian. This is the case either because it loses its symmetry or because it becomes complex. A general discussion of these transformation techniques can also be found in chapter 5 of [3].

---

[6]Commonly, the index s is used for Sobolev spaces of fractional orders, while k is used for integer orders. We only need $s \in \mathbb{N}$, but use s nevertheless instead of k because k is used to denote the momentum.

3.1   Similarity Transformations

The general situation considered is always this: Let $U_{\text{loc}}, W_{\text{loc}}$ be topological vector spaces, and U, W be Hilbert spaces. We are interested in a linear operator $\hat{H} : U_{\text{loc}} \to W_{\text{loc}}$. Since in particular we want to find eigenvalues, assume that $U_{\text{loc}} \subset W_{\text{loc}}$ as well as $U \subset W$. Then we want to transform $\hat{H}$ to $\hat{H}' : U \to W$, such that the eigenvalues remain the same for $\hat{H}$ and $\hat{H}'$, and can be calculated on the *Hilbert-space setting* of U and W instead of the topological spaces U_{loc} and W_{loc}. The most common scenario is $U = H^2(\mathbb{R}^n)$, $W = L^2(\mathbb{R}^n)$, $U_{\text{loc}} = H^2_{\text{loc}}(\mathbb{R}^n)$ and $W_{\text{loc}} = L^2_{\text{loc}}(\mathbb{R}^n)$, but also more general choices may be useful in practice. Next we develop our strategy more precisely and in detail, partially following [3].

Definition 1. Let $\hat{S} : W_{\text{loc}} \to W$ be linear and bijective with the inverse $\hat{S}^{-1}$. Let furthermore U_{loc} be "invariant" in the sense that $\hat{S}(U_{\text{loc}}) = U$ and $\hat{S}^{-1}(U) = U_{\text{loc}}$. Then $\hat{S}$ is called a *similarity transformation*[7] and we write

$$\psi' = \hat{S}\psi, \quad \hat{H}' = \hat{S}\hat{H}\hat{S}^{-1}. \tag{22}$$

Here, for the transformed vector and operator we have clearly $\psi' \in U$ for $\psi \in U_{\text{loc}}$ as well as $\hat{H}' : U \to W$.

As a fundamental property of such a similarity transformation (also known from the matrix case in linear algebra) on has:

Theorem 1. *Let $\hat{S}$ be a similarity transformation according to Definition 1. Then the sets of eigenvalues of $\hat{H}$ and $\hat{H}'$ are identical, with the eigenvectors to a common eigenvalue transformed to each other via $\hat{S}$ and $\hat{S}^{-1}$. In particular, eigenvectors of $\hat{H}'$ lie all in the Hilbert space U.*

Proof. Let $\lambda \in \mathbb{C}$ be an eigenvalue of $\hat{H}$. Then there exists $\psi \in U_{\text{loc}}$ with $\hat{H}\psi = \lambda\psi$, and consequently with $\psi' = \hat{S}\psi$ we have:

$$\hat{H}'\psi' = \hat{S}\hat{H}\hat{S}^{-1} \cdot \hat{S}\psi = \hat{S}\hat{H}\psi = \lambda\hat{S}\psi = \lambda\psi'.$$

Thus λ is also an eigenvalue of $\hat{H}'$ with eigenvector ψ'. The reverse direction also holds by the same argument using $\hat{S}^{-1}$. $\qquad\square$

3.2   The Zel'dovich Transformation

As we have seen in (15), a common cause for non-integrability of resonance eigenfunctions is exponential divergence like e^{kr} for $r \to \infty$. The so-called *Zel'dovich transformation* (see, for instance, section 5.1 of [3]) is a direct answer for functions that diverge like that. Since far away from the centre, where the potential is influential, not much relevant information is contained in the wave function anyway, it can be damped out without noticeable effects. The Zel'dovich transformation is thus simply a *multiplication operator* with a strongly decaying weight.

---

[7]This is similar in spirit to the notion of similarity transformations of matrices on finite-dimensional vector spaces, but we are not aware of a widely used generalisation of this concept for infinite-dimensional spaces. Thus Definition 1 should be understood as the particular meaning we want to give to this term for the present work only.

Definition 2. Fix some real $\epsilon > 0$. We define the Zel'dovich transformation as

$$\psi' = \hat{S}_\epsilon \psi = e^{-\epsilon r^2} \psi.$$

Thus in the spatial representation of a wave function, this is indeed just multiplication with a smooth, bounded and decaying weight. Furthermore, let us also introduce *Zel'dovich spaces*:

$$H_\epsilon^s(\mathbb{R}^n) = \left\{ \psi : \mathbb{R}^n \to \mathbb{C} \mid e^{-\epsilon r^2} \psi \in H^s(\mathbb{R}^n) \right\}, \quad s = 0, 1, \ldots.$$

For $L_\epsilon^2(\mathbb{R}^n)$, we can define an inner product as

$$\langle \psi | \phi \rangle_\epsilon = \left\langle e^{-\epsilon r^2} \psi | e^{-\epsilon r^2} \phi \right\rangle, \tag{23}$$

and, of course, analogous definitions are possible for higher-order spaces with $s > 0$.

Note that Definition 2 defines in fact a whole *family* of transformations, depending on the parameter ϵ, which can be varied in applications. The same will also be true later for complex scaling (in Section 3.3). It is easy to verify that this transformation behaves indeed as expected:

Theorem 2. *For $s = 0, 1, \ldots$ and $\epsilon > 0$, $H_\epsilon^s(\mathbb{R}^n)$ with the inner product defined in (23) is a Hilbert space. Furthermore, if $\epsilon_1 > \epsilon_2 > 0$, then*

$$H^s(\mathbb{R}^n) \subset H_{\epsilon_2}^s(\mathbb{R}^n) \subset H_{\epsilon_1}^s(\mathbb{R}^n)$$

with continuous embedding, and also $H_\epsilon^s(\mathbb{R}^n) \subset H_{loc}^s(\mathbb{R}^n)$. In the latter case, the embedding also "respects" the topology in the sense that convergence in $H_\epsilon^s(\mathbb{R}^n)$ implies convergence in $H_{loc}^s(\mathbb{R}^n)$.

Proof. It is enough to consider $s = 0$; the case $s > 0$ follows by the same arguments. It is trivial to see that (23) fulfils the linearity and anti-symmetry required for an inner product. Also,

$$\langle \psi | \psi \rangle_\epsilon = \left\langle e^{-\epsilon r^2} \psi | e^{-\epsilon r^2} \psi \right\rangle \geq 0$$

for every $\psi \in L_\epsilon^2(\mathbb{R}^n)$, with equality holding only if $e^{-\epsilon r^2} \psi = 0$. By the basic properties of the exponential function, the latter equivalently means $\psi = 0$. Hence, $\langle \cdot | \cdot \rangle_\epsilon$ is a valid inner product on $L_\epsilon^2(\mathbb{R}^n)$. Now, if $(\psi_n) \subset L_\epsilon^2(\mathbb{R}^n)$ is a Cauchy sequence, by (23) also $\left(e^{-\epsilon r^2} \psi_n \right)$ is a Cauchy sequence in $L^2(\mathbb{R}^n)$ and thus, since $L^2(\mathbb{R}^n)$ is complete, it is convergent. Call the limit $\psi' \in L^2(\mathbb{R}^n)$, and note that first $\psi = e^{\epsilon r^2} \psi' \in L_\epsilon^2(\mathbb{R}^n)$ and second

$$\|\psi_n - \psi\|_\epsilon^2 = \left\| e^{-\epsilon r^2}(\psi_n - \psi) \right\|^2 = \left\| e^{-\epsilon r^2} \psi_n - \psi' \right\|^2 \to 0,$$

wherefore also $(\psi_n) \subset L_\epsilon^2(\mathbb{R}^n)$ is convergent with limit ψ. This proves that $L_\epsilon^2(\mathbb{R}^n)$ is indeed a Hilbert space.

For the embeddings, note first that we always have

$$0 < e^{-\epsilon_1 r^2} \leq e^{-\epsilon_2 r^2} \leq 1.$$

Thus also for an arbitrary, measurable function ψ, it holds that

$$\|\psi\|_{\epsilon_1} \leq \|\psi\|_{\epsilon_2} \leq \|\psi\|.$$

This implies the continuous embeddings. If we, on the other hand, *fix a compact set* $K \subset \mathbb{R}^n$, then we find a uniform lower bound $0 < m \leq e^{-\epsilon r^2}$ valid on the whole of K. Thus the other way round, also

$$\|\psi\|_{L^2(K)} \leq \frac{1}{m} \|\psi\|_{L^2_\epsilon(K)},$$

which shows that $L^2_\epsilon(\mathbb{R}^n) \subset L^2_{\text{loc}}(\mathbb{R}^n)$ and convergence in the former implies convergence in $L^2(K)$ for each compact K, thus convergence in the topology of $L^2_{\text{loc}}(\mathbb{R}^n)$. $\qquad\square$

It is clear that a function behaving like e^{kr} for $r \to \infty$ belongs to $L^2_\epsilon(\mathbb{R}^n)$ for arbitrarily small $\epsilon > 0$. Therefore, this especially important class of eigenfunctions can indeed be "captured" using the Zel'dovich approach. However, working with the new, exotic spaces and inner products is cumbersome. Instead, one can also use the transformation $\hat{S}_\epsilon$ in the sense of the general framework in Section 3.1 to transform the situation back to the ordinary spaces $H^s(\mathbb{R}^n)$. This is made rigorous by the following result:

Theorem 3. *For every $\epsilon > 0$ and $s = 0, 1, \ldots$, the operator $\hat{S}_\epsilon : H^s_\epsilon(\mathbb{R}^n) \to H^s(\mathbb{R}^n)$ is linear and bijective. It is also bounded with $\left\|\hat{S}_\epsilon\right\| = 1$. In fact,*

$$\langle \phi | \psi \rangle_\epsilon = \left\langle \hat{S}_\epsilon \phi | \hat{S}_\epsilon \psi \right\rangle$$

and thus actually

$$\left\|\hat{S}_\epsilon \psi\right\| = \|\psi\|_\epsilon \ \text{ for all } \psi \in L^2_\epsilon(\mathbb{R}^n).$$

The same property also holds for $s > 0$. Furthermore, $\hat{S}_\epsilon$ fulfils the requirements of Definition 1 for every $\epsilon > 0$ when we set

$$U_{loc} = H^s_\epsilon(\mathbb{R}^n) \subset W_{loc} = L^2_\epsilon(\mathbb{R}^n), \ \ U = H^s(\mathbb{R}^n) \subset W = L^2(\mathbb{R}^n). \tag{24}$$

Proof. It is clear that $\hat{S}_\epsilon$ is linear, and also, that it maps $H^s_\epsilon(\mathbb{R}^n)$ to $H^s(\mathbb{R}^n)$. In fact, this is basically the definition of $H^s_\epsilon(\mathbb{R}^n)$ according to Definition 2. On the other hand, for every $\psi' \in H^s(\mathbb{R}^n)$ we know that

$$\psi = e^{\epsilon r^2} \psi' = \hat{S}_\epsilon^{-1} \psi' \in H^s_\epsilon(\mathbb{R}^n),$$

so that $\hat{S}_\epsilon$ is also bijective. The claim about the inner product and consequently the norm follows by (23), which also implies directly the boundedness of $\hat{S}_\epsilon$. Finally it is also clear that with the spaces of (24), $\hat{S}_\epsilon$ fulfils Definition 1. $\qquad\square$

Note that it is important to choose the parameter ϵ suitably. In theory, every $\epsilon > 0$ is good to transform away exponential divergence of wave functions. However, a large ϵ washes out relatively much information and may thus lead to inaccurate results. On the other hand, a small ϵ does not damp the divergence sufficiently, what may lead to numerical instability. Obviously, for $\epsilon \to 0$, the transformation converges to the identity mapping, which is shown by the following result:

Lemma 1. *Let $s = 0, 1, \ldots$ and $\psi \in H^s_{loc}(\mathbb{R}^n)$. Then $\hat{S}_\epsilon \psi \to \psi$ for $\epsilon \to 0$, where the convergence is in the topology of $H^s_{loc}(\mathbb{R}^n)$.*

Proof. Fix $K \subset \mathbb{R}^n$ compact. Then for the weight we have

$$e^{-\epsilon r^2} \to 1 \text{ for } \epsilon \to 0$$

at arbitrary $x \in K$ pointwise, thus also uniformly on K. This implies that $e^{-\epsilon r^2}\psi \to \psi$ in $H^s(K)$. $\qquad\square$

Finally, let us assume that we indeed have a Hamiltonian of the form (5) with a real potential. Then, as is commonly known, $\hat{H}$ is a Hermitian operator when considering the spaces $H^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$. This is, however, no longer the case if we apply the Zel'dovich transformation in order to extend[8] the eigenvalue problem to cover also eigenfunctions originally only in $H^2_\epsilon(\mathbb{R}^n)$. Let $\psi_\epsilon \in H^2(\mathbb{R}^n)$ be given. Then

$$
\begin{aligned}
(\hat{H}_\epsilon \psi_\epsilon)(x) &= (\hat{S}_\epsilon \hat{H} \hat{S}_\epsilon^{-1} \psi_\epsilon)(x) = -\frac{1}{2} e^{-\epsilon r^2} \Delta \left(e^{\epsilon r^2} \psi_\epsilon(x) \right) + V(x)\psi_\epsilon(x) \\
&= -\frac{1}{2} e^{-\epsilon r^2} \nabla \cdot \left(\left(\nabla \psi_\epsilon(x) + 2\epsilon \psi_\epsilon(x)x \right) e^{\epsilon r^2} \right) + V(x)\psi_\epsilon(x) \\
&= -\frac{1}{2} \left(\nabla \psi_\epsilon(x) \cdot 2\epsilon x + 4\epsilon^2 \psi_\epsilon(x) r^2 + 2\epsilon \nabla \psi_\epsilon(x) \cdot x + 2n\epsilon \psi_\epsilon(x) + \Delta \psi_\epsilon(x) \right) + V(x)\psi_\epsilon(x) \\
&= \hat{H}\psi_\epsilon(x) - 2\epsilon \nabla \psi_\epsilon(x) \cdot x - \left(2\epsilon^2 r^2 + n\epsilon \right) \psi_\epsilon(x).
\end{aligned}
$$

So the Zel'dovich-transformed operator is still real, but *no longer symmetric*. The term in the middle of the right-hand side, containing the plain gradient of ψ_ϵ (which appears because of the Laplacian being applied to the weight term $e^{\epsilon r^2}$ to the right of it), destroys the symmetry. This is the price to be paid in order to allow exponentially divergent eigenfunctions, corresponding to complex resonance eigenvalues.

3.3 The Complex-Scaling Transformation

Another very successful transformation is *complex scaling*, often also called *complex-coordinate method*. It was already introduced in [11], and is extensively discussed in section 5.2 of [3]. A treatment of its mathematical properties can be found in [12]. For simplicity, we will restrict ourselves here to only one dimension, but all can also be extended to multiple coordinates in a straight-forward way. To show the basic idea behind complex scaling, let us consider again the example of a resonance wave function behaving like

$$\psi(x) \sim e^{ikx} \tag{25}$$

for $x \to \infty$, where $k \in \mathbb{C}$ is some complex momentum. As discussed already on page 10, the question of whether or not this diverges is *directly tied to the sign of $\mathrm{Im}(k)$*: Divergence occurs if and only if the imaginary part of k is negative. The goal of complex scaling is now to apply a transformation that makes $\mathrm{Im}(k)$ positive via a *rotation in the complex plane*. This goal can be reached if we replace k by $k' = e^{i\theta}k$, where $\theta > 0$ is a positive

---

[8]Note that the transformed operator $\hat{H}_\epsilon$ is not an operator extension of $\hat{H}$. In fact, $\hat{H}_\epsilon$ is defined only on $H^2(\mathbb{R}^n)$, which is a *smaller* space than the domain of $\hat{H}$ where the original eigenfunctions lie.

angle large enough for some particular k (at least $\theta > |\arg k|$). In light of (25), this transformation can equally be achieved by replacing x with $x' = e^{i\theta}x$; this makes it clear why the transformation is also called "complex-coordinate" method: The spatial coordinate is extended into the complex plane. If the wave function is analytic, this can be done via analytic continuation. The necessary mathematical details will be developed in the following.

Definition 3. For $\theta > 0$, we first define

$$D_\theta = \left\{ e^{i\theta'}x \mid x \in \mathbb{R}, \theta' \in [0, \theta] \right\}, \quad D_{-\theta} = \left\{ e^{i\theta'}x \mid x \in \mathbb{R}, \theta' \in [-\theta, 0] \right\}.$$

D_θ and $D_{-\theta}$ are both sectors in the complex plane. Assume that V can be analytically extended onto both of them, and set

$$U_{\pm\theta} = \{\psi : \mathbb{R} \to \mathbb{C} \mid \mathrm{Re}\,(\psi) \text{ and } \mathrm{Im}\,(\psi) \text{ can be analytically extended onto } D_{\pm\theta}\}.$$

Then we define $\hat{S}_\theta : U_\theta \to U_{-\theta}$ by

$$(\hat{S}_\theta \psi)(x) = \psi'(x) = \tilde{\psi}\left(e^{i\theta}x\right), \tag{26}$$

where $\tilde{\psi}$ is the analytic continuation of ψ (i. e., its real and imaginary parts) onto D_θ. This is the *complex-scaling transformation*.

Theorem 4. S_θ of (26) is a linear and bijective operator with inverse $\hat{S}_\theta^{-1} = \hat{S}_{-\theta}$. If $\psi \in U_\theta$ and $\psi(x) \to \psi_\infty(x)$ for $x \to \infty$ with ψ_∞ like (25), where $\arg k \in (-\theta, \pi - \theta)$, then $\hat{S}_\theta \psi \in H^s([0, \infty))$ for all $s = 0, 1, \dots$. In particular, $\hat{S}_\theta \psi$ decays exponentially for $x \to \infty$. A similar property is, of course, also valid for $x \to -\infty$ with $\psi_{-\infty}(x) \sim e^{-ikx}$.

