

Dynamics of Driven Quantum Systems: A Search for Parallel Algorithms

Mehrdad Baghery

Geboren am 10.02.1988 in Rafsanjan, Iran

Dissertation

zur Erlangung des akademischen Grades

Doctor rerum naturalium (Dr. rer. nat.)

Eingereicht am 31.05.2017

Verteidigt am 24.11.2017

Abstract

This thesis explores the possibility of using parallel algorithms to calculate the dynamics of driven quantum systems prevalent in atomic physics. In this process, new as well as existing algorithms are considered.

The thesis is split into three parts. In the first part an attempt is made to develop a new formalism of the time dependent Schrödinger equation (TDSE) in the hope that the new formalism could lead to a parallel algorithm. The TDSE is written as an eigenvalue problem, the ground state of which represents the solution to the original TDSE. Even though mathematically sound and correct, it turns out the ground state of this eigenvalue problem cannot be easily found numerically, rendering the original hope a false one.

In the second part we borrow a Bayesian global optimisation method from the machine learning community in an effort to find the optimum conditions in different systems quicker than textbook optimisation algorithms. This algorithm is specifically designed to find the optimum of expensive functions, and is used in this thesis to 1. maximise the electron yield of hydrogen, 2. maximise the asymmetry in the photo-electron angular distribution of hydrogen, 3. maximise the higher harmonic generation yield within a certain frequency range, 4. generate short pulses via combining higher harmonics generated by hydrogen.

In the last part, the phenomenon of dynamic interference (temporal equivalent of the double-slit experiment) is discussed. The necessary conditions are derived from first principles and it is shown where some of the previous analytical and numerical studies have gone wrong; it turns out the choice of gauge plays a crucial role. Furthermore, a number of different scenarios are presented where interference in the photo-electron spectrum is expected to occur.

Declaration/Versicherung

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Die Arbeit wurde am Max-Planck-Institut für Physik komplexer Systeme in der Abteilung “Endliche Systeme” angefertigt und von Prof. Dr. Jan-Michael Rost betreut.

Ich erkenne die Promotionsordnung der Fakultät Mathematik und Naturwissenschaften der Technischen Universität Dresden vom 23.02.2011 an.

Datum

Unterschrift

Acknowledgements

The last four years have been a very influential period of my life. I believe I have developed as a person in ways that I would never have expected. And for this I am most thankful to my colleagues and friends.

I am deeply indebted to Jan-Michael Rost for putting his trust in me, and giving me the opportunity to take advantage of a new environment on the scientific as well as the social front.

I probably cannot thank Ulf Saalman enough for putting up with me for such a long time. I definitely consider him a friend before I call him my supervisor. I would also like to thank Thomas Pohl for many hours of mostly non-scientific discussion, and for all the joy and life he brought to our lunch group.

It wouldn't have been nearly as great a time without my close friends Colm Murray and Callum Mulhern. You guys definitely took my social life to a whole new level. On a side note, I don't even remember any more how many times I had to explain to newcomers how to distinguish between your names. My Christmas holiday in England and Ireland was by far one of the best trips I have ever taken. The memories of our trips to Marktredwitz, Innsbruck, and so many others will always stay with me.

My Canadian friend, Dylan Rose, belongs to the same group; a simple country boy with whom I had an unforgettable time. The story of muffins is definitely one I will tell my grandchildren.

And of course this list wouldn't be complete without Karsten Leonhardt (aka Wet Hand), who was an integral part of our group. Our trip to Berlin for the 2014 world cup final was one of our highlights.

My favourite Spanish couple, Anna Deluca & Daniel Viscor (aka Don Daniel), you guys are just amazing the way you are, and I hope you are doing well wherever

you are.

And last but not least, Younes Javanmard. I'm happy you were around, and it was definitely comforting to get together every now and then and talk out the frustration we both had once experienced.

Contents

Abstract	iii
Declaration	iv
Acknowledgements	v
1. Introduction	1
2. McLachlian method: TDSE as an eigenvalue problem	4
2.1. Imaginary time Schrödinger equation	6
2.2. Full Configuration Interaction Quantum Monte Carlo	7
2.3. McLachlian operator: definition	12
2.3.1. Proof of hermiticity of $\widehat{i\partial_t}$	15
2.3.2. Non-zero eigenvalues of the McLachlian	15
2.3.3. Basis required for the McLachlian	16
2.4. McLachlian operator: implementation	17
2.4.1. Initial conditions	20
2.5. Ground state of the McLachlian: direct diagonalisation	21
2.5.1. Two-level system	21
2.5.2. Three-level system	22
2.5.3. Atomic hydrogen in a time dependent electric field	23
2.6. Ground state of the McLachlian: FCIQMC	24
2.7. Analysis of the McLachlian eigenvalue problem	30
2.7.1. Complexity	30
2.7.2. Parallelisability	31
2.8. McLachlian as a time propagation scheme	31
2.9. Conclusion	35
3. Optimisation of complex systems using Gaussian processes	36
3.1. Probability: the basics	39

3.2.	Bayes' theorem as a guide to learning	41
3.3.	Fully Bayesian inference	43
3.3.1.	Inferring the influence of uncertainty in measurement	44
3.3.2.	Inferring the hyperparameters	45
3.3.3.	Inferring the covariance function	46
3.3.4.	Model selection	46
3.4.	Gaussian distribution	47
3.4.1.	Conditional and marginal distributions	48
3.5.	Gaussian processes	49
3.5.1.	Conditional and marginal distributions	50
3.6.	Gaussian processes as a Bayesian model	51
3.6.1.	Inferring the covariance function	51
3.6.2.	Inferring the hyperparameters	52
3.6.3.	Inferring the influence of uncertainty in measurement	57
3.7.	Heuristic methods of optimisation	59
3.7.1.	Upper Confidence Bound (UCB)	61
3.7.2.	Machine Learning Online Optimisation (MLOO)	62
3.7.3.	Probability of Improvement (PI)	63
3.7.4.	Expected Improvement (EI)	64
3.7.5.	One-Step Lookahead	64
3.8.	General procedure of optimisation	65
3.9.	Maximising the electron yield	69
3.9.1.	Pulse envelope	70
3.9.2.	Pulse energy	71
3.9.3.	Target function	72
3.9.4.	Procedure & Results	72
3.9.5.	Analysis	74
3.10.	Maximising the asymmetry in PAD	76
3.10.1.	Target function	79
3.10.2.	Procedure & Results	80
3.10.3.	Analysis	82
3.11.	Maximising the HHG yield	82
3.11.1.	Target function	84
3.11.2.	Procedure & Results	85
3.11.3.	Analysis	85