Proof. Linearity is clear by Definition 3, as is the inverse mapping. For proving the decay properties it is enough to consider $\psi_\infty(x) = e^{ikx}$. Clearly,

$$(\hat{S}_\theta \psi_\infty)(x) = \exp\left(ie^{i\theta}kx\right) = e^{ik'x},$$

where $k' = e^{i\theta}k$ and thus $\mathrm{Im}\,(k') > 0$. From this it is clear that $\hat{S}_\theta \psi_\infty \in H^s([0, \infty))$, since the necessary smoothness follows from $\psi \in U_\theta$ with its requirement about analyticity. $\square$

With Theorem 4, it is easy to see that also $\hat{S}_\theta$ fulfils the basic requirements of Definition 1. For the spaces, similarly to Definition 2, we can take the space of all those $\psi \in U_\theta$ such that $\hat{S}_\theta \psi \in L^2(\mathbb{R})$, as well as the resulting image space. While the image is not strictly a Hilbert space, it can, of course, be embedded (by definition) into $L^2(\mathbb{R})$ and we do not really need surjectivity of the similarity transformation onto a complete space. Still remaining is, however, the issue of analyticity: In order to apply $\hat{S}_\theta$, we need to be able to extend ψ to complex coordinates. Thus it is important to know whether or not the resonance eigenfunctions we are interested in actually fulfil this requirement. At least a partial regularity result in this direction can be deduced by elliptic regularity:

Theorem 5. Let $V \in C^\infty(\mathbb{R})$, $\hat{H} = -\frac{1}{2}\Delta + \hat{V}$ and assume that E is an eigenenergy for some state $\psi \in H^2_{loc}(\mathbb{R})$, i. e., $\hat{H}\psi = E\psi$. Then also $\psi \in C^\infty(\mathbb{R})$.[9]

---

[9]Unfortunately, Theorem 5 does not imply that ψ is even real-analytic, let alone that it can be continued analytically onto some region D_θ. There exists a result about elliptic regularity (see page 21 of [13]) that implies real-analyticity of the solution of (27), *if the right-hand side is real-analytic*. This cannot be used in the argument of the proof of Theorem 5, though.

Proof. Pick $x_0 \in \mathbb{R}$ and an open interval $I = (a, b)$ around x_0. Then $\psi \in L^2(I)$, and we know that it solves

$$-\frac{1}{2}\Delta\psi(x) = (E - V(x))\psi(x) \text{ in } I, \quad \psi(a), \psi(b) \text{ given on the boundary } \partial I \text{ of } I. \quad (27)$$

Since the right-hand side is in $L^2(I)$, elliptic regularity[10] tells us that in fact $\psi \in H^2(I)$. Continuing this with a bootstrap argument implies that $\psi \in H^s(I)$ even for all $s \in \mathbb{N}$ and thus by the Sobolev embedding theorem (Theorem 6 on page 270 of [10]) we find $\psi \in C^\infty(I)$. Alternatively, this result is also directly implied using Theorem 3 on page 316 of [10], which is based on the same arguments as presented above. $\qquad\square$

Similarly to the discussion in the end of Section 3.2, we now show how the transformed operator $\hat{H}_\theta$ looks like (see also to (5.144) in [3]):

$$
\begin{aligned}
(\hat{H}_\theta\psi_\theta)(x) &= (\hat{S}_\theta\hat{H}\hat{S}_\theta^{-1}\psi_\theta)(x) = -\tfrac{1}{2}\hat{S}_\theta\Delta\psi_\theta\left(e^{-i\theta}x\right) + \hat{S}_\theta V(x)\psi_\theta\left(e^{-i\theta}x\right) \\
&= -\tfrac{1}{2}e^{-2i\theta}\Delta\psi_\theta(x) + V\left(e^{i\theta}x\right)\psi_\theta(x)
\end{aligned}
\quad (28)
$$

Thus the kinetic-energy term gains a complex phase factor, and the potential is subjected to a complex rotation; especially for the latter one needs that V is analytic. This is an important restriction, since it rules out, for instance, the square-well potential (18). Contrary to the Zel'dovich transformation, the Hamiltonian now becomes complex. On the other hand, the operator stays "symmetric" in some sense (see Definition 4 with (36) for a more precise statement).

Since the application of the complex-scaling operator $\hat{S}_\theta$ requires analytic continuation and this method is thus not directly applicable to non-analytic potentials, it is interesting to look for generalisations of this idea that work for a wider range of problems. Accepting some additional complications, it is possible to perform complex scaling only further away from the scattering centre. Thereby, the exponential divergence is still avoided, while the transformation can be applied as long as V is analytic *outside of some compact region*. This makes the procedure applicable also to the square-well potential, for instance. These corresponding techniques can be found in sections 5.3 and 5.4 of [3] and are called *(smooth) exterior-scaling* transformations. Also numerically the analytic continuation is difficult, while it is usually not necessary if an analytic potential is given. However, if it is performed "backwards", it can be used to find the *original* wave function from the eigenfunction of $\hat{H}_\theta$. This can be useful in cases where also the eigenfunction and not only the energy is of interest.

In the following, we want to briefly describe two alternative approaches to complex scaling, although neither one worked satisfactory in our test calculations. Nevertheless, they are interesting from a theoretical point of view.

3.3.1 Complex Scaling as a Time Evolution

Using the power-series expansion of some analytic function ψ, it can be seen that $\hat{S}_\theta$ can be formally written as

$$\hat{S}_\theta = \exp\left(i\theta x\frac{\partial}{\partial x}\right). \quad (29)$$

---

[10]See, for instance, section 6.3 of [10] and in particular Theorem 2. While these results are usually used for real-valued functions, we can apply them here to the real and imaginary parts separately.

This can also be found with more detail in (5.14) in [3]. In the form (29), however, $\hat{S}_\theta$ can also be interpreted as the time-evolution semi-group of the degenerate parabolic differential equation

$$\frac{\partial}{\partial t}\phi(x,t) = ix\frac{\partial}{\partial x}\phi(x,t), \quad \phi(x,0) = \psi(x).$$ (30)

If the solution is known, then $\hat{S}_\theta\psi = \phi(\cdot,\theta)$. This can be readily verified by plugging the candidate solution $\phi(x,t) = \psi(e^{it}x)$ into (30). Also the method of characteristics (see section 3.2 of [10]) applied to (30) results in exactly this solution—*if it is known* already that ψ *can* be extended analytically to the points $e^{it}x$. One can go even further and rewrite the complex equation (30) into a system of two *completely real* functions:

$$\frac{\partial}{\partial t}\begin{pmatrix} v \\ w \end{pmatrix} = x\begin{pmatrix} & -1 \\ 1 & \end{pmatrix}\cdot\frac{\partial}{\partial x}\begin{pmatrix} v \\ w \end{pmatrix},$$ (31)

where $\phi(x,t) = v(x,t) + iw(x,t)$ and the same for the initial condition. It is easy to see that (31) is completely equivalent to (30) if this splitting of ϕ into real and imaginary parts is done. Each equation can be propagated numerically in time using some time-stepping method coupled with a spatial discretisation, but it turns out that this is quite instable; it works to transform $\hat{S}_\theta\psi$ back into ψ on a coarse grid and if not too many time steps are performed (although consequently with low accuracy), but the numerical propagation broke down in our calculations for smaller steps. Standard theory for parabolic equations (see section 7.1 of [10]) does not apply here either, because the equations are degenerate and do not contain any elliptic operator. One might try to add a regularising term $\epsilon\Delta\phi$ on the right-hand side of (30) similarly in spirit to the vanishing-viscosity method (see, for instance, page 540 of [10] and in particular (2) there), but we have not tried that. Viscosity solutions in the usual sense[11] also cannot be applied here because they crucially depend on the *ordering* of the real numbers. Thus one is not able to use this notion (to our knowledge) for complex equations or for systems of equations.

3.3.2 Complex Scaling via the Fourier Transform

Another approach to performing the complex-scaling transformation on some function ψ is via the *Fourier transform*: For $\psi \in L^2(\mathbb{R})$, the Fourier transform is given by

$$\widehat{\psi}(\xi) = \int_{\mathbb{R}} \psi(x)e^{-2\pi i\xi x}\, dx.$$

Then, by the Fourier inversion theorem, we can recover ψ from $\widehat{\psi}$ via

$$\psi(x) = \int_{\mathbb{R}} \widehat{\psi}(\xi)e^{2\pi i\xi x}\, d\xi.$$ (32)

This result as well as other useful properties of the Fourier transform are well-known, and can, for instance, be found in section 4.3 of part III of [1]. Note that in the latter formula

---

[11]See, for instance, [14] or [15]. For the standard definition, the notions of *sub-* and *supersolutions* are necessary, which in turn depend on comparisons. Also the *comparison principle* is a major tool, which cannot be applied to complex equations or to systems of equations.

x appears only in the exponential, wherefore we can use (32) to make the substitution $x \mapsto e^{i\theta}x$ in a straight-forward way and arrive at

$$\left(\hat{S}_\theta\psi\right)(x) = \int_{\mathbb{R}} \widehat{\psi}(\xi)e^{2\pi i\xi \cdot e^{i\theta}x}\, d\xi. \tag{33}$$

This method (not specific to complex scaling, though) is also proposed in [16] as a strategy to perform an analytic continuation numerically, although it did not work in our numerical tests for complex scaling (even less so than propagating (31) backwards in time which worked to some extent as described above in Subsection 3.3.1). However, we can, at least, give a sufficient condition for (33) being well-defined, which is based on the remarkable connection between the smoothness of a function and the decay properties of its Fourier transform:

Theorem 6. *Let $*$ denote the convolution operation between two functions, and assume[12] that there exist $\Psi \in L^2(\mathbb{R})$ and $\epsilon > 0$ such that $\psi = \Psi * e^{-\epsilon x^2}$. Then (33) is well-defined for all x and*

$$\left|\left(\hat{S}_\theta\psi\right)(x)\right| \leq C_\epsilon \left\|\Psi\right\| e^{\epsilon(\sin\theta \cdot x)^2}, \tag{34}$$

where C_ϵ is a (generic) constant depending on ϵ but not on Ψ, ψ, θ or x.

Proof. We shall show the estimate (34), since this then implies integrability in (33). Note first that the Fourier transform of the heat kernel $e^{-\epsilon x^2}$ equals

$$\widehat{e^{-\epsilon x^2}} = C_\epsilon e^{-\epsilon'\xi^2}, \text{ where } \epsilon' = \frac{\pi^2}{\epsilon} > 0.$$

Thus by the convolution property of the Fourier transform and the fact that it preserves the L^2-norm, we get

$$\widehat{\psi}(\xi) = C_\epsilon\widehat{\Psi}(\xi)e^{-\epsilon'\xi^2} \quad \Rightarrow \quad \left\|\widehat{\psi}e^{\epsilon'\xi^2}\right\| = C_\epsilon \left\|\widehat{\Psi}\right\| = C_\epsilon \left\|\Psi\right\|. \tag{35}$$

This expresses the well-known fact that the smoothness property of ψ translates to the Fourier transform in such a way that $\widehat{\psi}$ falls off quickly. Let us now introduce

$$e^{i\theta}x = c + is, \text{ where } c = \cos\theta \cdot x \text{ and } s = \sin\theta \cdot x,$$

for ease of notation. If we use this in (33), we find

$$\left(\hat{S}_\theta\psi\right)(x) = \int_{\mathbb{R}} \widehat{\psi}(\xi)e^{2\pi i\xi(c+is)}\, d\xi = \int_{\mathbb{R}} \widehat{\psi}(\xi)e^{2\pi i\xi c}e^{-2\pi\xi s}\, d\xi.$$

Taking the absolute value gives

$$\left|\left(\hat{S}_\theta\psi\right)(x)\right| \leq \int_{\mathbb{R}} \left|\widehat{\psi}(\xi)\right| \left|e^{2\pi i\xi c}\right| e^{-2\pi\xi s}\, d\xi = \int_{\mathbb{R}} \left|\widehat{\psi}(\xi)\right| e^{\epsilon'\xi^2} \cdot e^{-\epsilon'\xi^2}e^{-2\pi\xi s}\, d\xi.$$

Furthermore, we can use (35) and a simple calculation[13] yielding

$$\left\|e^{-\epsilon'\xi^2}e^{-2\pi\xi s}\right\| = C_\epsilon \exp\left(\frac{\pi^2 s^2}{\epsilon'}\right) = C_\epsilon e^{\epsilon s^2}$$

together with the Cauchy-Schwarz inequality to get the claimed estimate (34). $\quad\square$

---

[12] The assumption here means that ψ can be regarded as solution of the heat equation with L^2-initial data at some later time. In particular, $\psi \in C^\infty(\mathbb{R})$ must hold.

[13] This, as well as the Fourier transform of the heat kernel used above, can be done symbolically with Maxima [17].

3.4 Numerical Demonstrations

In this section we want to demonstrate the performance of the transformations (Zel'dovich transformation and complex scaling) considered above by means of numerical examples. We adhere again to the one-dimensional problems introduced in Section 2.5.

Example 4. While the Zel'dovich transformation works in principle, it is not very efficient and stable in practice. At least for the cases we tried out, the results were not satisfactory. Consider the square-well potential of Example 1 again. It was already discussed above, how the resonance wave functions can be calculated in the usual way. Of course, it is easy to also calculate the transformed wave functions, simply by multiplying them with the Zel'dovich weight. Before we can evaluate the Zel'dovich-scaled Hamiltonian (expressed in a finite basis of Gaussians, like those described in Subsection 4.2.1 below) on it, we also have to express the known exact wave function in this basis. That was done by means of a fit with minimal L^2-norm error (i. e., by a least-squares criterion). Given the resulting coefficient vector, we can verify how well it satisfies the eigenvalue equation of the transformed operator in our basis. If H and S denote the matrices of the Hamiltonian and overlap matrix elements, respectively, and $a \in \mathbb{C}^n$ is the vector of coefficients calculated by the fit, then we want the discretised eigenvalue equation

$$Ha = ESa$$

to be fulfilled for the resonance energy E in each component.[14] If we do not yet know E, the energy can be calculated from the quotient of the left- and right-hand sides of each component equation separately. It turns out that most of those values actually approximate the correct energy well, but some are completely wrong. By taking the median of all real and imaginary parts appearing for all components, this gives quite good an approximation, confirming that the Zel'dovich transformation must actually be somewhat correct. On the other hand, calculating the energies via the eigenvalue equation directly does not work well because of some instabilities. It is also crucial to choose the correct parameter ϵ, as can be seen from Figure 8a: The parameter values shown there are the ones for which the result is near the exact resonance position, but they are from a rather narrow range. If ϵ is outside of this range, the resulting approximate eigenvalues are far from the exact position. Figure 8b shows the corresponding (transformed) eigenfunctions. It can be seen that they fall off quickly for $|x| \to \infty$, if ϵ is large.

Example 5. Consider again the problem of Example 3 in Section 2.5. As mentioned already there, it can be solved numerically with the method of complex scaling. Note, however, that for $\theta \geq \frac{\pi}{4}$ the potential becomes unbounded as then $\mathrm{Re}\left(-\lambda e^{2i\theta}\right) > 0$. While this is a problem in theory, we did not observe any problems caused by it in practice. Furthermore, even if we chose only $\theta < \frac{\pi}{4}$, this would already be large enough to make a lot of the lower resonances accessible. For the numerics, we discretised in space with finite differences and 2,000 grid-points. For the Laplacian, we assumed homogeneous Dirichlet boundary conditions. While we are in fact solving on the whole real line, this is not a problem because after scaling, per construction the solutions *will* have (nearly) zero values at the boundary of a sufficiently large but finite spatial interval.

---

[14]See also (40) and (41) below, although the context there is for complex scaling and not the Zel'dovich transformation.
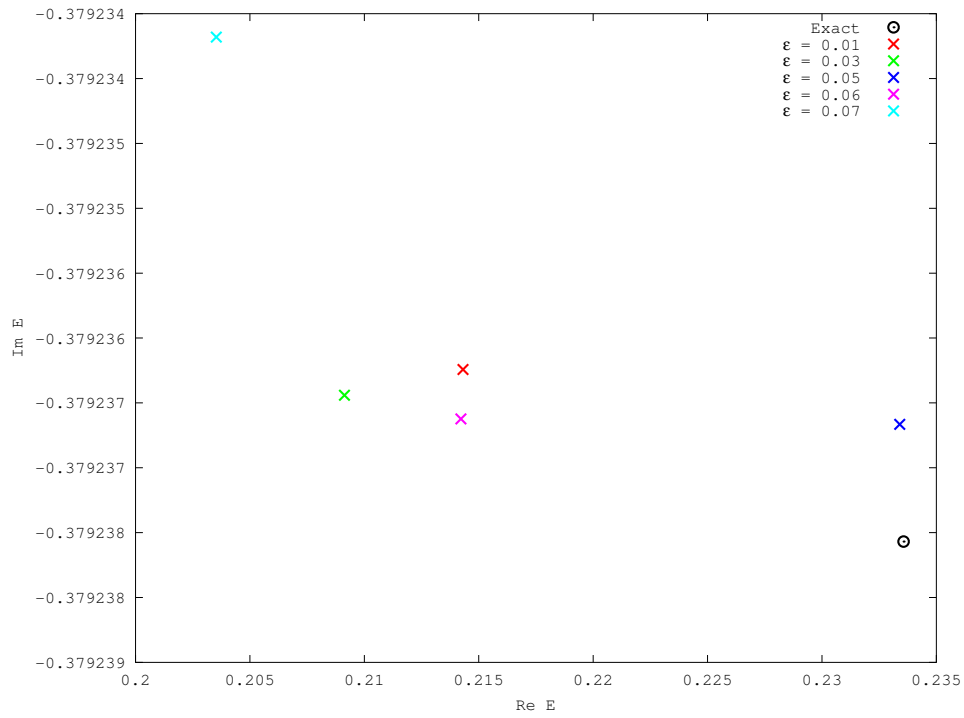
(a) Resonance energies. Note the scaling of the imaginary axis.
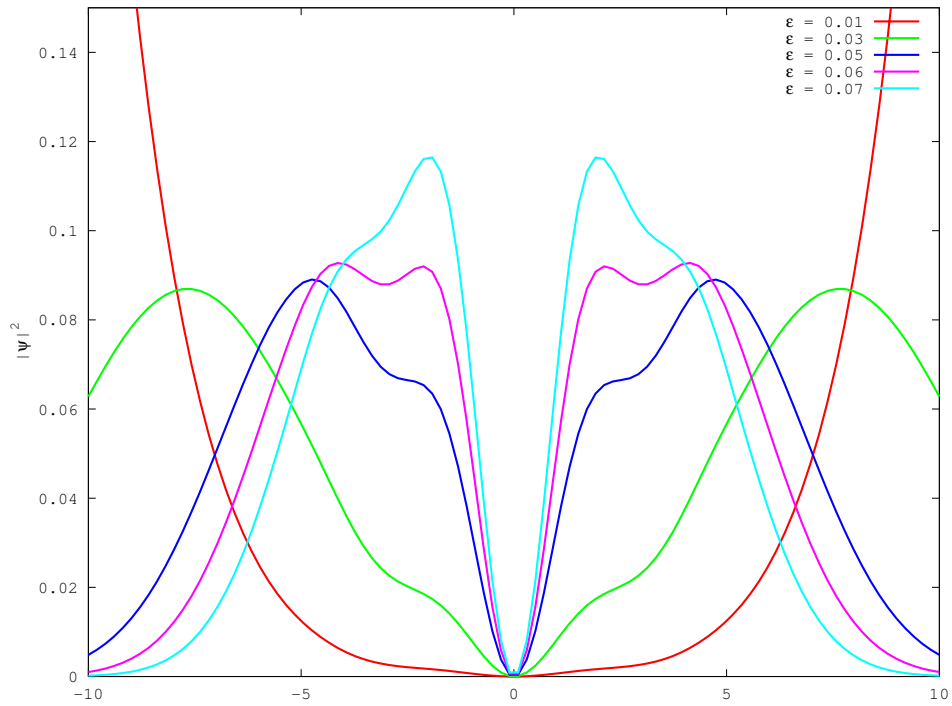
(b) The transformed wave functions corresponding to Figure 8a.