3.12. Generating short pulses via combining higher harmonics	91
3.12.1. Pulse composition	92
3.12.2. Pulse energy	92
3.12.3. Target function	93
3.12.4. Procedure & Results	94
3.12.5. Analysis	95
3.13. Conclusion	96
3.14. Outlook	98
4. Dynamic interference	101
4.1. Interaction picture	104
4.2. Perturbation theory	104
4.3. Method of essential states	106
4.4. Stark shift of the ground and continuum states	109
4.5. Choice of gauge in atomic physics	111
4.6. Notion of dynamic interference: analytics	113
4.7. Notion of dynamic interference: numerics	118
4.8. Necessary conditions for dynamic interference	123
4.9. Other scenarios for interference in PES	125
4.9.1. Double pulse	125
4.9.2. Stabilisation	127
4.9.3. Chirped pulse	129
4.9.4. Ionisation from an excited state	130
4.9.5. Resonance frequency	134
4.10. Conclusion	138
A. Atomic units	140
B. Finite difference method	142
B.1. Basics	142
B.2. Boundary conditions	145
C. Further details regarding Gaussian processes	152
C.1. Derivatives of $m^*(\vec{x})$ and $k^*(\vec{x}, \vec{x})$	152
C.2. Squared-exponential covariance function	153
C.3. Derivative of $\Lambda_1(\vec{x})$	154

D. Fourier transform	155
D.1. Discrete Fourier transform	156
D.2. Continuous limit of DFT	156
D.2.1. Fourier series	157
D.2.2. Fourier transform	158
D.3. Parseval's theorem	158
D.4. Convolution theorem	158
D.5. Windowed Fourier transform	159
D.6. Gabor transform	161
E. Laser pulses: essentials	163
E.1. Realistic laser pulses	163
E.2. Energy density and intensity of laser pulses	164
E.3. Energy density and intensity in the slowly-varying envelope regime .	165
E.4. Spectral and temporal widths of pulses	167
E.5. High harmonic generation: inhomogeneous wave equation	167
F. Numerical propagation	170
G. Four-level system	172
References	175

1. Introduction

Physics has evolved exceptionally over the past century, so much so that the tools required today are very different from those needed not so long ago. While all that was available to a 20th-century theoretical physicist might have been a pen and paper, today we regularly take advantage of fast computers, the Internet, instant access to journals from all over the globe, Skype calls, etc.¹ Computers have revolutionised our approach to understanding the world surrounding us so massively that it wouldn't be an exaggeration to say that they have helped us tackle problems we would never have dreamed of without them.

In particular, we have been able to approach systems of unprecedented complexity. For instance, we can now predict the behaviour of large groups of interacting quantum particles with high accuracy, something that would have been considered impossible half a century ago. The branch of physics that deals with such complicated systems is known as *complex systems* [1], and as its name may suggest, it is concerned with systems showing rich and complex behaviour which is hard (if not impossible) to predict using only the knowledge of the behaviour of isolated individual constituents. Clusters are a classic example of complex systems; while a small cluster of only argon atoms may show a trivial behaviour, a large cluster of a mixture of argon and hydrogen atoms behaves completely differently [2].

Throughout the second half of the twentieth century, engineers pushed for CPUs with higher and higher clock speeds in an effort to achieve higher computational power. But in recent years this trend seems to have stagnated [3] (at the time of this thesis a standard CPU has a clock speed of around 3 GHz). The main obstacle appears to be the large energy consumption of higher speed architectures which are then followed by complicated cooling problems. Even though one way of overcoming this problem might be to use more complicated cooling techniques, en-

1. We cannot escape the fact that in 20 years from now all these currently exciting developments will sound just as primitive as the tools from the 19th century sound to us now.

gineers have successfully alleviated the problem by simply stacking multiple slower CPUs in a single unit; such units are known as multi-core processors. It turns out a dual-core CPU with a speed clock of 3 GHz consumes considerably less energy than a single-core CPU with a clock speed of 6 GHz. But the downside is that the key to computational speed-up in this case is to have programs intricately designed to take advantage of this parallel architecture.

If done correctly, parallel programming can drive us towards more computational power and therefore tackling bigger and more complicated problems. It cannot be stressed enough that having parallel algorithms has never been so crucial before. It might not have been necessary to write a parallel programme 40 years ago, as the luckiest of universities had only access to at most one computer, whereas in today's world even the slowest smartphones have most likely multi-core processors. This thesis is an attempt at finding faster and potentially parallel algorithms. Even though these algorithms might be used for a wide variety of problems, we only take as examples systems from the field of theoretical atomic physics as this is the background of our research group. We will take two different approaches here.

The first approach appearing in chapter 2 is an effort to develop a more efficient algorithm to solve the time dependent Schrödinger equation (TDSE) on a parallel architecture. We reformulate the TDSE as an eigenvalue problem in the hope of using efficient parallel eigenvalue solvers, in particular the FCIQMC method [4]. This attempt is met by an unfortunate consequence causing the FCIQMC method (or any other diffusion method for that matter) to fail to solve highly quasi-degenerate eigenvalue problems. Later in the chapter, an analysis of the new eigenvalue formulation of the TDSE is given, proving that in order for this new approach to be computationally more efficient than textbook time-propagation methods, an eigenvalue solver with complexity less than that of vector-matrix multiplication is required. Such an algorithm would be revolutionary. While working on this project, it came to our attention that in a then recently published paper [5] it was claimed that it was possible to solve the new TDSE formalism on several processors in parallel. We end the chapter by showing why such a claim is untrue. It is worth noting that despite the computational failure of this formalism, it may still be of academic interest, as it treats time and space on an equal footing.

Given the complexity of the systems of practical interest, finding the optimal settings can be time consuming and resource intensive. It is of utmost importance to find such optimal conditions in the fastest possible way, and this is the topic

of chapter 3. In the particular case of quantum physics, one such attempt has been the so-called Quantum Optimal Control Theory [6, 7]. Even though elegant, the output of this algorithm usually shows unrealistic features such as cusps in the electric component of a laser field as discussed in ref. [8]. To circumvent this problem, we will take a somewhat different route. We will borrow an algorithm originally developed in the machine learning community for the optimisation of expensive functions, and try to apply it to physical systems. Even though the original algorithm is designed for one CPU, we will make a slight modification so that multiple CPUs can be used simultaneously. All the systems considered will be expressed in terms of a few parameters which the optimisation algorithm will have permission to control in order to find the optimum values. It shouldn't be forgotten that the system should be parametrised such that the non-physical artefacts are prevented to the best of our ability. For instance, instead of allowing unconstrained pulses we can force the electric component of a laser field to be smooth. Even though this restriction might reduce the pool of available laser pulses, it is the price that we have to pay in order to get physically meaningful results. We apply this optimisation algorithm to four different systems and try in each case to make sense of the obtained results.

While working on this project, we came across a few papers whose final results we could not reproduce [9–12], an observation that left us intrigued. We consequently started investigating what might have gone wrong in the original papers. The answer turns out to be more complicated than anticipated, and this is the story of chapter 4, where it is discussed how dynamic interference (the temporal equivalent of the double-slit experiment) can happen, and what its prerequisites are. A very delicate balance between different properties of the system seems to be the key. The mathematical relation governing this balance is derived, which is then followed by a discussion of different gauges in atomic physics, and how the choice of gauge can affect numerical convergence. The chapter ends with a list of different scenarios where an interference pattern in the photo-electron spectrum can be observed.

In view of the very different characters of the chapters of this thesis, we summarise each chapter separately in a section at the end of the corresponding chapter.

This work is non-relativistic throughout. And furthermore, atomic units are used unless stated otherwise.

2. McLachlan method: time dependent Schrödinger equation as an eigenvalue problem

The dream of an all-round efficient solver of the Schrödinger equation is to a physicist what the dream of a generic drug for all illnesses is to a pharmacologist. It is simply too ambitious. This disappointment, however, has forced physicists over the years to creatively develop system-specific methods, just as it has forced pharmacologists to devise illness-specific drugs.