Figure 8: Approximations to one of the resonances in the square-well potential (black dot in the top figure) for different Zel'dovich scaling parameters ϵ, as described in Example 4.

Figure 9 shows the results, with the approximated eigenvalues in the top figures and the eigenfunctions corresponding to low resonances in the bottom. It can be seen how different scaling angles affect the outcome. The red line denotes -2θ, which gives the region in the energy plane (since $E \sim k^2$) that is already accessible with the given scaling angle. It can be observed very nicely how with increasing scaling angle more and more resonances are captured correctly. Note also how the eigenfunctions behave; it is particularly interesting to compare $\theta = 0.1\pi$ and 0.2π, where the resonance pole corresponding to the wave function shown in black is just crossed by the scaling angle. While the eigenfunctions change with the scaling angle, the *eigenvalues* stay the same. This is a fundamental property of a similarity transformation, as explained in Section 3.1 above.
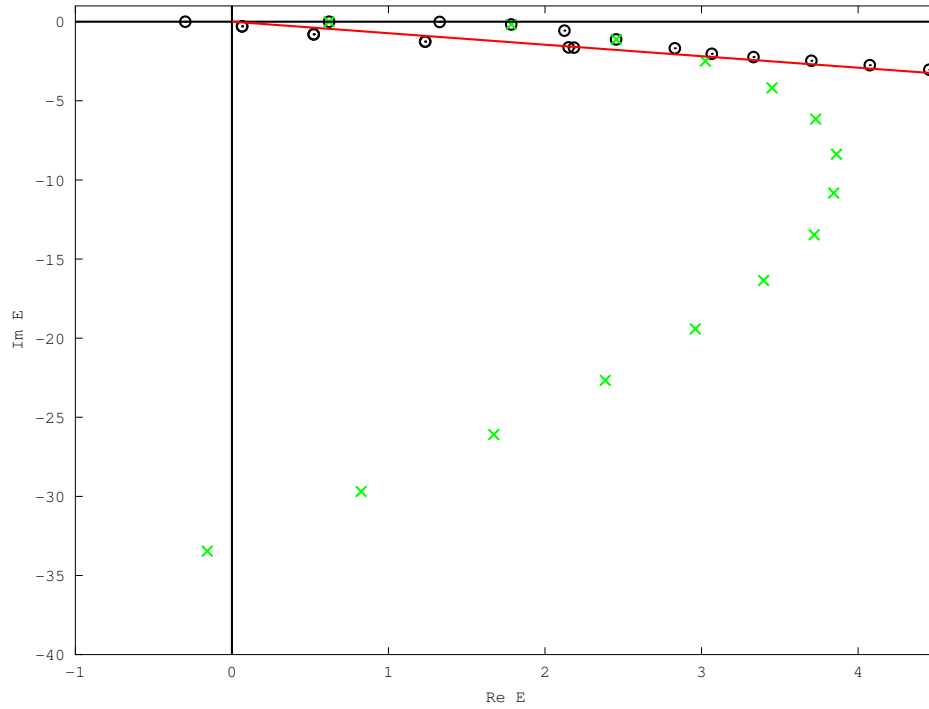
(a) Energies E.
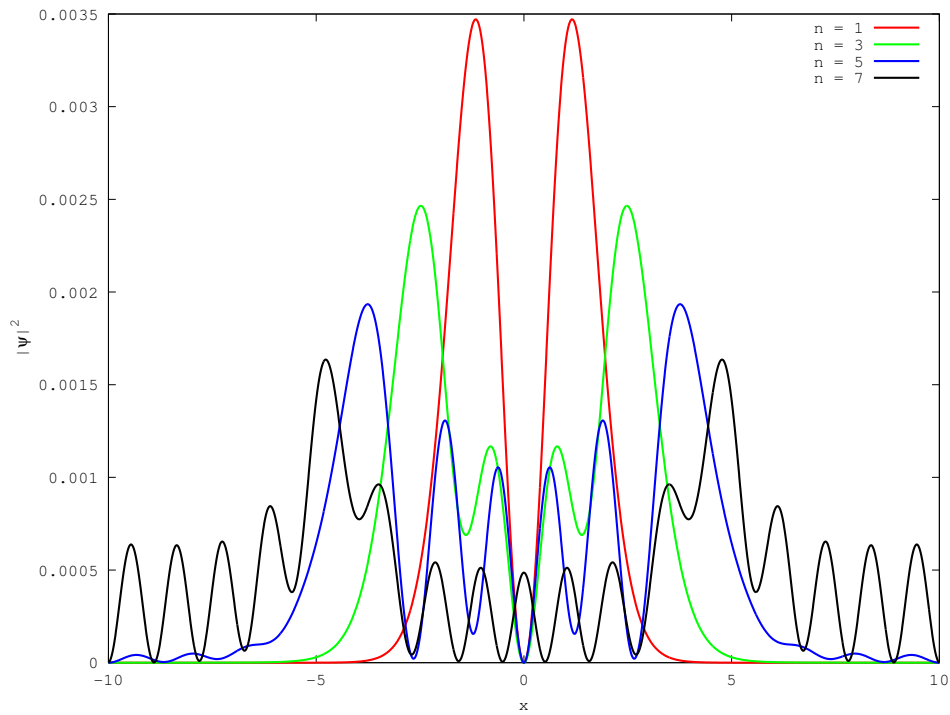
(b) Eigenfunctions $|\psi|^2$.

Figure 9: Results for Example 5 with the complex-scaling method. Green crosses show the values of [9], while black circles are the eigenvalues of the discrete, scaled operator. The red line denotes the scaling angle. In this figure, we show $\theta = 0.1\pi$.
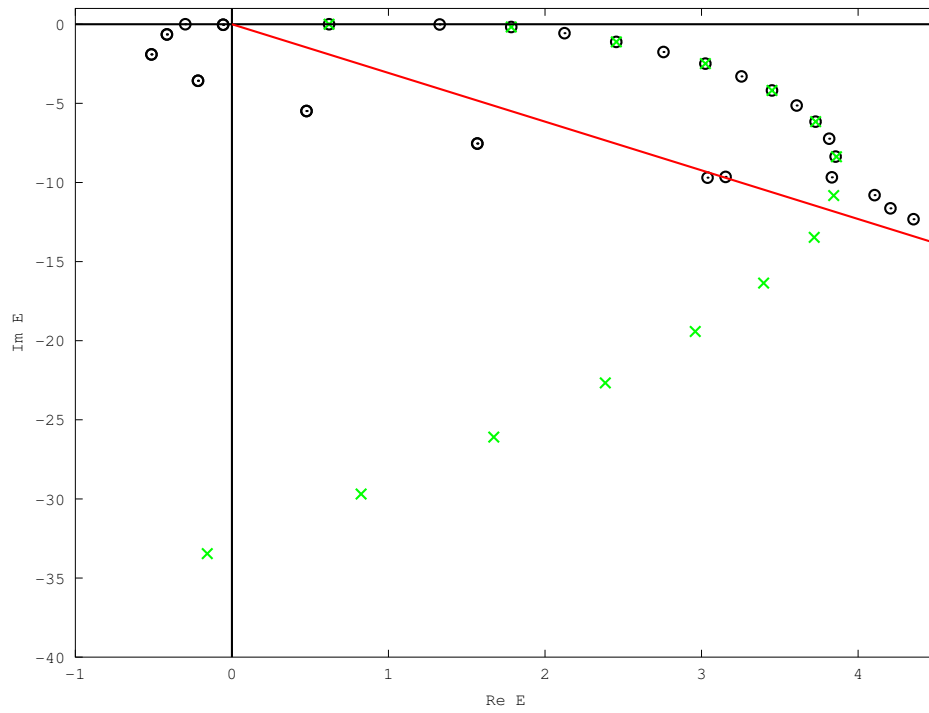
(c) Energies E.
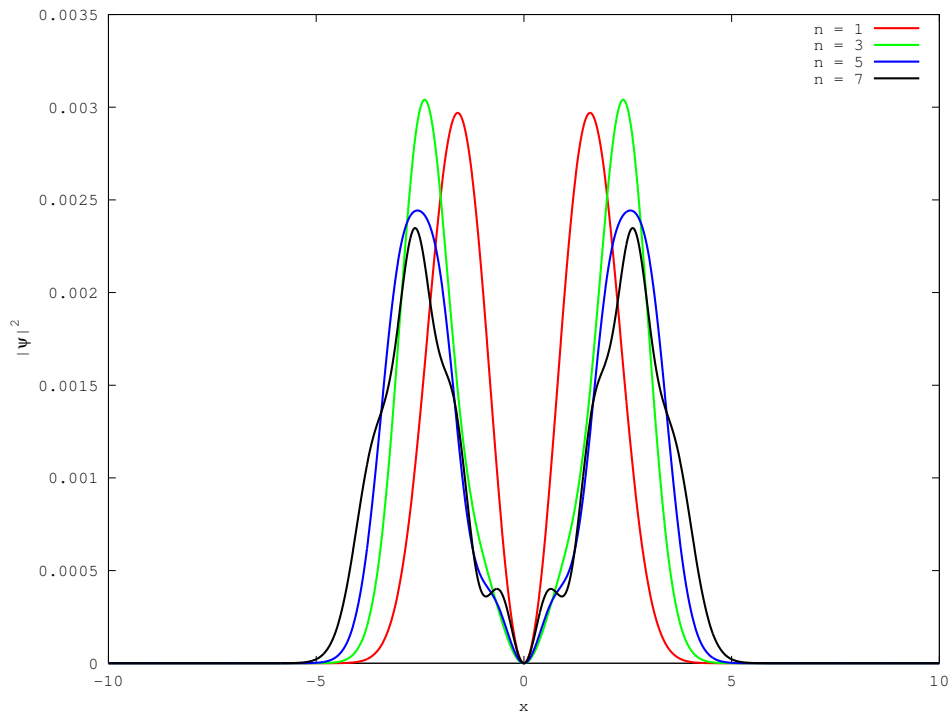
(d) Eigenfunctions $|\psi|^2$.

Figure 9: Continuation of Figure 9. Scaling angle $\theta = 0.2\pi$.
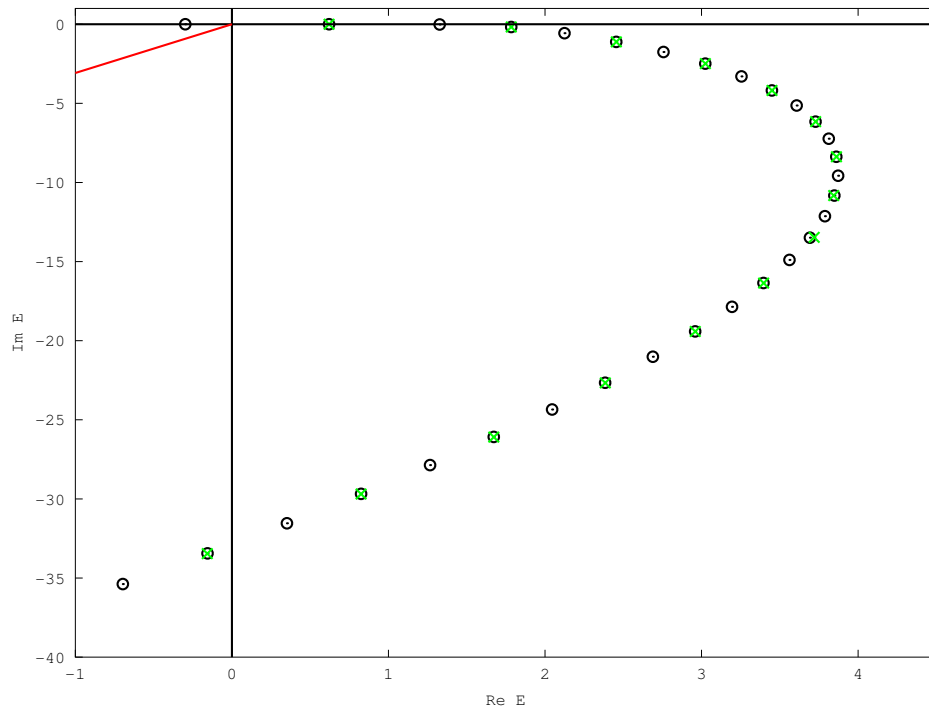
(e) Energies E.
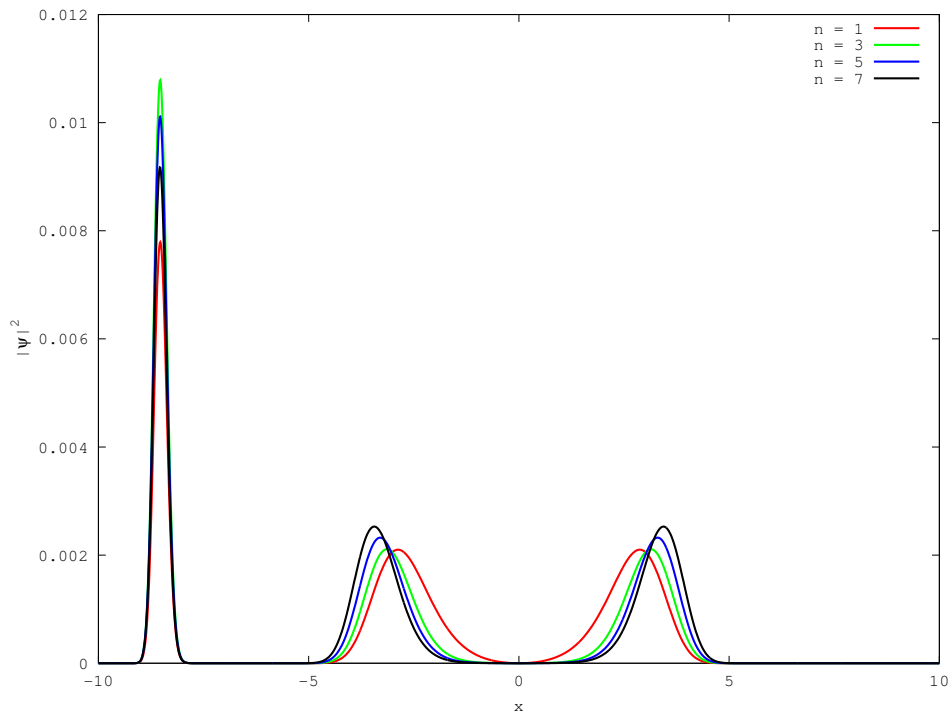
(f) Eigenfunctions $|\psi|^2$.

Figure 9: Continuation of Figure 9. Scaling angle $\theta = 0.3\pi$.

33

4 Variational Methods for Resonances

When the transformations described above in Chapter 3, and in particular the complex-scaling method, have been applied, it remains to solve the resulting non-Hermitian eigenvalue problem. A simple way to do so is an ordinary spatial discretisation, for instance with finite differences as has been done in Example 5. However, in particular for higher-dimensional spaces, the computational costs of those methods rise quickly and exponentially with the number of space dimensions (the so-called "curse of dimensionality"). For these cases, *variational methods* are often more efficient and practical. We want to focus in particular on the *stochastic variational method* [2], which is able to handle *non-linear* variational parameters. The idea is to stochastically (in spirit somewhat similar to a Monte-Carlo method) generate a finite basis of functions adapted to a particular given problem, rather than using a basis corresponding to a straight-forward systematic discretisation. If a good criterion for selecting the basis functions can be devised, this ensures that the chosen basis will indeed be very well suited for the problem at hand, similarly to applying adaptive methods on a spatial grid. In this chapter we want to develop such a variational method for complex-scaled Hamiltonians. Section 4.1 discusses a variational principle that generalises the Rayleigh-Ritz theorem to complex eigenvalues. Section 4.2 introduces the specific choice of basis functions we will use, while Section 4.3 is devoted to the crucial issue of selecting some basis functions over others. Finally, Section 4.4 gives numerical results for an example problem in order to demonstrate the actual performance of the discussed selection procedures.

4.1 A Generalised Variational Principle

As the name already tells, a *variational* method is based on the idea to "vary" an ansatz in order to find the best approximation of an eigenfunction searched for. The theoretical foundation for the classical variational method is the Rayleigh-Ritz theorem:

Theorem 7. *Let $\hat{H}$ be Hermitian and $E_1 \in \mathbb{R}$ be the lowest eigenenergy. If ψ is an arbitrary state, it always holds that*

$$\frac{\left\langle \psi | \hat{H} | \psi \right\rangle}{\langle \psi | \psi \rangle} \geq E_1.$$

Equality holds if and only if ψ is an eigenstate of $\hat{H}$ with energy E_1, i. e., $\hat{H}\psi = E_1\psi$.

Proof. See Theorem 3.1 in [2]. □

However, already at first glance it should be clear that this result cannot be directly generalised to a non-Hermitian operator with complex energies, because, when the eigenvalues are no longer real, it does not make any sense at all to talk about "minimisation" or an ordering. Instead, we will see in Theorem 10 that a possible generalisation replaces the minimisation property by the requirement for a stationary solution; see section 7.2 of [3] for an introduction to this principle. In order to formulate it, Moiseyev first introduces what he calls the "c-product" in chapter 6. Instead of defining the c-product

in one way or the other,[15] we aimed at constructing it axiomatically as far as possible. In general, we have a non-Hermitian (for instance, complex-scaled) linear operator $\hat{H} : H^s(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$, which is assumed to be bounded. The usual Hamiltonian (5) satisfies this when considered as an operator defined on $H^2(\mathbb{R}^n)$ and so does the complex-scaled variant of it, as long as the potential has the necessary regularity properties (in particular, $\hat{S}_\theta V \in H^{-2}(\mathbb{R}^n)$).

Definition 4. The form $(\cdot|\cdot) : L^2(\mathbb{R}^n) \times L^2(\mathbb{R}^n) \to \mathbb{C}$ defines an *appropriate c-product* for $\hat{H}$, if

1. $(\cdot|\cdot)$ is bilinear and symmetric,[16]

2. the Cauchy-Schwarz inequality holds:
$$|(\phi|\psi)| \leq \|\phi\| \, \|\psi\|$$
   for all $\phi, \psi \in L^2(\mathbb{R}^n)$, and

3. $\left(\phi|\hat{H}\psi\right) = \left(\hat{H}\phi|\psi\right)$ for all $\phi, \psi \in H^s(\mathbb{R}^n)$. In other words, $\hat{H}$ is "Hermitian" with respect to the c-product. To emphasise this symmetry, that expression will be denoted as $\left(\phi|\hat{H}|\psi\right)$ in the following.

It is *not* required (and in general not true) that the c-product will be positive-definite or non-degenerate.

If $\hat{H}$ is a complex-scaled Hamiltonian of the usual form (28), possibly even with an a-priori complex potential, then we can define

$$(\phi|\psi) = \int_{\mathbb{R}^n} \phi(x)\psi(x) \, dx. \tag{36}$$

See page 181 of [3] for why this then fulfils Definition 4 and in particular satisfies the "generalised Hermiticity" of $\hat{H}$ (as in the last item in Definition 4 above). Basically, for the kinetic-energy term we can apply integration by parts twice, and the potential can be put to the other operand simply because (36) uses just a commutative product. Since there is no complex conjugation involved, as would be the case for the ordinary L^2-inner product, this also works if the potential is complex (after complex scaling, for instance).

By Definition 4, it is clear that $(\phi|\cdot)$ is a bounded, linear functional on $L^2(\mathbb{R}^n)$ for arbitrary, fixed $\phi \in L^2(\mathbb{R}^n)$. Thus by the Riesz representation theorem (Theorem 2.3 on page 184 of [1]) we can represent it via the ordinary L^2-inner product. For the c-product (36), the representative would be ϕ^*. However, the mapping between a dual element and its representative is no longer linear (or even the identity as in the case of Hilbert spaces), but is now anti-linear. This is of no practical importance, however.

---

[15]In fact, there are several different definitions in the literature. See, for instance, equations (6.11) and (6.14) in [3], the explanation on page 182 therein, and (2.1) in [18].

[16]Note that contrary to the ordinary scalar product, we do *not* want *anti*-linearity or complex conjugates when switching the order!

Theorem 8. *Let* $(\cdot|\cdot)$ *be a c-product for* $\hat{H}$ *according to Definition 4. Assume that* ψ_1 *and* ψ_2 *are both eigenfunctions corresponding to non-degenerate energies:*

$$\hat{H}\psi_1 = E_1\psi_1, \ \ \hat{H}\psi_2 = E_2\psi_2, \ \ E_1 \neq E_2$$

Then $(\psi_1|\psi_2) = 0$. *In other words, the well-known orthogonality property for eigenfunctions carries over to the non-Hermitian case with the c-product.*

Proof. The proof goes just as in the Hermitian case, taking advantage of Definition 4. See also page 181 of [3]. We have:

$$\left(\psi_1|\hat{H}|\psi_2\right) = \left(\psi_1|\hat{H}\psi_2\right) = E_2\left(\psi_1|\psi_2\right), \ \ \left(\psi_1|\hat{H}|\psi_2\right) = \left(\hat{H}\psi_1|\psi_2\right) = E_1\left(\psi_1|\psi_2\right).$$

Subtracting both relations yields

$$0 = (E_1 - E_2)\left(\psi_1|\psi_2\right),$$

thus showing c-orthogonality, since $E_1 - E_2 \neq 0$ by assumption. $\qquad\square$

Theorem 9. *Let* $\hat{H}\psi = E\psi$, *with* $\psi \in H^s(\mathbb{R}^n)$ *and* $(\psi|\psi) \neq 0$. *Let* $\delta\psi \in H^s(\mathbb{R}^n)$ *be an arbitrary perturbation with* $\|\delta\psi\|_{H^s} \leq 1$.[17] *We define the Rayleigh quotient as functional* $D \subset H^s(\mathbb{R}^n) \to \mathbb{C}$:

$$R(\phi) = \frac{\left(\phi|\hat{H}|\phi\right)}{(\phi|\phi)}, \tag{37}$$

which is well-defined for the set D *of c-normalisable functions.*[18] *Then for*

$$\phi = \psi + \epsilon\delta\psi,$$

we have that

$$R(\phi) = E + O\left(\epsilon^2\right)$$

as $\epsilon \to 0$, *uniformly in* $\delta\psi$. *In particular, this implies that all directional derivatives of* R *vanish in* $\psi \in H^s(\mathbb{R}^n)$, *as well as its Fréchet derivative. In other words,* ψ *is a critical point of* R, *where the critical value is just the exact eigenvalue.*[19]

Proof. First of all, we can write $\delta\psi = \alpha\psi + \delta\psi^\perp$, where we set

$$\alpha = \frac{(\delta\psi|\psi)}{(\psi|\psi)}, \ \ \delta\psi^\perp = \delta\psi - \alpha\psi.$$

Then clearly
$$\left(\delta\psi^\perp|\psi\right) = (\delta\psi|\psi) - \alpha\left(\psi|\psi\right) = (\delta\psi|\psi) - (\delta\psi|\psi) = 0,$$

---

[17]We do not need that $\delta\psi$ is c-normalisable, since an upper bound is enough and the c-product can be bounded by the ordinary L^2-norm. Furthermore, of course, $\|\delta\psi\| \leq \|\delta\psi\|_{H^s} \leq 1$ by the embedding of $L^2(\mathbb{R}^n)$ into $H^s(\mathbb{R}^n)$.

[18]Since $\phi \mapsto (\phi|\phi)$ is continuous in the L^2-topology, D contains at least an L^2- (and thus also H^s-) neighbourhood of $\psi \in D$. This is enough for our interests.

[19]See also (7.17) in [3].

so that we found a c-orthogonal decomposition. Furthermore, if we define $C = \frac{1}{(\psi|\psi)} \neq 0$, we have the bound

$$|\alpha| \leq |C| \, \|\delta\psi\| \, \|\psi\| \leq |C| \, \|\psi\| \leq |C| \, \|\psi\|_{H^s}$$

and consequently also

$$\left\|\delta\psi^\perp\right\|_{H^s} \leq \|\delta\psi\|_{H^s} + |\alpha| \, \|\psi\|_{H^s} \leq 1 + |C| \, \|\psi\|_{H^s}^2 = C'.$$

Take note that the constant C' is uniform in $\delta\psi$ and only depends on $\hat{H}$ and ψ. Clearly, also the identity $\phi = (1 + \epsilon\alpha)\psi + \epsilon\delta\psi^\perp$ holds.

Now we calculate the Rayleigh quotient using this decomposition and the fact that $\delta\psi^\perp$ is c-orthogonal to ψ:

$$
\begin{aligned}
R(\phi) &= \frac{(1 + \epsilon\alpha)^2 \left(\psi|\hat{H}|\psi\right) + 2\epsilon(1 + \epsilon\alpha)\left(\delta\psi^\perp|\hat{H}|\psi\right) + \epsilon^2\left(\delta\psi^\perp|\hat{H}|\delta\psi^\perp\right)}{(1 + \epsilon\alpha)^2\,(\psi|\psi) + 0 + \epsilon^2\,(\delta\psi^\perp|\delta\psi^\perp)} \\
&= \frac{E(1 + \epsilon\alpha)^2\,(\psi|\psi) + \epsilon^2\left(\delta\psi^\perp|\hat{H}|\delta\psi^\perp\right)}{(1 + \epsilon\alpha)^2\,(\psi|\psi) + \epsilon^2\,(\delta\psi^\perp|\delta\psi^\perp)}
\end{aligned}
$$

If ϵ is small enough, $1 + \epsilon\alpha$ will be non-zero (and in fact, since α is bounded, converge to 1 uniformly in $\delta\psi$ as $\epsilon \to 0$). Let us define

$$C_\epsilon = \frac{1}{(1 + \epsilon\alpha)^2\,(\psi|\psi)},$$

which has the limit $C_\epsilon \to C$ for $\epsilon \to 0$. Then we get

$$R(\phi) - E = \frac{E - E(1 + C\epsilon^2\left(\delta\psi^\perp|\delta\psi^\perp\right))}{1 + C\epsilon^2\left(\delta\psi^\perp|\delta\psi^\perp\right)} + C\epsilon^2\frac{Q}{1 + C\epsilon^2\left(\delta\psi^\perp|\delta\psi^\perp\right)} = C\epsilon^2\frac{Q - \left(\delta\psi^\perp|\delta\psi^\perp\right)}{D},$$

where

$$|Q| = \left|\left(\delta\psi^\perp|\hat{H}|\delta\psi^\perp\right)\right| \leq \left\|\delta\psi^\perp\right\|\left\|\hat{H}\delta\psi^\perp\right\| \leq C'\left\|\hat{H}\right\|\left\|\delta\psi^\perp\right\|_{H^s} \leq (C')^2\left\|\hat{H}\right\|$$

is bounded and

$$\lim_{\epsilon \to 0} D = \lim_{\epsilon \to 0}\left(1 + C\epsilon^2\left(\delta\psi^\perp|\delta\psi^\perp\right)\right) = 1,$$

both uniformly in $\delta\psi$. Consequently we know that the limit

$$\lim_{\epsilon \to 0}\frac{R(\phi) - E}{\epsilon^2} = \frac{Q - \left(\delta\psi^\perp|\delta\psi^\perp\right)}{(\psi|\psi)}$$

exists and is finite, showing the claim. $\qquad\square$

Note that Theorem 9 is still valid if $\delta\psi$ is not normalised, while, of course, it always needs to be in $H^s(\mathbb{R}^n)$. Then, however, the result is not uniform in $\delta\psi$ and instead the constant in front of the $O(\epsilon^2)$ term depends on $\|\delta\psi\|_{H^s}$.

Theorem 10. *Let $\psi(\alpha) \in H^s(\mathbb{R}^n)$ be a parameter-dependent wave function, where we assume for simplicity $\alpha \in \mathbb{C}$.[20] Let $\psi(\cdot)$ be differentiable and assume that the eigenvalue equation $\hat{H}\psi(\alpha_0) = E\psi(\alpha_0)$ is satisfied for a particular α_0 as well as $(\psi(\alpha_0)|\psi(\alpha_0)) \neq 0$. We consider the Rayleigh quotient of (37) now as a function of α. Then*

$$\frac{\partial R(\alpha_0)}{\partial \alpha} = 0. \tag{38}$$

Proof. We begin by defining $\delta\psi = \frac{\partial\psi(\alpha_0)}{\partial\alpha}$. It follows that

$$\psi(\alpha_0 + \epsilon) = \psi(\alpha_0) + \epsilon\delta\psi + r(\epsilon),$$

where $r(\epsilon) \in H^s(\mathbb{R}^n)$ and $r(\epsilon) = o(\epsilon)$ in the H^2-norm; this is evident, since we require differentiability of $\psi(\cdot)$ in $H^2(\mathbb{R}^n)$. From Theorem 9 we know already that the stationarity property $R(\psi(\alpha_0) + \epsilon\delta\psi) = E + O(\epsilon^2)$ holds (as remarked after Theorem 9 above, it is not important to normalise $\delta\psi$ for the argument we have in mind here). For ease of notation, define

$$Q = \frac{(\psi(\alpha_0) + \epsilon\delta\psi|\psi(\alpha_0) + \epsilon\delta\psi)}{(\psi(\alpha_0 + \epsilon)|\psi(\alpha_0 + \epsilon))}.$$

Note that for $\epsilon \to 0$, we have

$$|Q - 1| = \left| \frac{(\psi(\alpha_0) + \epsilon\delta\psi|\psi(\alpha_0) + \epsilon\delta\psi) - (\psi(\alpha_0 + \epsilon)|\psi(\alpha_0 + \epsilon))}{(\psi(\alpha_0 + \epsilon)|\psi(\alpha_0 + \epsilon))} \right| \leq \|r(\epsilon)\| C_\epsilon$$

with C_ϵ bounded for $\epsilon \to 0$. Then

$$|R(\alpha_0 + \epsilon) - R(\psi(\alpha_0) + \epsilon\delta\psi)| \leq \left| \frac{(Q-1)\left(\psi(\alpha_0) + \epsilon\delta\psi|\hat{H}|\psi(\alpha_0) + \epsilon\delta\psi\right)}{(\psi(\alpha_0) + \epsilon\delta\psi|\psi(\alpha_0) + \epsilon\delta\psi)} \right| + \|r(\epsilon)\| C_\epsilon' \leq \|r(\epsilon)\| C_\epsilon'',$$

where $C_\epsilon', C_\epsilon''$ are also bounded for $\epsilon \to 0$. Hence

$$R(\alpha_0 + \epsilon) = R(\psi(\alpha) + \epsilon\delta\psi) + o(\epsilon) = E + o(\epsilon)$$

for $\epsilon \to 0$, showing the claim. $\qquad\square$

Theorem 10 gives the statement that we must look for a stationary energy expectation value to find the best (non-linear) variational parameters, and its proof shows rigorously[21] the validity of (7.28) in [3]. Finally, we want to deduce the *linear* variational principle from Theorem 10. For this, assume that we expand the exact eigenfunction ψ with $\hat{H}\psi = E\psi$ into a basis of $H^s(\mathbb{R}^n)$, say

$$\psi = \sum_{k=1}^{\infty} a_k\chi_k \approx \sum_{k=1}^{N} a_k\chi_k.$$

---

[20]Of course, it is trivial to extend this to the situation of a multi-dimensional parameter space.

[21]Note, however, that this is, in fact, only true if the ansatz is chosen in such a way that it is possible to *exactly* achieve the correct eigenfunction with a suitable choice of parameters. This is not true in practice, but the secular equations (38) still hold approximately.

Consider again the Rayleigh quotient (37):

$$E = R(\psi) = \frac{\sum_{k,l=1}^{\infty} a_k a_l \left(\chi_k | \hat{H} | \chi_l \right)}{\sum_{k,l=1}^{\infty} a_k a_l \left(\chi_k | \chi_l \right)} \quad \Leftrightarrow \quad \sum_{k,l=1}^{\infty} a_k a_l \left(\chi_k | \hat{H} | \chi_l \right) = \left(\sum_{k,l=1}^{\infty} a_k a_l \left(\chi_k | \chi_l \right) \right) R(\psi)$$

Now take the derivative with respect to a_k on both sides, and note that $\frac{\partial R(\psi)}{\partial a_k} = 0$ because of Theorem 10. This yields

$$\sum_{l=1}^{\infty} \left(\chi_k | \hat{H} | \chi_l \right) a_l = E \sum_{l=1}^{\infty} \left(\chi_k | \chi_l \right) a_l. \tag{39}$$

Assuming that (39) still holds if we restrict the sums to be finite, we can proceed to the corresponding statement for a finite basis: Define the two matrices

$$H_{kl} = \left(\chi_k | \hat{H} | \chi_l \right), \quad S_{kl} = \left(\chi_k | \chi_l \right), \tag{40}$$

then we find that for the vector a of coefficients we have to require

$$Ha = ESa. \tag{41}$$

Consequently, if we have a finite basis, we can approximate the infinite-dimensional eigenvalue problem of $\hat{H}$ by calculating the matrices H, S and then solving the generalised eigenvalue problem (41) for the expansion coefficients a_k in the used basis. For a treatment of the variational principle in the conventional Hermitian setting, see, for instance, section 5.4 of [4]. Note that the assumption made, namely that (39) holds even when truncated to a finite expansion, is usually not really justified. In practice, however, (39) still holds approximately in this case, and eigenvalues of (41) approximate the eigenvalues of $\hat{H}$ well if the basis is chosen large enough.

4.2 Stochastic Basis Functions

Since the goal is to use (41) to approximate the eigenvalues of a non-Hermitian (scaled) Hamiltonian, we have to find some basis and calculate the matrices H and S first. In this section we will address these points for a specific choice of basis functions. As the calculations here involve a lot of Gaussian integrals, not everything is detailed below. The necessary calculations were instead done with Maxima [17].

4.2.1 Gaussian Basis Functions

Definition 5. Let $\alpha > 0$ and $\mu \in \mathbb{R}^n$. We define the Gaussian basis function corresponding to these parameters as

$$\chi_{\alpha,\mu}(x) = e^{-\alpha |x - \mu|^2}, \tag{42}$$

where $x \in \mathbb{R}^n$ can be (at the moment) multi-dimensional.

The Gaussian basis functions (42) have several nice properties: They are given by analytic expressions and are smooth (in fact, whole functions), dense in $H^s(\mathbb{R}^n)$, localised
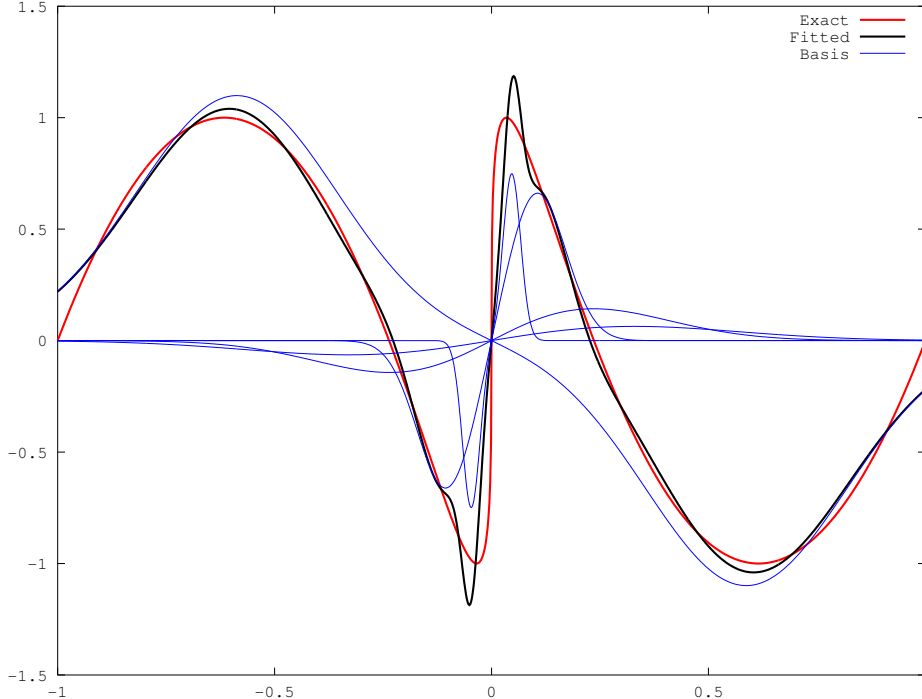
Figure 10: Least-squares fit (black) of a given function (red) in a basis of five Gaussians according to Definition 5. Since the target function is odd, the basis functions have been mirrored accordingly to make them odd, too. The basis functions times their corresponding coefficients are shown in blue (i. e., the sum of the blue curves gives the black approximation curve). Note that this is not strictly the best possible fit for any five Gaussian basis functions, but one that illustrates the power of the approximation by means of Gaussians nicely.