It's been nearly a century since the discovery of quantum mechanics, and the search for efficient methods for solving the Schrödinger equation goes on. Quantum Monte Carlo (QMC) and Rayleigh-Ritz Variational¹ (RRV) methods are two of the main branches of these methods. For reviews of QMC and RRV methods please refer to refs. [13, 14] and [15, 16], respectively. While QMC methods make use of random variables in quantum simulations, RRV methods rely on the variational principle.

The Full Configuration Interaction method (FCI) is one of the methods belonging to RRV methods. In this method, the full Hamiltonian is expanded in terms of possible configurations of the system. The energy of the ground state of this Hamiltonian is an upper bound on the energy of the ground state of the system under study. One main drawback of any FCI method is that there are usually an “astronomical” number of configurations available, even for the smallest of systems.

Diffusion Monte Carlo methods (DFC) are on the other hand a branch of QMC methods. They are based on treating the Schrödinger equation as a diffusion equation. These algorithms use methods that simulate the phenomenon of diffusion

1. sometimes simply referred to as variational methods

in order to find the ground state of a quantum system.

In 2009 Booth et al. developed a method based on a mixture of the QMC and FCI methods [4], which outperformed many if not all of the pre-existing methods [17–19]. This algorithm has been called Full Configuration Interaction Quantum Monte Carlo (FCIQMC) and has come a long way ever since. This algorithm has been extended in many different ways to accommodate different types of systems [20].

We will take a step in this chapter towards extending the FCIQMC method. Motivated by the fact that this method is shown to be a method for finding the ground state of a hermitian matrix [21, 22], we will write down the time dependent Schrödinger equation (TDSE) in a form similar to the diffusion equation. An attempt is then made to solve this new diffusion equation using the FCIQMC method.

The chapter is organised as follows. After introducing the imaginary time Schrödinger equation in section 2.1, a discussion of the basics of the FCIQMC method is followed in section 2.2. The new formalism of the TDSE is written in terms of an operator called McLachlian introduced in section 2.3, which is followed by a discussion on its properties in the same section. It is explained in section 2.4 how the McLachlian is implemented, and how the initial conditions are applied. The resulting eigensystem is solved in two ways, 1. direct diagonalisation of the McLachlian matrix as described in section 2.5, and 2. using the FCIQMC method as described in section 2.6. While the direct diagonalisation works as hoped, it'll be seen that the FCIQMC method fails hopelessly.

It will be shown in section 2.6 that the ground state of the McLachlian eigensystem is highly degenerate in the absence of any initial condition, and the degeneracy is only lifted when the initial conditions are applied. However, the difference between the eigenvalues of the ground state and the second state is unfortunately very small. The smallness of this difference is indeed the main culprit in the failure of any diffusion based method including the FCIQMC method in finding the ground state of the McLachlian matrix.

Furthermore, while it has been claimed before that finding the ground state of such an eigensystem can be easily done in parallel [5], a quick analysis appearing in section 2.7 shows that the solution of the TDSE starts forming from the initial

time step (which is where the initial conditions are applied) towards the later time steps. This fact makes this new formalism in practice no different from a time propagation method. In fact, as seen in section 2.8 any propagation method can be reformulated as an eigenvalue problem.

2.1. Imaginary time Schrödinger equation

The time independent Schrödinger equation (TISE) written in atomic units reads

$$H\phi_i(x) = E_i\phi_i(x), \quad (2.1)$$

where x denotes the generalised (possibly multidimensional) coordinates of the particle under study, and H is the time-independent Hamiltonian. E_i and $\phi_i(x)$ are the the eigenenergies and eigenfunctions, respectively. The eigenfunctions form an orthonormal basis, in terms of which the solution of the time dependent Schrödinger equation (TDSE)

$$i\frac{\partial\psi(x,t)}{\partial t} = H\psi(x,t), \quad (2.2)$$

can be expanded

$$\psi(x,t) = \sum_j a_j\phi_j(x)\exp(-iE_jt), \quad (2.3)$$

where the exponentials show the time evolution of each eigenfunction. By replacing t with $-i\tau$, we obtain the so-called imaginary time Schrödinger equation (ITSE) [23, 24]

$$\frac{\partial\psi(x,\tau)}{\partial\tau} = -H\psi(x,\tau). \quad (2.4)$$

In order to see why this substitution is helpful, we'll take a closer look at the solution of this equation. The solution can be found simply by making the same substitution in the solution of the TDSE in eq. (2.3),

$$\begin{aligned} \psi(x,\tau) &= \sum_j a_j\phi_j(x)\exp(-E_j\tau) \\ &= \exp(-E_0\tau) \left(a_0\phi_0(x) + \sum_{j\neq 0} a_j\phi_j(x)\exp(-(E_j - E_0)\tau) \right), \end{aligned} \quad (2.5)$$

where in the second line the component along the ground state has been taken out of the sum, and its time evolution has been factored out. In the asymptotic limit

of $\tau \rightarrow \infty$ we note that all the terms in the sum vanish due to the presence of exponentials, leaving us with

$$\lim_{\tau \rightarrow \infty} \psi(x, \tau) \rightarrow a_0 \exp(-E_0 \tau) \phi_0(x). \quad (2.6)$$

By renormalising the wave function regularly, the ground state is obtained

$$\lim_{\tau \rightarrow \infty} \psi(x, \tau) \rightarrow a_0 \exp(-E_0 \tau) \phi_0(x) \rightarrow \phi_0(x). \quad (2.7)$$

In sum, when the ground state of a Hamiltonian is desired, any initial wave function propagated according to the ITSE for a sufficiently long time converges to the ground state. Even though the choice of initial wave function does affect the convergence efficiency, a randomly generated wave function is usually used as the initial guess. The final wave function is the ground state of the Hamiltonian.²

2.2. Full Configuration Interaction Quantum Monte Carlo

The method of Full Configuration Interaction Quantum Monte Carlo (FCIQMC) was first introduced in 2009 by Booth et al. [4]. Since then it's been applied to a wide range of chemical systems [17–19], and been extended to cover complex wave functions of solids [20]. The parallelisability of this method makes it very appealing to a broad range of scientists [27].

The FCIQMC method is in fact a method developed for multi-particle systems possessing an extensive configuration space. A configuration is a multi-particle

2. There are many different methods which make use of the ITSE; they are usually called diffusion methods (for example Diffusion Monte Carlo [25, 26]) as the ITSE is similar in form to the diffusion equation. More clearly, if the term representing the potential energy in the Hamiltonian in eq. (2.4) is set to zero, we get

$$\frac{\partial \psi(x, \tau)}{\partial \tau} = \frac{1}{2m} \frac{\partial^2 \psi(x, t)}{\partial x^2}, \quad (2.8)$$

which is reminiscent of the diffusion equation

$$\frac{\partial \varphi(x, t)}{\partial t} = D \frac{\partial^2 \varphi(x, t)}{\partial x^2}, \quad (2.9)$$

where $\varphi(x, t)$ is the density of the diffusing material, and D is the diffusion coefficient.

state constructed by the direct product of single-particle states. The eigenfunctions (e.g. the ground state) of the multi-particle system can then be expressed in terms of these configurations. Unfortunately, the number of required configurations grows very quickly even for the smallest of systems. This is the reason why single-particle methods tend to become inefficient if naively extended to multi-particle scenarios. Some methods try to overcome this problem by focusing only on a small subset of the configuration space. On the contrary, the FCIQMC method has essentially access to the entire configuration space, which is where the first part of its name (“FCI” short for Full Configuration Interaction) comes from. Given the enormity of such a space, the FCIQMC method can only sample it randomly, hence the second part of the name (“QMC” short for Quantum Monte Carlo).