(not exactly with compact support but "almost" so) and they can resolve fine details where necessary by choosing a large α. This is nicely illustrated in Figure 10: Through a combination of wide and narrow Gaussians, the given function (in red) can be approximated reasonably well. In the centre, where a lot of details are present, more basis functions have their support, while in the outer regions not so many are necessary. See also section II/A of [19]. One can even add a more general correlation matrix to (42), but this is only a technical detail and is not considered here.

Now assume the one-dimensional case for simplicity. In order to express the integrals that represent the elements of the matrices S and H of (40) in closed form, let us first introduce a short-hand notation for a common sub-expression:

$$E(\alpha, \beta, \mu, \nu) = \exp\left(-\frac{\alpha\beta}{\alpha + \beta}(\mu - \nu)^2\right)$$

Then the overlaps of basis functions (42) (the entries of S in (40)) are:

$$(\chi_{\alpha,\mu}|\chi_{\beta,\nu}) = \int_{\mathbb{R}} e^{-\alpha(x-\mu)^2 - \beta(x-\nu)^2}\,dx = \sqrt{\frac{\pi}{\alpha + \beta}} \cdot E(\alpha, \beta, \mu, \nu) \tag{43}$$

The kinetic-energy part of the Hamiltonian matrix element is:

$$
\begin{aligned}
\left(\chi_{\alpha,\mu}|(-\tfrac{1}{2}\Delta)|\chi_{\beta,\nu}\right) &= -\tfrac{1}{2}\int_{\mathbb{R}} e^{-\alpha(x-\mu)^2} \cdot \tfrac{d^2}{dx^2} e^{-\beta(x-\nu)^2} \, dx \\
&= \tfrac{1}{2}\int_{\mathbb{R}} \tfrac{d}{dx} e^{-\alpha(x-\mu)^2} \cdot \tfrac{d}{dx} e^{-\beta(x-\nu)^2} \, dx \\
&= \tfrac{\sqrt{\pi}}{(\alpha+\beta)^{\frac{5}{2}}} \alpha\beta \left(\alpha + \beta - 2\alpha\beta(\mu-\nu)^2\right) \cdot E(\alpha,\beta,\mu,\nu)
\end{aligned}
\tag{44}
$$

All what remains is to calculate the (scaled) potential part of the Hamiltonian matrix element, which can be done—depending on the potential—also symbolically or simply by some numerical quadrature scheme. Then one can use (28) together with the above results to build up the matrix H and solve (41) numerically.

4.2.2 Orthonormalisation of the Basis

Let us assume for the moment that we have an *arbitrary* basis of functions; they need not be the Gaussians (42). Let us call it (χ_k), where $k = 1, 2, \ldots$. They are, of course, linearly independent (because we want them to be a *basis*), but usually neither orthogonal nor normalised (although it would not be hard to normalise them). For instance, they may be Gaussians according to Definition 5 with randomly chosen parameters α_k and μ_k. However, in some circumstances it is preferable to express the variational method in terms of an *orthonormal basis* rather than an arbitrary one. To make (χ_k) orthonormal, the well-known *Gram-Schmidt* procedure[22] can be applied. If we denote the correspondingly orthonormalised basis by $(\tilde{\chi}_k)$, then this procedure means nothing else than[23]

$$
\tilde{\chi}_k = C_k \left(\chi_k - \sum_{i=1}^{k-1} \langle \tilde{\chi}_i | \chi_k \rangle \, \tilde{\chi}_i \right),
\tag{45}
$$

where C_k is the proper normalisation constant (namely just the reciprocal of the norm of the parenthesised expression). Since the (χ_k) are linearly independent, it is clear that (45) only produces *non-zero* functions $(\tilde{\chi}_k)$. Hence it is guaranteed that we can normalise them. Furthermore, it is easy to see that the span of the first N basis functions is not affected by the Gram-Schmidt procedure (45).

Lemma 2. *For $(\tilde{\chi}_k)$ according to (45), one has $\langle \tilde{\chi}_k | \tilde{\chi}_l \rangle = \delta_{kl}$ for all $k, l = 1, 2, \ldots$.*

Proof. For $k = l$ this is clear per definition of C_k as the proper normalisation constant. So assume $l < k$. We use a proof by induction in k, thus assume further that $\langle \tilde{\chi}_l | \tilde{\chi}_i \rangle = \delta_{li}$ for all $i < k$. But then from (45) and by linearity of the inner product in the second argument:

$$
\langle \tilde{\chi}_l | \tilde{\chi}_k \rangle = C_k \left(\langle \tilde{\chi}_l | \chi_k \rangle - \sum_{i=1}^{k-1} \langle \tilde{\chi}_i | \chi_k \rangle \langle \tilde{\chi}_l | \tilde{\chi}_i \rangle \right) = C_k \left(\langle \tilde{\chi}_l | \chi_k \rangle - \langle \tilde{\chi}_l | \chi_k \rangle \right) = 0
$$

$\square$

---

[22] A description can be found, for instance, in subsection 4.8.2 of [20] and in many other textbooks.

[23] In this subsection and in general for orthonormalisation, the *ordinary inner product* and not the c-product is used. This has to be done because we still want to measure differences (or approximation errors) in the ordinary L^2-norm induced by it and want to have Parseval's identity (see Theorem 12 below) available after orthonormalisation. This will be applied in Subsection 4.3.3. However, since the basis functions are usually real, there is no difference to the c-product for them anyway.

Thus it is shown that (45) indeed does what we expect it to do: Transforming an arbitrary basis into an orthonormal one without changing the basic properties of it, i. e., being a basis that spans the same subspace.

Next we want to express this procedure in terms of matrices. Let S be as before the (symmetric) matrix of overlaps in the basis, $S_{kl} = \langle \chi_k | \chi_l \rangle$. This is, of course, not the identity as it would be if we already had an orthonormal basis, but it is a symmetric, positive-definite matrix. By S^k we will denote the $k \times k$ submatrix of only the first k basis functions. We want to construct M^k (again denoting the matrix for the first k basis functions) such that

$$\begin{pmatrix} \tilde{\chi}_1 \\ \vdots \\ \tilde{\chi}_k \end{pmatrix} = M^k \begin{pmatrix} \chi_1 \\ \vdots \\ \chi_k \end{pmatrix} \tag{46}$$

holds according to (45). Clearly this matrix will be lower triangular because of the "forward propagation" nature of the Gram-Schmidt process. If we have $l < k$, then assuming we already have at least l orthonormalised basis functions, we can write

$$\langle \tilde{\chi}_l | \chi_k \rangle = \left\langle \sum_{i=1}^{l} M_{li}^k \chi_i \Big| \chi_k \right\rangle = \sum_{i=1}^{l} M_{li}^k \langle \chi_i | \chi_k \rangle .$$

Note that M is a real matrix because we use real basis functions and thus linearity also holds for the first argument in the inner product. Hence the products on the right-hand side of (45) can be expressed as

$$p_k = \begin{pmatrix} \langle \tilde{\chi}_1 | \chi_k \rangle \\ \vdots \\ \langle \tilde{\chi}_{k-1} | \chi_k \rangle \end{pmatrix} = M^k \cdot (S_{ik})_{i=1}^{k} .$$

Here, $(S_{ik})_{i=1}^{k}$ denotes the vector formed of the first k elements in the k-th column of the matrix S. This then gives the final reformulation of (45) as

$$\tilde{\chi}_{k+1} = C_{k+1}(-p_{k+1}^\top \ \ 1) \begin{pmatrix} \tilde{\chi}_1 \\ \vdots \\ \tilde{\chi}_k \\ \chi_{k+1} \end{pmatrix} = C_{k+1}(-p_{k+1}^\top \ \ 1) \cdot \begin{pmatrix} M^k & 0 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \chi_1 \\ \vdots \\ \chi_{k+1} \end{pmatrix},$$

which also gives the last row to add to M^k in order to produce M^{k+1} from it. For the norm, observe that we can rewrite (45) as

$$\sum_{i=1}^{k-1} \langle \tilde{\chi}_i | \chi_k \rangle \, \tilde{\chi}_i + \frac{1}{C_k} \tilde{\chi}_k = \chi_k.$$

If we now "square" both sides with the inner product, we can apply Pythagoras' theorem on the left-hand side (since the vectors there are orthonormal), which yields

$$\sum_{i=1}^{k-1} \langle \tilde{\chi}_i | \chi_k \rangle^2 + \frac{1}{C_k^2} = \| p_k \|_2^2 + \frac{1}{C_k^2} = \langle \chi_k | \chi_k \rangle .$$

From this it is trivial to calculate C_k. So now we have a nice theoretical description of the Gram-Schmidt procedure for abstract basis functions expressed in terms of matrix and vector operations. All we have to know about the basis functions is the matrix S.

Lemma 3. *Consider a finite basis of the first m elements, and let ψ be in their span:*

$$\psi = \sum_{i=1}^{m} a_i \chi_i = \sum_{i=1}^{m} \tilde{a}_i \tilde{\chi}_i$$

Then the relation between the coefficients in the original and orthonormalised bases is given as

$$a = M^\top \tilde{a}, \quad \tilde{a} = M^{-\top} a, \tag{47}$$

where $M = M^m$ is the matrix from (46).

Also, if $\hat{O}$ is an arbitrary operator on the finite-dimensional subspace spanned by this part of the basis, we can define the matrix elements[24] similarly to (40):

$$O_{kl} = \left(\chi_k | \hat{O} | \chi_l \right), \quad \tilde{O}_{kl} = \left(\tilde{\chi}_k | \hat{O} | \tilde{\chi}_l \right)$$

Then these basis representations of $\hat{O}$ are related by

$$\tilde{O} = MOM^\top, \quad O = M^{-1}\tilde{O}M^{-\top}. \tag{48}$$

Proof. If we denote the "vectors" of the first m basis functions in either basis by χ and $\tilde{\chi}$, respectively, note that per (46) we get $\tilde{\chi} = M\chi$. Thus

$$\psi = \tilde{a}^\top \tilde{\chi} = \tilde{a}^\top M\chi = a^\top \chi,$$

and hence we can conclude (also take uniqueness of the basis representation into account):

$$\tilde{a}^\top M = a^\top \iff a = M^\top \tilde{a} \iff \tilde{a} = M^{-\top} a$$

Note that M and $M^\top$ are invertible since they are triangular matrices with non-zero entries on the diagonal.

For the matrix elements, we have similarly

$$
\begin{aligned}
\tilde{O}_{kl} &= \left(\tilde{\chi}_k | \hat{O} | \tilde{\chi}_l \right) = \left(\sum_{i=1}^{m} M_{ki}\chi_i | \hat{O} | \sum_{j=1}^{m} M_{lj}\chi_j \right) = \sum_{i,j=1}^{m} M_{ki}M_{lj} \left(\chi_i | \hat{O} | \chi_j \right) \\
&= \sum_{i,j=1}^{m} M_{lj}M_{ki}O_{ij} = \sum_{j=1}^{m} (MO)_{kj}M_{lj} = (MOM^\top)_{kl}.
\end{aligned}
$$

Thus also the second claim follows. $\square$

We can now conclude two important results about how to find the orthonormal basis in practice and how it can be employed; see also Subsection 4.3.3 below for another possible use.

---

[24]Here we indeed use the *c-product* again because the operators we need (in particular H and S from (40)) will be made up from matrix elements in the c-product. This is a completely different and independent issue to the inner product used for orthonormalisation.

Theorem 11. *Let $S = LL^\top$ be the Cholesky factorisation*[25] *of S, then $M = L^{-1}$.*

E is a generalised eigenvalue to (41) if and only if E is an ordinary eigenvalue of $\tilde{H} = MHM^\top$. The eigenvector $\tilde{a}$ of coefficients in the orthonormalised basis corresponds to the coefficients a in (41) via $\tilde{a} = M^{-\top}a$ as per (47).

Proof. We first apply (48) to the identity. Note that the corresponding matrix is just S per (40). Because the $\tilde{\chi}_k$ are per definition orthonormal, the transformed matrix $\tilde{S}$ built up from c-inner products of the orthonormalised basis is instead the identity matrix. Thus

$$S = M^{-1}\tilde{S}M^{-\top} = M^{-1}M^{-\top} = LL^\top \tag{49}$$

with $L = M^{-1}$. Because M is lower-triangular, the same also holds for L as can be seen easily from the procedure of Gaussian elimination. The form (49) is, however, nothing else than the (unique) Cholesky decomposition of S.

For the second part, let $E \in \mathbb{C}$ and $a \in \mathbb{R}^m$ be given, and set $\tilde{a} = M^{-\top}a$. Then:

$$\tilde{H}\tilde{a} = E\tilde{a} \quad \Leftrightarrow \quad MHM^\top \cdot M^{-\top}a = EM^{-\top}a \quad \Leftrightarrow \quad Ha = EM^{-1}M^{-\top}a = ESa$$

$\square$

4.3 Selection Principles for the Basis

We are still trying to use (41) derived above in Section 4.1 to calculate eigenvalues. In Section 4.2 we already described which class of basis functions we want to use for that. In this section we will discuss how one can choose the parameters α_k and μ_k in (42) in order to find a "good" basis. Of course, the most straight-forward procedure is to simply generate them altogether *randomly*. In our implementation, the μ's are uniformly distributed on some finite interval, which is chosen a-priori according to the expected support of the scaled eigenfunction one is trying to find. The α's are selected uniformly from a *logarithmic* interval of plausible widths. This seems better than a uniform distribution, because the interval of seemingly possible α's is something like 10^{-2} to 10^1 (for the test problem we considered). Since α must be positive, a logarithmic range makes more sense than a linear one.

In order to do better, the general idea in the stochastic variational method (see also section II/B of [19]) is to build up the basis step by step, and for each element that should be added to the basis, *multiple* candidate functions are generated. Then each of those possibilities is evaluated according to some selection criterion, and the new basis function that seems to be best is chosen as the real next one and added to the basis. In conventional Hermitian quantum mechanics, the obvious selection criterion is minimisation of the eigenvalues according to the Rayleigh-Ritz principle. In other words, at each step that candidate function is chosen, which generates the smallest eigenvalue (or for which a suitable "basket" of the lowest few eigenvalues is minimal). With complex eigenvalues and non-Hermitian operators, this is unfortunately no longer possible. Below we discuss some possible alternative strategies. In Section 4.4 those will be compared to each other for a practical calculation.

---

[25] A general discussion of the Cholesky factorisation and its properties can be found in section 4.5 of [20]. Note that the explicit procedure for calculating M given above is very similar to the well-known Cholesky algorithm, which can also be found in Figure 4.5 of [20].

4.3.1 Improving Poorly Conditioned Matrices

When one starts to do numerics with randomly chosen basis functions of the form (42), one quickly finds out that by chance some of the chosen functions will have quite a large overlap. In other words, the chosen basis will become "less and less linearly independent". Of course, in mathematical terms there is only linear independence or dependence, with no intermediates. However, in practice the matrix S from (40) may become worse and worse conditioned. In fact, its condition number[26] grows exponentially with the basis size. For the orthonormalisation procedure described in Subsection 4.2.2 this is a problem, as with already about 50 basis functions the Cholesky factorisation is no longer possible numerically due to inaccuracies. In order to combat this problem, one can add an additional selection step for new basis functions as described above, and pick from a number of candidate functions that one, which results in the lowest condition number of the new S. While this step does, of course, nothing to help us find a basis that is well-suited for solving the eigenvalue problem, it helps with, and even allows us again, to use orthonormalisation. As we will see below in Figure 13a, this selection criterion *alone* does about as well as a purely random selection. However, we can add an additional selection *on top* of the condition-number-based criterion, which will *then* allow improvements. Figure 11a shows how the condition number grows with the basis size, where the shown curves correspond to different numbers of candidates generated at each step ("trials"). The black curve with only a single candidate corresponds to no selection, and we can see that only five candidates (green curve) reduce the condition number already to about the square root of the original one. This is the number of candidates we used in our calculations every time the condition-number selection was applied. Figure 11b shows the parameters α and μ of an example basis with and without selection. One can see that the condition number gets better when larger α are preferred. This is true because more localised Gaussians enter the basis in this case, which have consequently less overlap with the other functions of the basis.