This algorithm has been analysed by W. M. C. Foulkes’s group [21, 22], and it’s been revealed how the well-known sign problem of fermionic systems is automatically alleviated in the case of the FCIQMC method. An extension to this method has been put forward by the same group to sample the density matrix instead of the Hamiltonian [28], paving the way for treating mixed systems. Ten-no has suggested another modified version of the FCIQMC method [29], where the basis is divided into two parts, the core and the rest. The core usually includes the configurations which are suspected to have the biggest role in the ground state. The Hamiltonian is directly diagonalised in the core, while the FCIQMC method is used for the rest of the basis.

We will now give a brief description of the inner workings of this method in its most general form. As discussed in the previous section the general solution to the ITSE can be expanded in terms of the eigenfunctions of the corresponding Hamiltonian,

$$\psi(x, \tau) = \sum_k c_k(\tau) \phi_k(x), \quad (2.10)$$

where c_k are complex numbers $c_k = c_k^{\mathcal{R}} + i c_k^{\mathcal{I}}$. These coefficients have to be complex as they entail the time dependent exponentials representing the time evolution of the eigenfunctions in eq. (2.5). Upon substituting eq. (2.10) in the ITSE we get

$$\frac{dc_k(\tau)}{d\tau} = - \sum_j H_{kj} c_j(\tau), \quad (2.11)$$

where in general the hermitian Hamiltonian can be complex $H_{kj} = H_{kj}^{\mathcal{R}} + i H_{kj}^{\mathcal{I}}$. If the derivative on the left hand side is approximated by the first order forward

difference,³ we get an equation governing the evolution of c_k ,

$$c_k(\tau + \Delta\tau) = c_k(\tau) - \Delta\tau \sum_j H_{kj} c_j(\tau), \quad (2.12)$$

which can be decomposed into real and imaginary parts

$$\begin{aligned} c_k^{\mathcal{R}}(\tau + \Delta\tau) &= c_k^{\mathcal{R}}(\tau) - \Delta\tau \sum_j \{H_{kj}^{\mathcal{R}} c_j^{\mathcal{R}}(\tau) - H_{kj}^{\mathcal{I}} c_j^{\mathcal{I}}(\tau)\} \\ c_k^{\mathcal{I}}(\tau + \Delta\tau) &= c_k^{\mathcal{I}}(\tau) - \Delta\tau \sum_j \{H_{kj}^{\mathcal{I}} c_j^{\mathcal{R}}(\tau) + H_{kj}^{\mathcal{R}} c_j^{\mathcal{I}}(\tau)\}. \end{aligned} \quad (2.13)$$

These are the main equations behind the FCIQMC method. The weight of the wave function along each component (called sites), i.e. $c_j(\tau)$, is represented by signed walkers⁴ which can have either a (+) or a (-) sign. The number of walkers on site j determines the value of c_j up to a global normalising factor. The terms inside the curly brackets are interpreted as spawning probabilities. Every time step each walker has the opportunity to spawn child walkers on other sites (or on its own site), during which process a new walker is placed on the corresponding site. When two walkers on the same site have opposite signs, they cancel out, a process called annihilation. Initially a number of walkers are placed on some sites (which may be randomly chosen, but don't have to be), and if the simulation is allowed to run for long enough, the final distribution of the walkers represents the ground state of the Hamiltonian.

To correctly describe the complex weights c_k , two sets of walkers are implemented, one representing $c_k^{\mathcal{R}}$ (called real walkers) and one representing $c_k^{\mathcal{I}}$ (called imaginary walkers). Going back to eqs. (2.13), the first term in the sum on the right hand side of the first equation shows the probability of a real walker on site j spawning a real child walker on site k . The second term shows the probability of an imaginary walker on site j spawning a real child walker on site k , and so on.

It is stated in the original paper [4] that for the sake of efficiency, instead of including all the terms in the sums in eqs. (2.13), any walker on site j is only allowed to appear twice on the right hand side, once in the equation governing the evolution of site j , and once in the equation governing the evolution of site k where k is randomly chosen. In other words, every walker is allowed to spawn twice every time step, once on its own site, and once on a site randomly chosen. The random

3. The first order forward difference is $\dot{f}(t) = \frac{f(t+\Delta t) - f(t)}{\Delta t}$.

4. Ironically they don't walk at all.

number generator in charge of picking the second site generates numbers according to the probability distribution $P(j \rightarrow k)$ provided by the user. This probability distribution represents the probability that site k is chosen to be spawned a child walker on by the walker on site j . Therefore, the closer $P(j \rightarrow k)$ mimics the structure of the j -th row of the Hamiltonian, the more efficient the algorithm is.⁵ However, most of the time the structure of the Hamiltonian is not known, or is time consuming to discover, thus a uniform probability is used, i.e. $P(j \rightarrow k) = 1/M$ where M is the number of spatial basis functions. Under this modification, a walker on site j spawns a child walker on site k with slightly different probabilities

$$\begin{aligned}
\mathcal{P}_{j \rightarrow k}^{\mathcal{R} \rightarrow \mathcal{R}} &= -\frac{H_{kj}^{\mathcal{R}} \Delta\tau}{P(j \rightarrow k)} \\
\mathcal{P}_{j \rightarrow k}^{\mathcal{R} \rightarrow \mathcal{I}} &= -\frac{H_{kj}^{\mathcal{I}} \Delta\tau}{P(j \rightarrow k)} \\
\mathcal{P}_{j \rightarrow k}^{\mathcal{I} \rightarrow \mathcal{R}} &= +\frac{H_{kj}^{\mathcal{I}} \Delta\tau}{P(j \rightarrow k)} \\
\mathcal{P}_{j \rightarrow k}^{\mathcal{I} \rightarrow \mathcal{I}} &= -\frac{H_{kj}^{\mathcal{R}} \Delta\tau}{P(j \rightarrow k)},
\end{aligned} \tag{2.14}$$

where, for example, $\mathcal{P}_{j \rightarrow k}^{\mathcal{R} \rightarrow \mathcal{I}}$ is the probability that a real walker on site j spawns an imaginary child walker on site k . For better convergence, it is advised to use small enough $\Delta\tau$ such that none of the probabilities is greater than one. Bigger values might cause the algorithm to diverge [4].

One point worth mentioning is that if any of the above probabilities turns out to be negative, the sign should merely serve as a means of determining what the sign of the child walker will be, and therefore should be ignored when evaluating the probabilities. The sign of the newly born walker is determined by the product of three signs, 1. the sign of the mother walker, 2. the sign appearing in front of the corresponding probability in eqs. (2.14), and 3. the sign of the Hamiltonian.

In the case of a purely real Hamiltonian there will be only real walkers and only one spawning probability

$$\mathcal{P}_{j \rightarrow k}^{\mathcal{R} \rightarrow \mathcal{R}} = -\frac{H_{kj}^{\mathcal{R}} \Delta\tau}{P(j \rightarrow k)}. \tag{2.15}$$

5. A similar situation is encountered when using Monte Carlo integration. The choice of probability distribution plays a key role in the efficiency of this algorithm. And if a good knowledge of the structure of the integrand is available, the efficiency can be further increased by incorporating importance sampling.

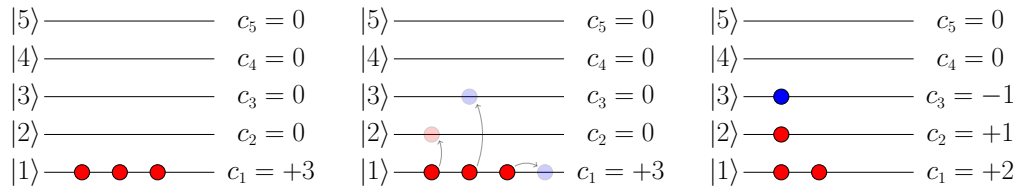


Figure 2.1. These three diagrams show the three constituent stages of the FCIQMC method for a real Hamiltonian. The color of each walker represents its sign. The diagram on the far left shows the population of the walkers remaining from the previous time step. Each walker happens to spawn one child walker. In the middle diagram the three walkers have spawned on three different sites, the first one on $|2\rangle$ with like sign, the second one on $|3\rangle$ with opposite sign, and the third one on the same site with opposite sign. The two unlike signs on the same site cancel out. The third diagram shows the distribution of walkers at the end of the time step.

A simple illustration of the algorithm is shown fig. 2.1.

The FCIQMC method is in fact a diffusion Monte Carlo method for finding the ground state of a hermitian matrix [21, 22], see fig. 2.2. However, being a Monte Carlo method based on the diffusion equation, it has two major downsides. The first one is that assuming the ground state is non-degenerate, the method can resolve the ground state and the next state separated by an energy difference of ΔE only if it is allowed to run for $\tau \propto 1/\Delta E$. More precisely, by taking a look at eq. (2.5) we see that if the weight of the ground state is one, the weight of the second state is $\exp(-\Delta E \tau)$, and in order for this exponential prefactor to be small, τ has to be at least of the same order as $1/\Delta E$.

The second drawback comes from the Monte Carlo nature of the FCIQMC method, i.e. the accuracy of the ground state of a matrix is dictated by the number of walkers. Imagine the current wave function found by the FCIQMC method is written as $\psi = a_0\phi_0 + a_\perp\phi_\perp$, where ϕ_0 is the true ground state and ϕ_\perp is everything else. If there are N walkers and we are lucky enough to have all walkers but one

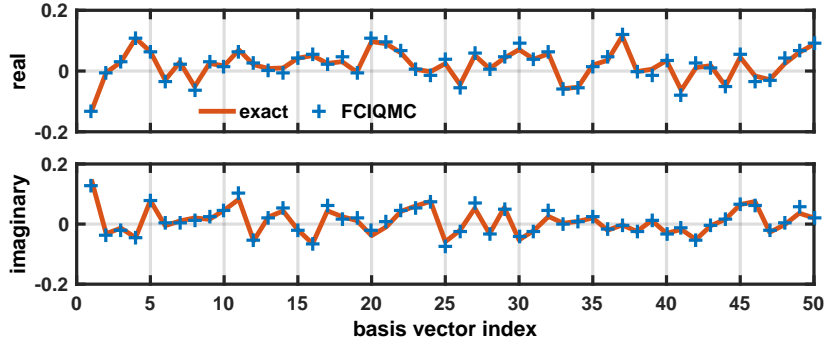


Figure 2.2. The figure shows the real and imaginary components of the ground state of a random 50×50 hermitian matrix found using direct diagonalisation and the FCIQMC method. It is seen from the agreement that the FCIQMC method is indeed an algorithm for finding the ground state of a hermitian matrix.

on the true ground state, the weights will be given by

$$\begin{aligned}
 a_0 &= \frac{N-1}{\sqrt{(N-1)^2+1}} \approx 1 \\
 a_\perp &= \frac{1}{\sqrt{(N-1)^2+1}} \approx \frac{1}{N}.
 \end{aligned} \tag{2.16}$$

Therefore, the inaccuracy (determined by the weight of ϕ_\perp) is of the order of $1/N$. For instance, for an accuracy of 10^{-6} at least 10^6 walkers are needed.

We will soon see how these two drawbacks make it practically impossible to use the FCIQMC method to solve the dynamics of a system.

2.3. McLachlian operator: definition

Once again we start from the TDSE and move everything to the left hand side to get $0 = H\psi(x, t) - i\partial_t\psi(x, t)$ (for brevity, the time dependence of the Hamiltonian is not explicitly written). This equation has to be exactly satisfied every single time step in order for us to have the exact time evolution of the wave function. But we know that using a finite time step Δt prevents this relation from being exactly satisfied, and therefore always leaves us with some residue which can be positive or negative. As becomes apparent soon, we need to redefine this residue

to make sure it is always non-negative, which is done by squaring it,⁶

$$\begin{aligned}\mathcal{R}(x, t) &= |H\psi(x, t) - i\partial_t\psi(x, t)|^2 \\ &= (H\psi(x, t) - i\partial_t\psi(x, t))^\dagger (H\psi(x, t) - i\partial_t\psi(x, t)) .\end{aligned}\tag{2.17}$$

It is obvious that if two different approximate evolutions of the wave function are given, the one having a smaller residue over the entire space $\int \mathcal{R}(x, t) dx$ is a better approximation; the residue of the exact solution is zero. To find the wave function at the next time step, it is only necessary to minimise this quantity over the entire space. Since this has to be done at each time step separately and the residue is never negative, the integral of the residue over space and time has to be minimised, that is

$$I = \int \mathcal{R}(x, t) dx dt \geq 0, \tag{2.18}$$

where the equality sign is only satisfied by the exact solution.

Using the Dirac bra-ket notation in quantum mechanics, we can write $\psi(x) \equiv \langle x|\psi\rangle$. Similarly, we define a new notation $\psi(x, t) \equiv (\langle x| \otimes \langle t|) |\psi\rangle$,⁷ in terms of which we can rewrite eq. (2.18),

$$I = \int \langle\langle \psi | (H - \widehat{i\partial_t})^\dagger |x\rangle \otimes |t\rangle \langle t| \otimes \langle x| (H - \widehat{i\partial_t}) |\psi\rangle \rangle dx dt, \tag{2.19}$$

where $\widehat{i\partial_t}$ is an operator⁸ whose representation in the time domain is differentiation with respect to time, $\langle t|\widehat{i\partial_t}|\psi\rangle \equiv i\partial_t \langle t|\psi\rangle$. This expression can be simplified with the help of the closure relation in the position space $\int |x\rangle \langle x| dx = \mathbb{1}$,

$$I = \int \langle\langle \psi | (H - \widehat{i\partial_t})^\dagger |t\rangle \langle t| (H - \widehat{i\partial_t}) |\psi\rangle \rangle dt, \tag{2.20}$$

6. In the later stages of the project we found out that this equation had already been derived by McLachlan in a slightly different way [30]. This is why we have called the operator to follow the McLachlian.