4.3.2 Stationarity of the Energy Expectation Value

We know already from Theorem 10 that the energy expectation value given by the Rayleigh quotient (37) is stationary for perturbations of the exact eigenfunction. We can try to exploit this property in order to find a good basis. For this, let us first actually calculate the derivatives of the Rayleigh quotient. Let our basis be given as (χ_{α_k,μ_k}) according to (42), where $k = 1, \ldots, N$. Let H, S be the matrices of (40), and recall in particular Subsection 4.2.1 for the form of the matrix elements of H and S. If we express an approximate wave function ϕ in this basis by
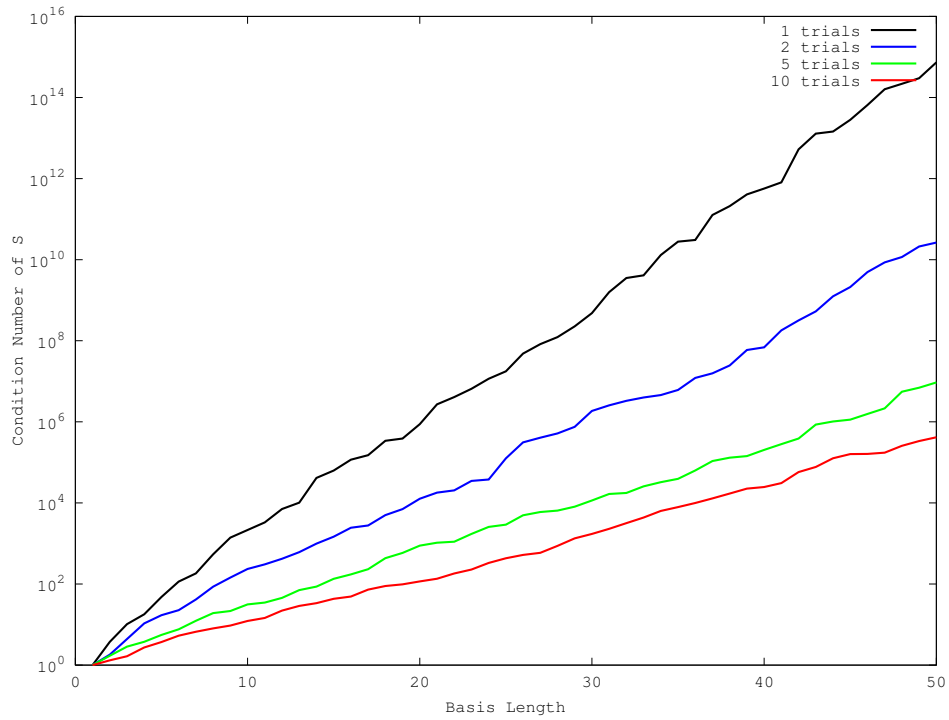
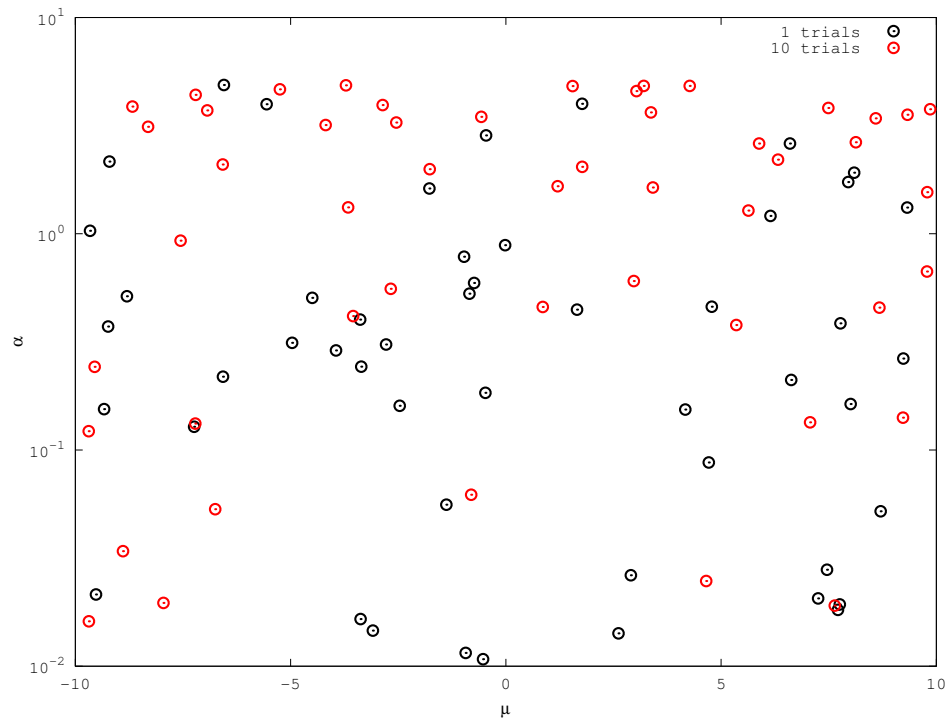$$\phi = \sum_{i=1}^{N} a_i \chi_{\alpha_i,\mu_i},$$

the Rayleigh quotient is

$$R(\phi) = \frac{\left(\phi|\hat{H}|\phi\right)}{(\phi|\phi)} = \frac{a^\top H(\alpha,\mu)a}{a^\top S(\alpha,\mu)a}.$$

---

[26]Here and in the following, we always mean the condition number in the usual Euclidean 2-norm.

(a) Condition number for growing bases. Each line is the average of 10 runs.

(b) Parameters of a representative basis.

Figure 11: Behaviour of the condition number of the matrix S for bases of random Gaussians, and for bases after applying condition-number selection as described in Subsection 4.3.1.

Note that, since the c-product is *bilinear*, we really have $a^\top$ here and *no* Hermitian conjugates! As indicated, the dependence on the non-linear parameters α_k and μ_k of the basis comes in through the matrix elements of H and S. Thus the derivatives of $R(\phi)$ with respect to these parameters are given simply by the quotient rule as

$$\frac{\partial R(\phi)}{\partial \alpha_k} = \frac{a^\top \frac{\partial H}{\partial \alpha_k} a}{a^\top S a} - R(\phi) \frac{a^\top \frac{\partial S}{\partial \alpha_k} a}{a^\top S a}. \tag{50}$$

Of course, (50) has the same form for the derivatives with respect to the μ's. Thus for (50) to be applicable, we have to calculate the derivatives of the matrix elements. For $\frac{\partial S}{\partial \alpha_k}$ and $\frac{\partial S}{\partial \mu_k}$, note that (43) can be symbolically differentiated arbitrarily often; while the expressions are lengthy, it is nevertheless easy to calculate them using a computer algebra system like Maxima [17]. The same also holds for the derivatives of the kinetic-energy parts (44) in H. For the potential parts, note that

$$V_{kl} = \left(\chi_{\alpha_k,\mu_k} | \hat{V} | \chi_{\alpha_l,\mu_l} \right) = \int_{\mathbb{R}^n} \chi_{\alpha_k,\mu_k}(x) \chi_{\alpha_l,\mu_l}(x) V(x) \, dx = \int_{\mathbb{R}^n} e^{-\alpha_k |x-\mu_k|^2 - \alpha_l |x-\mu_l|^2} V(x) \, dx.$$

From this form, it is easy to see that the only dependence on the parameters α_k and μ_k is in the exponential. Thus it can, of course, again be easily differentiated arbitrarily often and symbolically, yielding $\frac{\partial V}{\partial \alpha_k}$ and $\frac{\partial V}{\partial \mu_k}$ if we have a general-purpose quadrature routine that can numerically integrate arbitrary functions against the potential. In a practical implementation, we need such a routine anyway in order to calculate the potential matrix elements, thus also these derivatives can easily be calculated. In the end, we have everything to use (50) in order to evaluate derivatives of $R(\phi)$ with respect to the non-linear parameters.

With this in hand, we can now formulate a basis selection criterion based on Theorem 10: Define the gradient of the Rayleigh quotient with respect to all parameters,

$$\nabla R^\top = \begin{pmatrix} \frac{\partial R}{\partial \alpha_1} & \cdots & \frac{\partial R}{\partial \alpha_N} & \frac{\partial R}{\partial \mu_1} & \cdots & \frac{\partial R}{\partial \mu_N} \end{pmatrix}.$$

From all candidate basis functions, choose the one that minimises some norm of ∇R. In our implementation it was the $\|\cdot\|_\infty$ norm (i. e., the *maximal* single partial derivative that appears as a component of the gradient), but, of course, any other norm would also be suited; especially because all vector norms are equivalent anyway in the underlying finite-dimensional space $\mathbb{C}^{2n}$.

One can also make use of the stationarity provided by Theorem 10 in a stochastic way: Assuming for a moment that N basis functions are enough to represent the exact eigenfunction suitably well (such that also in this basis a stationary point is achieved at least approximately), we can consider a probability space on the set of all possible non-linear parameters (which is a subset of $\mathbb{R}^{2N}$ for the case of basis functions according to (42)). Solving (41) for a given set of parameters then induces a functional dependence between those parameters and the resulting approximate eigenvalues. In stochastic terms, the eigenvalues are *random variables* on that probability space. By Theorem 10 we know that for those parameters that lead to a good approximation of the real eigenfunction, the derivatives of this mapping vanish. This implies that the probability density for the approximate eigenvalues has a maximum at the exact eigenvalue. Thus, one could "simply" generate a lot of random bases and then choose some mean of all occurring approximate

eigenvalues as approximation to the exact eigenvalue. According to the central limit theorem one can expect that these approximations converge to the real one with a rate of $\frac{1}{\sqrt{K}}$, where K is the number of runs.[27] This is not really great convergence, and numerical results also indicate that we have indeed this or even worse a rate. Furthermore, the whole discussion above is based on the crucial assumption that the basis length N is already large enough (so that we can really approximate the exact eigenfunction well with only N basis functions). Thus, with this approach, *in addition* to lots of runs we *still* also need a large basis to begin with. All in all, this approach is not very practical. It is far more desirable to have a selection criterion which allows us to build a good basis step by step, rather than always replacing the *whole* basis for each run. This was our original idea behind the stochastic variational method, after all.

Alternatively one can, of course, vary just the last basis function and consider the resulting approximate eigenvalues for all candidates. In this case, however, the problem is that most probably this procedure does not give enough flexibility to actually represent the exact eigenfunction good enough with the existing and *fixed* $N-1$ basis functions plus the new one, such that Theorem 10 cannot be applied. Both cases are demonstrated in Figure 12: For each one, the "cloud" of approximate eigenvalues plus the exact positions are shown. The figures in the bottom of each page show a view zoomed around the third resonance. One can see nicely that in Figure 12b the exact position is really at the centre of the cloud, while in Figure 12d this is not the case.

So far we did not mention an additional difficulty with the rough idea of "trying to minimise the gradient of the energy expectation value": Namely that in order to calculate $\nabla R(\phi)$, we actually need *one* approximate wave function ϕ at which the Rayleigh quotient and also its gradient can be evaluated. However, when solving (41), one actually obtains N eigenfunctions, since it is the eigenvalue problem of an $N \times N$ matrix. Most of them are only spurious artefacts of the finite-dimensional discretisation and do not have any physical significance, but a-priori we do not know which of them really correspond to eigenvalues of the original problem. Nevertheless, in order to calculate ∇R with (50), we need *a single* coefficient vector a. Hence, we also have to decide upon which of the eigenvectors we use to apply the gradient criterion! Looking at Figure 12a, one can, however, see that the eigenvalues of the discretised problem have a rather nice structure: Spurious eigenvalues, that do not correspond to real resonances, may appear anywhere near the indicated scaling-angle line. Above it, the eigenvalues cluster nicely around the discrete resonance poles. Thus we can immediately mention at least two (heuristic) decision strategies for finding the "real" eigenvalues:

1. If we know *approximately* where a resonance is expected, we can simply choose the approximate eigenvalue that is closest to that position and use its corresponding eigenvector as a in (50). Since the resonances are clearly distinguished from each other, it is already enough to know the value up to very few digits. For our calculations where this approach was applied, we added Gaussian noise with a standard deviation of 0.1 (corresponding very roughly to a 10% perturbation) to the exact resonance position. This disturbed value was then used as the "known approximate
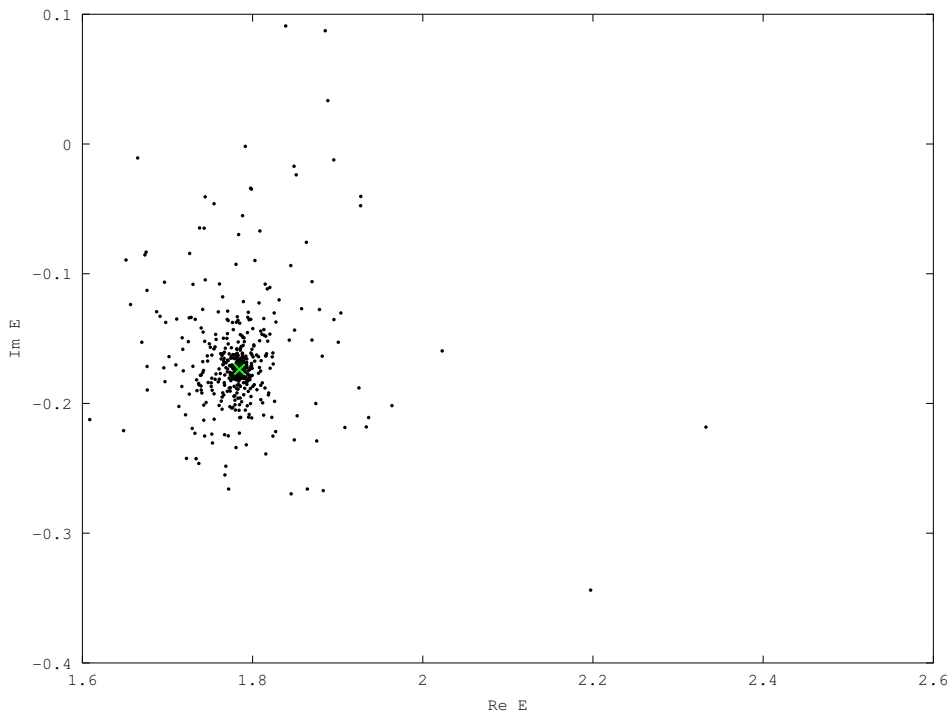
---

[27]See, for instance, chapter 7 of [21] and in particular Theorem 1.1 there. To justify this claim rigorously from the theorem, one needs that the random-basis eigenvalues form an *unbiased* estimator for the exact resonance energy. It is not clear whether this is really the case.
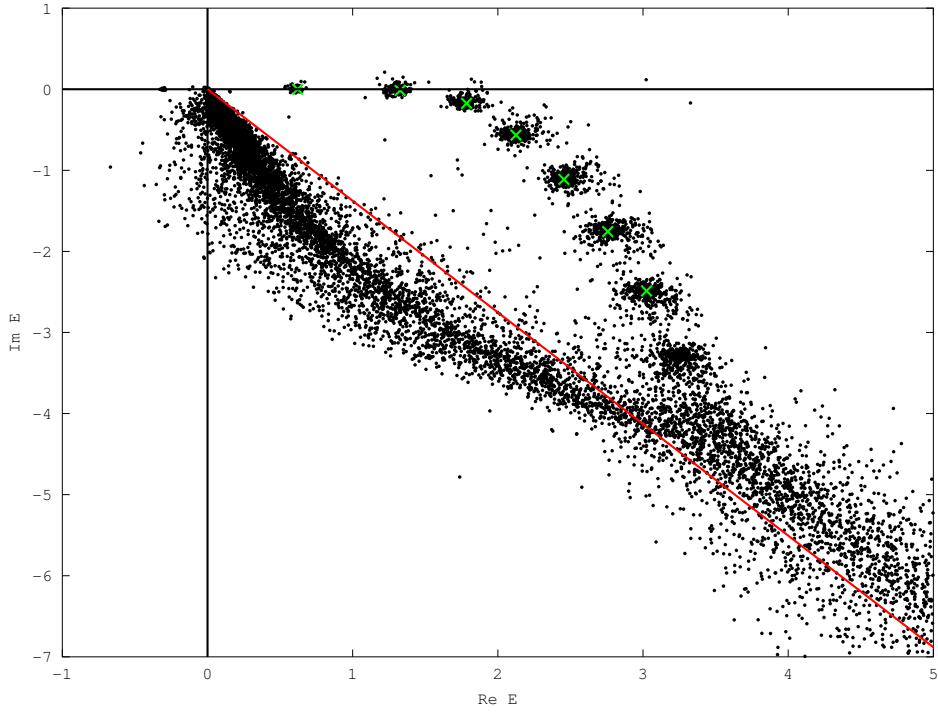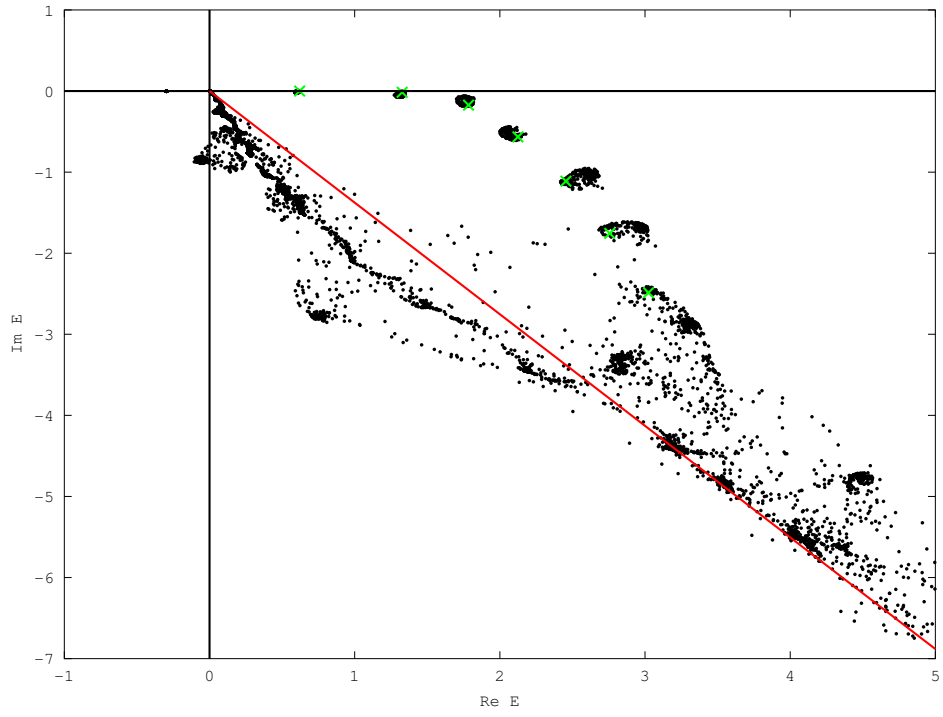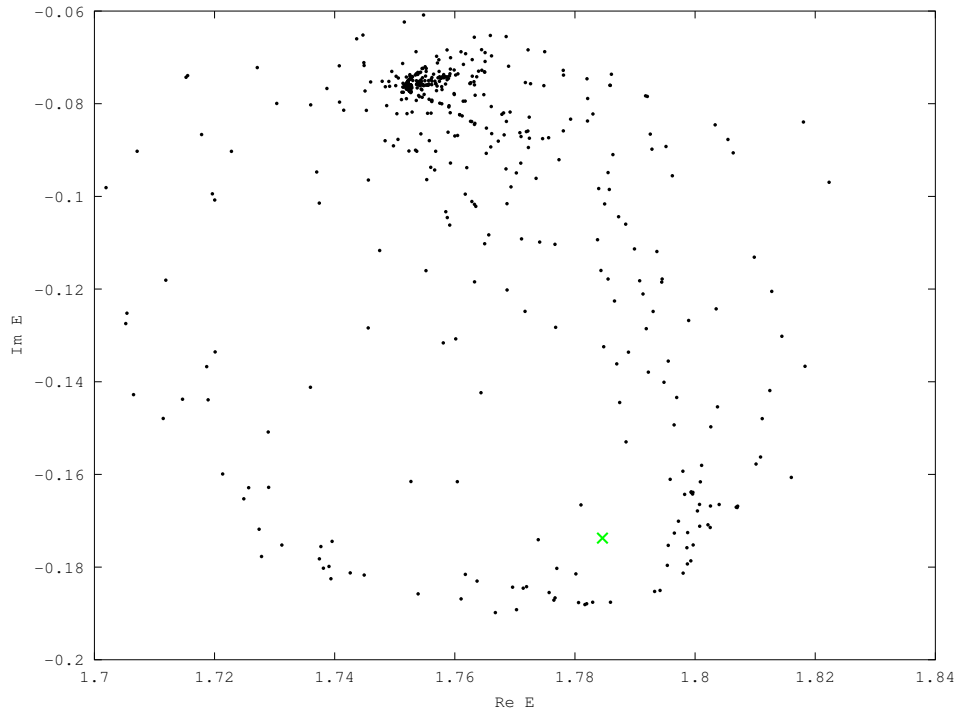
(a) Full view.