7. The reason why $|\dots\rangle\rangle$ is used instead of $|\dots\rangle$ will become clear in subsection 2.3.3. This notation is often used when tackling periodical systems with Floquet Theory [31].

8. Physicists mostly believe that time appears as a parameter in quantum mechanics, in contrast to position and momentum which appear as operators. One is usually referred to Wolfgang Pauli's book [32] when asking about the reason. Pauli states in a footnote in his book that since having a time operator requires an energy space covering all energies from $-\infty$ to ∞ (which any choice of Hamiltonian fails to satisfy), the idea of time operator should be forgotten altogether. However, as explained by Hilgevoord [33], the energy space does indeed cover all the energies from $-\infty$ to ∞ , and the problem arises when the energy space is confused with the energies a system can have. This is very similar to the momentum of a particle in an infinite potential well; the momentum is not demoted from an operator to a parameter just because its eigenvalues are discrete.

	Hamiltonian	McLachlian
eigenvalue problem	$H \psi_k\rangle = E_k \psi_k\rangle$	$\hat{I} \psi_k\rangle\rangle = \alpha_k \psi_k\rangle\rangle$
variational method	$\langle\psi H \psi\rangle \geq E_0$	$\langle\langle\psi \hat{I} \psi\rangle\rangle \geq 0$
diffusion equation	$\partial_\tau\psi(x, \tau) = -H\psi(x, \tau)$	$\partial_\tau\psi(x, t, \tau) = -\hat{I}\psi(x, t, \tau)$

Table 2.1. TDSE and TISE are similar in a number of ways. They both can be written as an eigenvalue problem. The ground states of both of them can be found by either using a variational method, or solving a diffusion equation.

and using a similar closure relation in the time domain⁹ $\int |t\rangle \langle t| dt = \mathbb{1}$,

$$I = \langle\langle\psi| \left(H - i\widehat{\partial}_t \right)^\dagger \left(H - i\widehat{\partial}_t \right) |\psi\rangle\rangle . \quad (2.21)$$

We are now in a position to define the term sandwiched between the two wave functions as a new operator acting in the position space as well as the time domain

$$\hat{I} = \left(H - i\widehat{\partial}_t \right)^\dagger \left(H - i\widehat{\partial}_t \right) , \quad (2.22)$$

which we shall call the McLachlian.¹⁰

Expressed in terms of the McLachlian, eq. (2.18) takes a compact form

$$\langle\langle\psi|\hat{I}|\psi\rangle\rangle \geq 0 , \quad (2.23)$$

which bears a clear resemblance to the main equation of the variational theorem $\langle\psi|H|\psi\rangle \geq E_0$ for the TISE [34]. Furthermore, the solution to the TDSE can be interpreted as the ground state of an eigenvalue problem $\hat{I} |\psi_k\rangle\rangle = \alpha_k |\psi_k\rangle\rangle$, which is similar to the TISE written as an eigenvalue problem. Analogously, we can write a diffusion equation which in the limit of $\tau \rightarrow \infty$ converges to the solution of the TDSE, $\partial_\tau\psi(x, t, \tau) = -\hat{I}\psi(x, t, \tau)$, very much like the ITSE. Notice that the diffusion time τ is indeed different from the real time t . A comparison between the Hamiltonian and the McLachlian is shown in table 2.1.

In the next two subsections some properties of the McLachlian operator will

9. The reader may convince themselves of the validity of this relation simply by thinking of $|t\rangle$ as a set of delta functions spanning the whole time interval.

10. We have chosen *McLachlian* over *McLachlanian* because of its easier pronunciation, even though technically the latter should be used.

be examined followed by a discussion on the basis needed for accommodating the McLachlian operator.

2.3.1. Proof of hermiticity of $\widehat{i\partial_t}$

One way for the McLachlian to be hermitian is to consist of individually hermitian terms. The Hamiltonian is hermitian, but let's see if the same can be said of $\widehat{i\partial_t}$. To prove the hermiticity of $\widehat{i\partial_t}$ it is sufficient to show

$$\langle\langle\psi|\overrightarrow{i\partial_t}|\psi\rangle\rangle = \langle\langle\psi|\overleftarrow{i\partial_t}|\psi\rangle\rangle, \quad (2.24)$$

where the arrows indicate if the operator is acting on the state on the right or the left. We start from the left hand side and will see how the right hand side emerges automatically. By inserting the closure relation in the time domain in the middle we get

$$\begin{aligned} \langle\langle\psi|\overrightarrow{i\partial_t}|\psi\rangle\rangle &= \int_{t_i}^{t_f} \langle\langle\psi|t\rangle\rangle \langle t|\widehat{i\partial_t}|\psi\rangle dt \\ &= i \int_{t_i}^{t_f} \langle\langle\psi|t\rangle\rangle (\partial_t \langle t|\psi\rangle) dt, \end{aligned} \quad (2.25)$$

where in going from the first line to the second we have used the notation introduced in the last section. Using integration by parts yields

$$i \int_{t_i}^{t_f} \langle\langle\psi|t\rangle\rangle (\partial_t \langle t|\psi\rangle) dt = i \langle\langle\psi|t\rangle\rangle \langle t|\psi\rangle \Big|_{t_i}^{t_f} - i \int_{t_i}^{t_f} (\partial_t \langle\langle\psi|t\rangle\rangle) \langle t|\psi\rangle dt, \quad (2.26)$$

where the first term vanishes because $|\langle t|\psi\rangle| = 1$ at all times, and the second term is nothing but the right hand side of eq. (2.24),

$$-i \int_{t_i}^{t_f} (\partial_t \langle\langle\psi|t\rangle\rangle) \langle t|\psi\rangle dt = \int_{t_i}^{t_f} (i\partial_t \langle t|\psi\rangle)^* \langle t|\psi\rangle dt = \langle\langle\psi|\overleftarrow{i\partial_t}|\psi\rangle\rangle. \quad (2.27)$$

Therefore operator $\widehat{i\partial_t}$ is hermitian.

2.3.2. Non-zero eigenvalues of the McLachlian

The McLachlian matrix has other eigenvalues than zero, which might beg the question of what do their eigenvectors represent. Given that $\widehat{i\partial_t}$ is a hermitian

operator, the McLachlian can be rewritten as

$$\begin{aligned}\hat{I} &= \left(H - i\widehat{\partial}_t\right)^\dagger \left(H - i\widehat{\partial}_t\right) \\ &= \left(H - i\widehat{\partial}_t\right)^2,\end{aligned}\tag{2.28}$$

giving rise to the eigenvalue equation

$$(H - i\partial_t)^2\psi = \alpha^2\psi,\tag{2.29}$$

where α is a real number. By taking the right hand side of this equation to the left, and rewriting the left hand side as the product of two separate terms we get

$$(H - i\partial_t - \alpha)(H - i\partial_t + \alpha)\psi = 0,\tag{2.30}$$

with the general solution¹¹

$$\psi(x, t) = c_1 e^{-i(H+\alpha)t}\psi_0(x) + c_2 e^{-i(H-\alpha)t}\psi_0(x).\tag{2.31}$$

It is seen that the eigenvectors corresponding to higher eigenvalues represent the same dynamics only with a global shift of energy.