(b) Zoomed around $n = 3$ resonance.

Figure 12: Approximate eigenvalues (black dots) and exact positions of resonances (green crosses) for Example 3 with complex scaling (angle 2θ indicated by red line). $N = 40$ random basis functions are used, and 500 runs are combined. This figure shows how the eigenvalues are distributed when the whole basis is changed.

(c) Full view.

(d) Zoomed around $n = 3$ resonance.

Figure 12: Continuation of Figure 12. Here, only the last basis function is changed.

position" of the calculation. This was accurate enough. In practice, an approximate position can, for instance, be obtained via a calculation with a single, random basis or with finite differences. Such a combination of methods can be useful in order to get a rough overview over the resonance structure first, and to calculate the positions of particular eigenvalues more precisely later on.

2. It is visible (and also clear from the explanation above based on Theorem 10) that eigenvalues of different runs are by far more stable around the real resonances than in case of spuriousness. Thus one can, for instance, look at eigenvalues in the new basis of N basis functions that are close to eigenvalues of the previous step with only $N-1$ basis functions, or devise heuristics that take into account the distribution of eigenvalues for all the current candidate functions generated. Below such a heuristic (based only on the results with N and $N-1$ basis functions, not the whole set of candidates) is also used and performs quite well. Note, however, that this strategy is not so well suited if a particular resonance is of interest, because the eigenvalue determined by the heuristic approach may correspond to different real resonance positions for different steps. Another approach utilising the same basic idea of "stationarity" of the real eigenvalues in contrast to artefact ones is to solve the discrete problem with different scaling angles θ in the same basis. Also in this case, the eigenvalues corresponding to real resonances are more stable than spurious ones. See, for instance, Figure 7 in [18].

See also Subsection 4.4.3 and Figure 16 below for a comparison of these decision strategies and their influences on the overall method performance in the case of a numerical example.

4.3.3 Weight-Based Update Strategies

Before we describe the "weight-based update strategies" further, let us recall Parseval's identity (which is basically a generalised version of the Pythagorean theorem):

Theorem 12. *Let U be an inner-product space with the norm $\|\cdot\|$ generated by the inner product. If $(\tilde{\chi}_k) \subset U$ is an orthonormal set[28] and $(\tilde{a}_k) \subset \mathbb{C}$ are arbitrary coefficients, then*

$$\left\| \sum_k \tilde{a}_k \tilde{\chi}_k \right\|^2 = \sum_k |\tilde{a}_k|^2 .$$

Proof. This follows by expanding the squared norm and using orthonormality. □

This is the starting point for the following consideration: Assume that we have a basis of $N+1$ orthonormalised functions to approximate a resonance eigenfunction, say

$$\psi \approx \psi_{N+1} = \sum_{k=1}^{N+1} \tilde{a}_k \tilde{\chi}_k.$$

If we now want to reduce the basis to just N functions by removing one element, because we are looking for a small basis still giving a good approximation ψ_N, the best choice is

---

[28]Usually, one formulates this result with a basis, but that is not necessary.

to remove $\tilde{\chi}_m$ where $|\tilde{a}_m| \leq |\tilde{a}_l|$ for all $l = 1, \ldots, N + 1$. This choice minimises the error introduced by reducing the basis, because

$$\|\psi_{N+1} - \psi_N\|^2 = |\tilde{a}_m|^2$$

is minimal among all possible choices. Of course, if we have in mind that the coefficients $\tilde{a}_k$ are chosen by solving (41), then the situation is not as simple as above, where we assumed them to not be changing at all. However, it is still suggested to be a good strategy to remove the particular basis function which has the minimal coefficient or *weight* among them all. Note, however, that for this argument to be valid, it is really crucial to have an *orthonormalised basis*, because otherwise Theorem 12 is simply wrong! Thus, when deciding about the relative weights of the basis functions for not yet orthonormalised Gaussians, one has to first employ (47) to find the weights *in the corresponding orthonormal basis* and then find the minimal $\tilde{a}_k$ rather than the minimal a_k directly. According to this reasoning, we propose the following strategy for generating a good basis:

1. Starting with N already chosen basis functions, add another one using a condition-number-based selection as described above in Subsection 4.3.1 to get a basis with $N + 1$ elements.

2. From *all* basis functions (including the "old" N ones and not only the new), find the one with minimum weight $\tilde{a}_k$ as described above. Replace it by a new basis function, again selecting the new one based on the resulting condition number of S. Repeat this procedure several times in order to adapt the basis to the problem at hand *without increasing its size*.

3. Continue with the first step to gradually grow the basis.

There are two important remarks we have to make about this procedure: First, as opposed to just selecting the best *new* candidate, it also allows to change old basis functions—this gives more flexibility, as one is not "locked in forever" with basis functions once they are selected. This is, however, still a gradual build-up rather than selecting completely new bases every step as discussed above on page 48. Second, we found it to be practically important to employ condition-number selection as a sub-process, because otherwise as mentioned in Subsection 4.3.1 it becomes impossible to numerically perform the Cholesky factorisation of S necessary to find the "real" weight of a basis function. Even with it, though, at some point the orthonormalisation still becomes instable as can be seen from the green curve in Figure 13b (see also the explanations in Subsection 4.4.1).

4.4 Numerical Results

In this section, we finally want to compare the methods described above to each other based on how fast (or slow) they approach the correct resonance energies with a growing number of basis functions. As our toy problem, Example 5 is used again, with the parameters chosen as $J = 0.8$, $\lambda = 0.1$, and $\theta = 0.15\pi$. Convergence is measured with respect to the $n = 3$ resonance energy, where $E_3 \approx 1.7846 - 0.1738i$. We use the absolute error in the eigenvalue as our measure of convergence, i. e., $\left|E_3 - \tilde{E}_3\right|$, where $E_3, \tilde{E}_3 \in \mathbb{C}$ are the exact and approximate resonance eigenvalues, respectively. The value

we calculated and used as comparison agrees with the one given in [9], at least up to the four digits of accuracy given there.[29] It was calculated by the gradient method (see the list below), in a single run, with $N = 100$ basis functions. This should be large enough a basis for this purpose, since we will see in Figure 13b (blue curve) that the maximum precision possible due to numerical issues is already reached around $N \approx 80$. Since the variational methods to be examined here are stochastic by design, the concrete results depend on the actual run. In order to compensate for that, the curves shown and discussed below will be averages over multiple runs (namely, 1,000 to be precise), where the average is taken of the logarithm of the error (since the plots are also logarithmic on the y-axis). We always use Gaussian basis functions according to Subsection 4.2.1, and will compare these methods below:

Random For this method, no actual basis selection is performed. Instead, the non-linear parameters of the basis functions are chosen purely randomly, and only the linear parameters (i. e., expansion coefficients in this basis) are varied. This helps to benchmark the performance of other methods.

Condition Number The selection of candidate basis functions is *only* based on the condition number of S as described in Subsection 4.3.1.

Error Minimisation We choose at each step that candidate basis function, which results in an approximate eigenvalue that is closest to the exact resonance position. This is of course "cheating" because one needs to know the result already (and thus this method is not applicable for a practical calculation), but it is nevertheless interesting to consider this method here.

Gradient The selection criterion based on minimisation of the gradient norm, as described in Subsection 4.3.2.

Weight-Based with ONB The weight-based update strategy (see Subsection 4.3.3), together with orthonormalisation of the basis. Furthermore, additional selection steps are used to reduce the condition number and improve the numerical stability. See also the algorithm outline given on page 52.

Besides the actual comparisons of these methods in Subsection 4.4.1, we will also demonstrate and discuss some more subtle issues related to some details of the methods in Subsection 4.4.2 and Subsection 4.4.3. In Subsection 4.4.4, we will finally analyse how the stochastic variational method compares to a finite-difference method, which is completely different in spirit. The numerical calculations were performed with GNU Octave 3.6.3 [22] on a GNU/Linux system of the `amd64` architecture. For the gradient calculations, we used both symbolic expressions derived with Maxima [17] as well as automatic differentiation with ADOL-C [23], both giving consistent results. See also [24] for more details on this.

---

[29]Note that (as was already mentioned on page 17) the potential in [9] has an additional offset of J, which has to be taken into account when comparing our results with the values there.
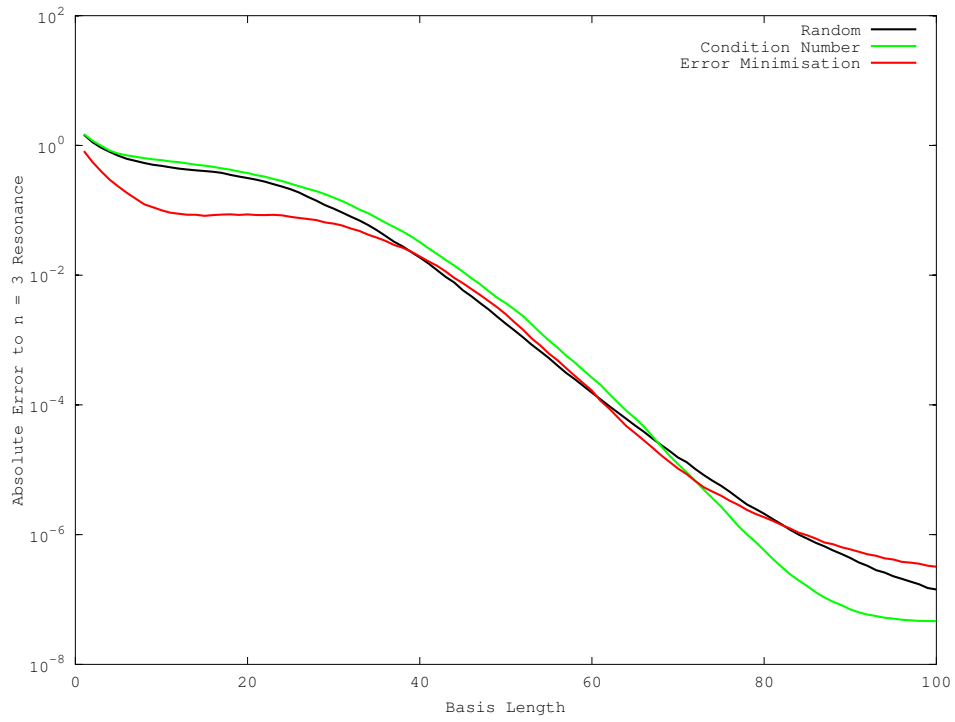
4.4.1 Performance of the Stochastic Variational Methods

Let us start directly with the main topic, namely the comparison of the performance of different stochastic variational methods to each other. How the error decreases for increasing basis size is shown in Figure 13: Some rather basic methods are compared to each other on top in Figure 13a, and below in Figure 13b the better methods we suggest for a practical calculation are shown. The black curve in both plots shows the random basis-selection method, so that other methods can be judged by it. As can be seen in Figure 13a from the green curve, the condition-number-based selection does not perform any better. The red curve, showing minimisation of the error, is also interesting: At the beginning, it behaves very well as might be expected "by design", but interestingly enough, later on this method is not better than random selection at all, even though it may be intuitive to assume it should be the best selection possible. This is probably due to the fact that a basis function which is "best" at some early step may well be worse later on. Without the possibility to replace functions already selected (see page 52), they may deteriorate the overall performance in subsequent steps. Especially, since basis functions are chosen only to satisfy a small error and not by some more general criterion (for instance, a small gradient as suggested by Theorem 10), this may lead to "overfitting" in later iterations.

Methods that actually show an improvement over random selection are shown in Figure 13b. In particular, the best method for the basis lengths tested seems to be the weight-based update, whose convergence is shown by the green curve in Figure 13b. It can be seen, however, that above a basis size of $N \approx 70$ the maximal accuracy is no longer achieved by this method. This might be due to the fact that in this region even with the condition-number selection steps the condition number of S grows so much, that the orthonormalisation becomes instable, wherefore the weight-based update also does not work well any more. The maximal precision of about 10^{-7} achieved with a basis size of $N \approx 70$ is probably already more than enough for most applications, though. The blue curve shows how the gradient method performs. Compare here also Figure 14, which shows how the gradient norm itself decreases accordingly and in good correlation to the decrease in the absolute error of the eigenvalue. After a certain initial basis size is reached at around $N = 30$, it shows the steepest descent of all methods tried. This might be due to the fact that it has the best theoretical justification (based on Theorem 10). The initially needed basis size may be explained by the assumption (see page 39) that we can actually express the exact wave function already with our basis. This assumption is necessary for Theorem 10, but only fulfilled accurately enough if the basis size is not too small. It can be seen that also this method "flattens out" for a large basis, but this is very probably due to the fact that at this point the numerical eigenvalue calculation itself is already the limiting factor in precision. Furthermore, also the "exact" comparison value (see the explanation at the beginning of this section on page 53) is, of course, not without inaccuracies, which is another limit from below for the shown error.

4.4.2 Different Weight-Based Methods

In particular for weight-based methods there are multiple things to consider in order to get them "right" as mentioned in Subsection 4.3.3. The influence of those is shown in Figure 15: There, the red curve shows the most naive implementation—not employing

(a) Methods with the same performance as random selection.
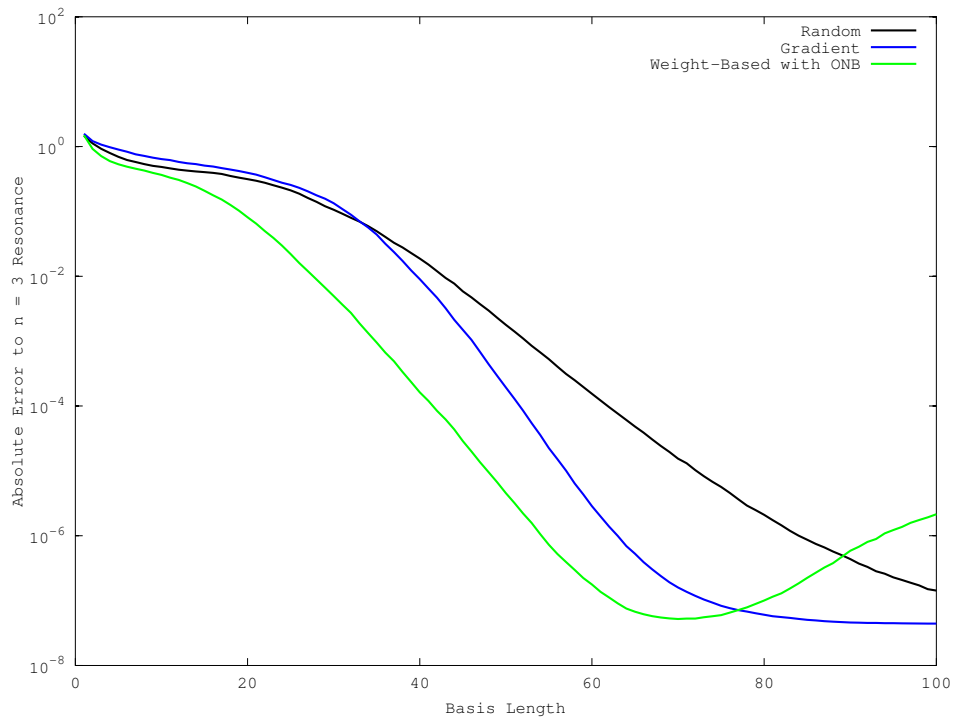
(b) Methods that outperform random selection.

Figure 13: Convergence of the absolute error for different selection criteria in the stochastic variational method.
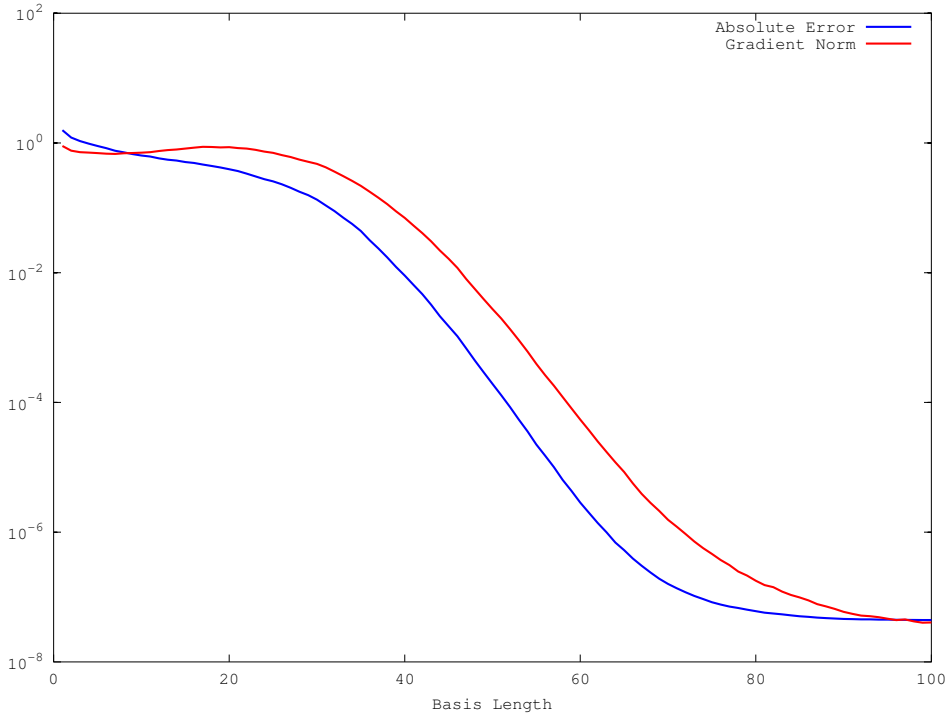
Figure 14: Convergence of the gradient method in absolute error (blue) in comparison to the decrease of the gradient norm (red).

orthonormalisation, but instead using the coefficients in the original, non-orthogonal basis (although the individual basis functions are, of course, normalised). It works at the beginning, when there is not yet much overlap, but it quickly fails for larger basis sizes. The blue curve shows what happens if still no real orthonormalisation is applied, but the condition-number criterion is used. It works surprisingly well up to a certain point, but then also fails. The initial success is due to the fact that by selecting basis functions in order to minimise the condition number of S, functions with not much overlap are preferred. This leads to "almost orthogonal" basis functions, and is seemingly good enough for smaller basis sizes. The green curve is the same as the green curve in Figure 13b. It does real orthonormalisation as it should be. The black curve is again the same one as before in Figure 13 (random selection).