It'll be shown in the next section that in order for us to be able to apply the proper boundary conditions in time, which is a requirement for the implementation of the McLachlian matrix, we will have to sacrifice the hermiticity. As a side-effect, the eigenvectors corresponding to higher eigenvalues will no longer describe the same dynamics, as the proof above hinges on the McLachlian being hermitian. However, if periodic boundary conditions in time are used (see appendix B), the hermiticity of $i\partial_t$ is preserved. In this case, the higher eigenvectors indeed represent the same dynamics with a global shift of energy

2.3.3. Basis required for the McLachlian

In quantum mechanics the wave function is almost always expressed in terms of a set of complete spatial basis functions $\phi_i(x)$.¹² Even though not customary, it is also possible to express the time evolution of the wave function in terms of a set of temporal basis functions $T_j(t)$.

11. In the case of a time-dependent Hamiltonian the exponents have to be replaced with appropriate integrals.

12. In the simplest case Dirac delta functions $\delta(x)$ are used as the basis functions.

In order to express the McLachlian operator, a complete basis in space and time is needed. Such a basis is easily constructed by the outer product of the spatial and temporal bases,

$$\Lambda = T \otimes \phi, \quad (2.32)$$

or when written out explicitly

$$\Lambda_{i,j}(x, t) = T_i(t)\phi_j(x), \quad (2.33)$$

where $\Lambda(x, t)_{i,j}$ are the basis functions in space-time. The time-dependent wave function can now be expressed in terms of these basis functions

$$\psi(x, t) = \sum_{i,j} a_{i,j} \Lambda_{i,j}(x, t), \quad (2.34)$$

where $a_{i,j}$ have no spatial or temporal dependence.

Normally the wave function is written using the Dirac notation in the absence of a specific spatial basis, i.e. $|\psi(t)\rangle$. Since there is usually no temporal basis, the time dependence is expressed simply by adding an argument to the ket $|\psi\rangle \rightarrow |\psi(t)\rangle$. On the contrary, we have included a temporal basis in our approach, and in order to express the wave function in the absence of a specific spatio-temporal basis, we use a slightly different version of the Dirac notation, i.e. $|\dots\rangle\rangle$ (the same notation is used by physicists when applying Floquet theory to periodic systems [31]). When the temporal dependence is needed, the wave function has to be multiplied from the left by the suitable temporal basis function, i.e. $|\psi(t)\rangle = \langle t|\psi\rangle\rangle$. And when the spatio-temporal dependence is required, it has to be multiplied by the suitable spatio-temporal basis function, that is $\psi(x, t) = (\langle t|\otimes\langle x|)|\psi\rangle\rangle$. Now we can rewrite eq. (2.34) using the new notation,

$$|\psi\rangle\rangle = \sum_{i,j} a_{i,j} |\Lambda_{i,j}\rangle\rangle, \quad (2.35)$$

where $|\Lambda_{i,j}\rangle\rangle$ form the spatio-temporal basis, and $a_{i,j} = \langle\langle\Lambda_{i,j}|\psi\rangle\rangle$. To sum, the space in which $|\psi\rangle\rangle$ is defined has one more dimension than the space required to express $|\psi(t)\rangle$, as the latter is only one slice of the former at a particular time.

2.4. McLachlian operator: implementation

A suitable choice of spatial basis depends to a large degree on the system under study. A basis appropriate for one system may be completely inadequate for an-

other. The temporal basis on the other hand proves to be trickier, as the dynamics of the system is not known prior to solving the TDSE. The simplest temporal basis is a grid in time, where each basis function represents a point in time. As will be soon seen, this choice of basis makes it possible to apply the initial conditions without any complications.

The first step is to express the time-derivative appearing in the TDSE in the finite difference language; for a short primer on finite differences please refer to appendix B. Using the central difference [35], the time derivative is

$$\dot{f}_i \equiv \frac{f_{i+1} - f_{i-1}}{2\Delta t}, \quad (2.36)$$

or when expressed as a vector

$$\partial_t \mathbf{f} \equiv \begin{pmatrix} \dot{f}_0 \\ \dot{f}_1 \\ \vdots \\ \dot{f}_{N-2} \\ \dot{f}_{N-1} \end{pmatrix} = \frac{1}{2\Delta t} \begin{pmatrix} f_1 - f_{-1} \\ f_2 - f_0 \\ \vdots \\ f_{N-1} - f_{N-3} \\ f_N - f_{N-2} \end{pmatrix}, \quad (2.37)$$

where it is seen that when the derivative at $t = 0$ is desired, the value of the function at $t = -\Delta t$ is required. Depending on the boundary condition used, f_{-1} is replaced with different values. For example, in the case of the Neumann boundary condition f_{-1} is replaced with f_0 ; for more details see appendix B. But none of the boundary conditions in the literature is useful here, as each of them carries a special meaning unjustified in this case.

What we need to do is express the derivative at $t = 0$ in a way that no boundary conditions are assumed. This can be only done by expressing the derivatives at the boundaries in terms of the points inside the domain. In other words, we have to use one-sided finite differences to express the derivatives at points $t = 0$ and $t = N \Delta t$. The forward difference of second-order accuracy is

$$\dot{f}_i = \frac{-3f_i + 4f_{i+1} - f_{i+2}}{2\Delta t}, \quad (2.38)$$

and similarly the backward difference

$$\dot{f}_i = \frac{f_{i-2} - 4f_{i-1} + 3f_i}{2\Delta t}, \quad (2.39)$$

where these one-sided differences are of the same order as the central difference introduced earlier. If the derivatives at $t = 0$ and $t = N \Delta t$ in eq. (2.37) are

replaced with the one-sided differences we obtain

$$\begin{aligned} \partial_t \mathbf{f} &= \frac{1}{2\Delta t} \begin{pmatrix} -3f_0 + 4f_1 - f_2 \\ f_2 - f_0 \\ \vdots \\ f_N - f_{N-2} \\ f_{N-2} - 4f_{N-1} + 3f_N \end{pmatrix} \\ &= \frac{1}{2\Delta t} \begin{pmatrix} -3 & 4 & -1 \\ -1 & 0 & 1 \\ & \ddots & \\ & -1 & 0 & 1 \\ & & 1 & -4 & 3 \end{pmatrix} \begin{pmatrix} f_0 \\ f_1 \\ \vdots \\ f_{N-2} \\ f_{N-1} \end{pmatrix}, \end{aligned} \quad (2.40)$$

where the differentiation operator is clearly not hermitian any more. However, if periodic boundary conditions are used (see appendix B) the hermiticity will be retained.