4.4.3 The Problem of Eigenvalue Selection

As mentioned already on page 48, both the weight-based and the gradient methods rely on a decision, which one of the discrete eigenvalues one should use for calculating the basis weights or the gradient norm. In practice this seems not to be a big problem, though, since already a rough guess or some heuristic approach (as discussed there) gives good results. This is confirmed by Figure 16. The blue and red curves correspond to the gradient method, where for the blue curve (as before) the exact value with 10% of Gaussian noise was used in the calculation, while for the red one the exact value itself was used. They have virtually no difference, since it is already enough to know the value up to this precision in order to decide on the correct eigenvalue in each step. Green is the
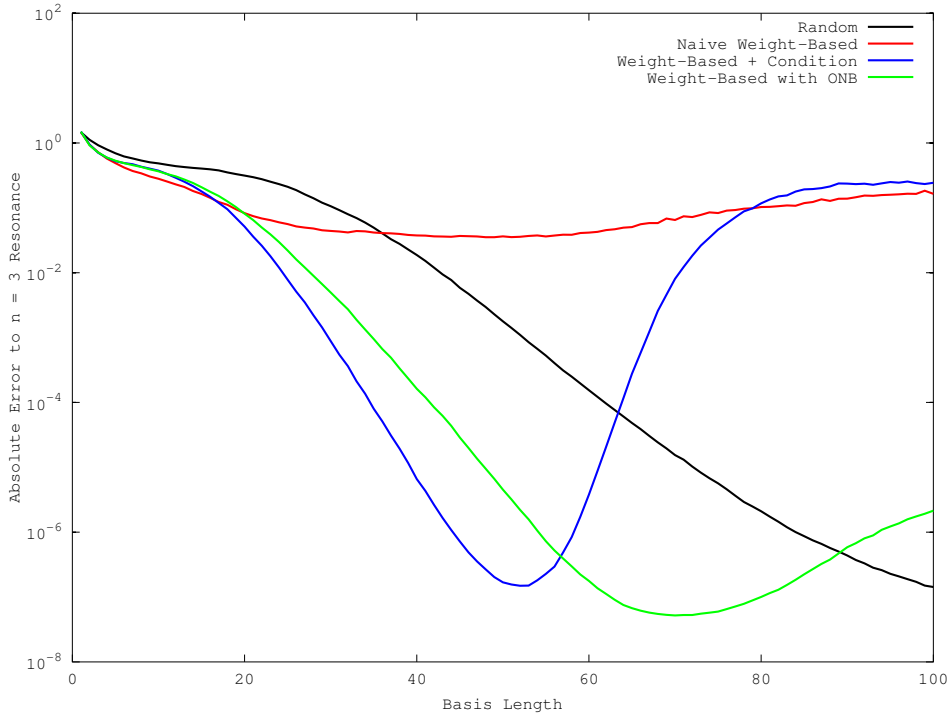
Figure 15: Effects that different "ingredients" to be considered for the weight-based method have on its convergence rate, see also Subsection 4.4.2.

weight-based curve shown already in other plots; it is also based on the exact value with 10% of noise. The magenta curve instead uses the heuristic approach described on page 51 based on assumed stationarity of the "real" eigenvalues with respect to multiple runs. Since this strategy decides not always on the $n = 3$ resonance, it is a little worse at the beginning, but interestingly it also does better in the end. The reason for this behaviour is probably that because of the switches, a basis is chosen that is "more independent" and thus slightly better when the calculations tend to become instable.

4.4.4 Stochastic Variational Methods versus Finite Differences

An alternative and more straight-forward approach as compared to variational methods is a finite-difference discretisation of the Hamiltonian. This was discussed already briefly in the introducing paragraph on page 34. Here we want to compare convergence results for finite differences to those already shown above for the stochastic variational methods. For the calculation, we used the same potential and scaling parameters as before, and discretised the space interval $x \in [-10, 10]$ into uniform subintervals. The Laplacian was discretised[30] with the standard second-order central-differences stencil, which gives an accuracy of $O(h^2)$ in the step-size h and results in a tri-diagonal matrix for the one-dimensional case (see, for instance, section 9.2 of [20]). The convergence to the $n = 3$ resonance is shown in Figure 17. It can be seen that the error follows precisely the expected power-law of h^2, and that around 100,000 grid-points (corresponding to the

---

[30]See also Example 5 above for more explanations about the finite-difference discretisation employed.
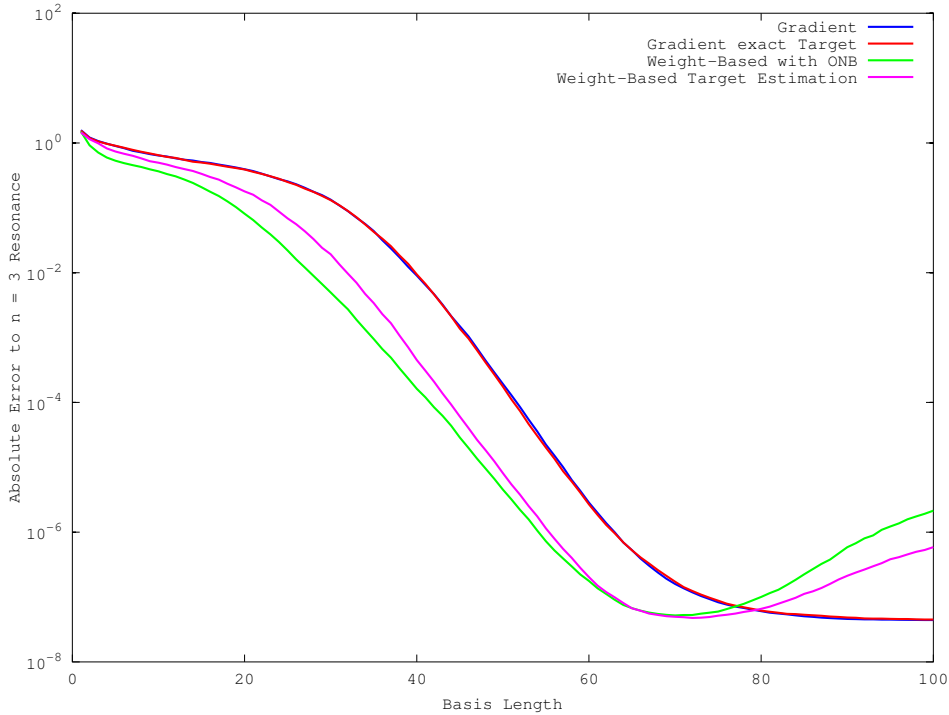
Figure 16: The influence that the eigenvalue selection strategy has on the weight-based and the gradient methods as discussed in Subsection 4.4.3.

interval $[-10, 10]$ divided into pieces of length $h = 2 \cdot 10^{-4}$ each) would be necessary to achieve the maximum accuracy of 10^{-8}, which the stochastic variational method yields. Our actual calculations were only done up to 5,000 subintervals, since that was already quite expensive.[31] With the stochastic variational method much smaller matrices are enough, with around 80 basis functions already giving the same accuracy as a finite-difference discretisation with 100,000 grid-points. However, it must not be forgotten that the finite-difference matrices are sparse, while the stochastic variational method produces dense matrices. In addition to solving the matrix eigenvalue problem, it is also quite costly to generate the stochastic basis for the latter; in fact, this is the most expensive part of the calculation. The finite-difference matrix, on the other hand, can be built readily. Nevertheless, note that two more orders of magnitude in the desired accuracy require increasing the number of grid-points by a factor of ten for finite-differences, while they require only *adding* 10 more basis functions in the stochastic variational method (when, for instance, the blue curve (gradient method) in Figure 13b is considered).

---

[31]Note that the calculation can surely be optimised further, for instance by using vector iteration (see in particular Definition 13.38 in [20]). If (similarly to what was assumed for the stochastic variational method) already an approximate eigenvalue is known, this method can be used to efficiently calculate its position more precisely.
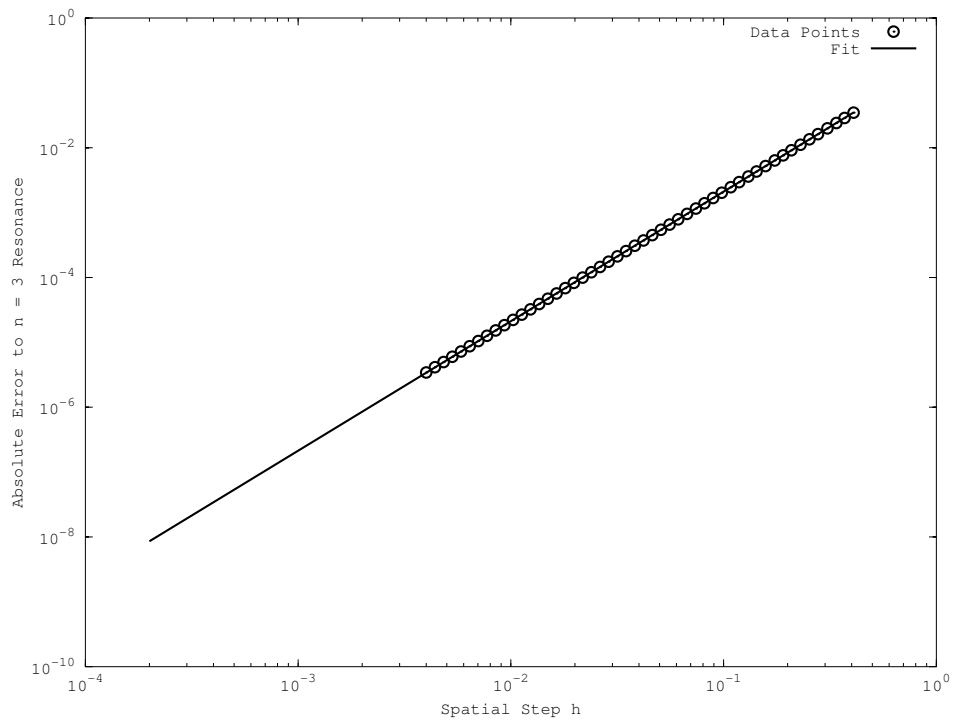
Figure 17: Convergence of a finite-difference approximation with actual results (circles) and the power-law extrapolation (line).

5 Conclusion

The main goal of the present work was the extension of the stochastic variational method to allow the treatment of resonances, which are associated with complex eigenvalues of quantum-mechanical systems. In this case, one is usually confronted with non-Hermitian Hamiltonians that arise, for instance, with optical potentials or after similarity transformations, such as the complex scaling or the Zel'dovich transformations. We based our considerations as far as possible on a rigorous mathematical foundation, since a lot of work on resonances in the literature is of predominantly practical nature and often lacks the desired mathematical rigour.

From the short introduction into the general theory of resonances in Chapter 2 it became particularly clear that resonance wave functions pose a difficulty within the usual Hilbert-space formulation of quantum mechanics, since this formulation is based on spaces of square-integrable functions (such as $L^2(\mathbb{R}^n)$), which do not contain the resonance wave functions of interest to us. Therefore, methods commonly used in standard quantum mechanics are no longer appropriate. In order to circumvent this problem, we resorted to similarity transformations such as those described in Chapter 3. They allowed us to transform the eigenvalue problem for resonances into an eigenvalue problem of a non-Hermitian operator. Evidently, its eigenvalues are in general complex, but its eigenfunctions are square-integrable and can thus be handled again in the usual mathematical framework.

The stochastic variational method has proven to be quite powerful for Hermitian problems. In this case it is based on the well-known Rayleigh-Ritz variational principle. Obviously, the latter cannot be applied for complex eigenvalues. Therefore it was necessary to derive a generalised variational principle that avoids minimisation and is based on *stationarity* instead. It states that the best values for linear and non-linear parameters in a parametrisation of the wave function are near a stationary point of the energy expectation value. Consequently, the gradient with respect to those parameters has to vanish when the correct eigenvalue is attained. Precise formulations and corresponding proofs of this result are given in Theorem 9 and Theorem 10 of Chapter 4. Based on this generalisation of the variational principle, also the criterion for selecting good candidate basis functions in the stochastic variational method had to be adapted. In addition to the stochastic selection of basis functions, we also designed a further criterion for improving the whole basis throughout the calculation. It consists in replacing basis functions with small expansion coefficients by more promising ones. As a result, we arrived at two particular variants of the stochastic variational method that can be applied to non-Hermitian problems. Both turned out to perform significantly better than purely random basis selection or a straight-forward finite-difference discretisation of the Hamiltonian.

Having arrived at these results, still a number of problems remain open with regard to the description of quantum-mechanical resonances. In particular, the following items could be the topic of further research:

- Theorem 5 is unfortunately not enough to fully justify the application of complex scaling to resonance wave functions, since it does not show analyticity of the eigenfunctions. It would be interesting to consider whether such a proof is possible. It could possibly incorporate deeper knowledge about resonance wave functions from scattering theory.

- As discussed during the treatment of Example 4, we were not able to use the Zel'dovich transformation for any meaningful practical results. Since the main focus of this work was complex scaling, we did not investigate further why this was the case and whether the problems can be fixed so that also the Zel'dovich transformation can be applied successfully.

- For simplicity and in order not to detract the focus from our work on variational methods in the complex case, we did not try to apply exterior scaling or any other more sophisticated variant. While that would be more complicated, it is probably only a matter of a sufficiently diligent implementation to apply these techniques together with our variational methods, for instance, to the square-well potential.

- It would be interesting to study problems of non-Hermitian Hamiltonians containing (complex) optical potentials (such as the ones resulting from Feshbach elimination in multi-channel systems) from the same point of view as adopted in the present work.

- Finally, it is certainly demanding to extend our methods and their implementations to higher-dimensional situations as well as to real-world problems of atomic, nuclear and particle physics.

References

[1] E. Prugovečki. *Quantum Mechanics in Hilbert Space.* Academic Press, New York, second edition, 1981.

[2] Y. Suzuki and K. Varga. *Stochastic Variational Approach to Quantum-Mechanical Few-Body Problems.* Springer, Berlin, 1998.

[3] N. Moiseyev. *Non-Hermitian Quantum Mechanics.* Cambridge University Press, Cambridge, 2011.

[4] J. J. Sakurai. *Modern Quantum Mechanics.* Addison-Wesley, revised edition, 1994.

[5] J. R. Taylor. *Scattering Theory: The Quantum Theory on Nonrelativistic Collisions.* John Wiley & Sons, New York, 1972.

[6] H. Feshbach. Unified Theory of Nuclear Reactions. *Ann. Phys. (N.Y.)*, 5:357–390, 1958.

[7] T. Hippchen, K. Holinde, and W. Plessas. Bonn Meson-Exchange Potential in the $N - \overline{N}$ System. *Phys. Rev. C*, 39(3):761–765, 1989.

[8] S. Flügge. *Rechenmethoden der Quantentheorie.* Springer, Berlin, fifth edition, 1993.

[9] H. J. Korsch, H. Laurent, and R. Möhlenkamp. Comment on "Weyl's Theory and the Complex-Rotation Method Applied to Phenomena Associated with a Continuous Spectrum". *Phys. Rev. A*, 26(3):1802–1803, 1982.

[10] L. C. Evans. *Partial Differential Equations*, volume 19 of *Graduate Studies in Mathematics.* American Mathematical Society, 1998.

[11] J. Nuttall and H. L. Cohen. Method of Complex Coordinates for Three-Body Calculations above the Breakup Threshold. *Phys. Rev.*, 188(4):1542–1544, 1969.

[12] E. Balslev and J. M. Combes. Spectral Properties of Many-Body Schrödinger Operators with Dilatation-Analytic Interactions. *Comm. Math. Phys.*, 22:280–294, 1971.

[13] J. Barros-Neto and R. A. Artino. *Hypoelliptic Boundary-Value Problems*, volume 53 of *Lecture Notes in Pure and Applied Mathematics.* Marcel Dekker, New York, 1980.

[14] M. G. Crandall. Viscosity Solutions: A Primer. In *Viscosity Solutions and Applications*, volume 1660 of *Lecture Notes in Mathematics.* Springer, Berlin, 1997.

[15] M. G. Crandall, H. Ishii, and P.-L. Lions. User's Guide to Viscosity Solutions of Second-Order Partial Differential Equations. *Bull. Am. Math. Soc.*, 27(1):1–67, 1992.

[16] J. Franklin. Analytic Continuation by the Fast Fourier Transform. *SIAM J. Sci. Stat. Comp.*, 11(1):112–122, 1990.

[17] Maxima: A Computer Algebra System. `http://maxima.sourceforge.net/`.

[18] N. Moiseyev, P. R. Certain, and F. Weinhold. Resonance Properties of Complex-Rotated Hamiltonians. *Mol. Phys.*, 36(6):1613–1630, 1978.

[19] K. Varga and Y. Suzuki. Precise Solution of Few-Body Problems with the Stochastic Variational Method on a Correlated Gaussian Basis. *Phys. Rev. C*, 52(6):2885–2905, 1995.

[20] R. Plato. *Concise Numerical Mathematics*, volume 57 of *Graduate Studies in Mathematics*. American Mathematical Society, 2003.

[21] A. Gut. *Probability: A Graduate Course*. Springer Texts in Statistics. Springer, New York, second edition, 2013.

[22] GNU Octave. `https://www.gnu.org/software/octave/`.

[23] A. Walther and A. Griewank. Getting Started with ADOL-C. In *Combinatorial Scientific Computing*, pages 181–202. Chapman-Hall CRC Computational Science, 2012.

[24] D. Kraft. The Variational Method for Non-Hermitian Quantum Mechanics: Automatic Differentiation with Complex Numbers. `http://www.domob.eu/research/AutoDiffComplex.pdf`, 2013.