Having expressed the time-derivative, it is time to write down the McLachlian matrix, which is easily done given the temporal basis just defined

$$\hat{I} = \left(\bigoplus_{i=0}^{N-1} H(t_i) - i\partial_t \otimes \mathbb{1}_\phi \right)^\dagger \left(\bigoplus_{i=0}^{N-1} H(t_i) - i\partial_t \otimes \mathbb{1}_\phi \right), \quad (2.41)$$

where \bigoplus is the direct sum, and \dagger means hermitian conjugation. $\mathbb{1}_T$ and $\mathbb{1}_\phi$ are the identity matrices in the time domain and position space respectively. To make sense of this expression, remember that $|\psi\rangle\rangle$ is constructed, using a temporal basis defined by a grid in time, simply by stacking up $|\psi(t)\rangle$ at different times, that is

$$|\psi\rangle\rangle = \begin{pmatrix} |\psi(t_0)\rangle \\ |\psi(t_1)\rangle \\ \vdots \\ |\psi(t_f)\rangle \end{pmatrix}. \quad (2.42)$$

In the case of a time-independent Hamiltonian the McLachlian matrix reduces to

$$\hat{I} = (\mathbb{1}_T \otimes H - i\partial_t \otimes \mathbb{1}_\phi)^\dagger (\mathbb{1}_T \otimes H - i\partial_t \otimes \mathbb{1}_\phi). \quad (2.43)$$

2.4.1. Initial conditions

Any solution to the TDSE should in principle act as an eigenvector of the McLachlian eigenvalue problem. In an M -dimensional spatial space there are M orthogonal initial conditions, in terms of which any arbitrary initial condition can be expanded. The solution to the TDSE for any arbitrary initial condition should be expressible using the solutions for these two orthogonal initial conditions. This is a direct consequence of the temporal linearity of the TDSE.

In this M -dimensional space, the ground state of the McLachlian matrix is M -fold degenerate, each of which is the solution to the TDSE for one of the M orthogonal initial conditions (see fig. 2.8). This suggests that one way to have the solution for an arbitrary initial condition is to find all the ground states, in terms of which any solution can be expressed later.

If the size of the system is large, however, it is not feasible to find all the ground states first and then find a combination satisfying the initial conditions. Therefore, it is imperative that we apply the initial conditions prior to finding the ground state. The initial conditions should lift the degeneracy such that there is only one ground state left. This can be easily done by adding a penalty term to the McLachlian matrix [36],

$$p \equiv \begin{array}{l} |t=0\rangle \\ |t=1\rangle \\ \vdots \\ |t=N\rangle \end{array} \begin{pmatrix} \langle t=0| & \langle t=1| & \dots & \langle t=N| \\ \mathbb{1} - |\psi_0\rangle\langle\psi_0| & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}, \quad (2.44)$$

where $|\psi_0\rangle$ is the initial state. Therefore, the new McLachlian matrix is

$$\hat{I} \rightarrow \hat{I} + p. \quad (2.45)$$

It can be confirmed by direct substitution that only the state satisfying the initial conditions serves as the ground state of the new McLachlian matrix with eigenvalue 0.

Applying the initial conditions using a penalty term has the side-effect that the solution is found up to a global phase. To clarify, consider the following element from the penalty term

$$\mathbb{1} - |\psi_0\rangle\langle\psi_0|, \quad (2.46)$$

where again $|\psi_0\rangle$ is the initial state. If a global phase is added to the initial state $|\psi_0\rangle \rightarrow \exp(i\theta)|\psi_0\rangle$, the penalty term remains unaltered. In other words, the solution obtained in this way may have a different global phase to the original initial state. But this doesn't have any effect on the validity of the solution, as a constant global phase is trivial in quantum mechanics.

2.5. Finding the ground state of the McLachlian matrix: direct diagonalisation

Having the new formalism of the TDSE at our disposal, we would like to solve for the dynamics of a few systems. We will consider three systems: a two-level system, a three-level system, and a hydrogen atom in the presence of a laser pulse. First we will write down the Hamiltonian for each system, and then form the corresponding McLachlian, the ground state of which is the solution of the TDSE. There are many methods in the literature for finding the ground state of a matrix. In this section we will find the ground state by directly diagonalising the McLachlian, which is then compared with the solution by textbook propagation methods. The next chapter is an attempt at using the FCIQMC method to find the ground state of the McLachlian, and it becomes apparent there that diffusion methods won't work in this particular case.

2.5.1. Two-level system

First we will consider a two-level system with its two levels at $E_1 = 0$ and $E_2 = 1$. The two levels are coupled through the term $\gamma e^{i\omega t}$ where $\omega = 4$ and $\gamma = 1$. The Hamiltonian in matrix form is

$$H = \begin{pmatrix} E_1 & \gamma e^{i\omega t} \\ \gamma^* e^{-i\omega t} & E_2 \end{pmatrix}. \quad (2.47)$$

This system can be solved analytically [34]. The numerical solution obtained by diagonalising the McLachlian matrix along with the analytic solution are drawn in fig. 2.3, where a perfect agreement is visible. The system starts off with all its population in the state corresponding to E_1 , and is then allowed to evolve from $t = 0$ to $t = 6$ with time step $dt = 0.01$.

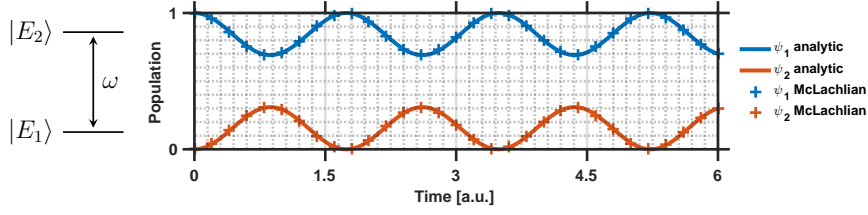


Figure 2.3. State populations of a two-level system solved by directly diagonalising the McLachlan matrix compared with the analytic solution [34].

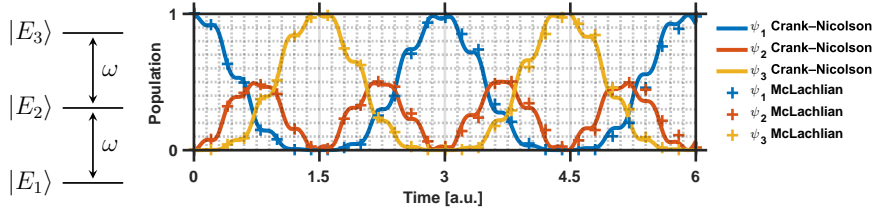


Figure 2.4. State populations of a three-level system solved by directly diagonalising the McLachlan matrix compared with Crank-Nicolson time propagation method.

2.5.2. Three-level system

The next system we will consider is a three-level system with its levels at $E_1 = 0$, $E_2 = \omega$, and $E_3 = 2\omega$, and each two consecutive levels are coupled through $\gamma \cos(\omega t)$. The parameters used are $\omega = 9$ and $\gamma = 3$. The Hamiltonian is

$$H = \begin{pmatrix} E_1 & \gamma \cos(\omega t) & 0 \\ \gamma \cos(\omega t) & E_2 & \gamma \cos(\omega t) \\ 0 & \gamma \cos(\omega t) & E_3 \end{pmatrix}. \quad (2.48)$$

The numerical solution obtained via directly diagonalising the McLachlan matrix, and a solution by the Crank-Nicolson scheme are shown in fig. 2.4. Again a good agreement between the two is visible, even though the agreement worsens as time progresses (no two numerical schemes agree forever, but it is possible to extend the agreement by decreasing the time step or simply by using higher order schemes). The system is initially in the state corresponding to E_1 , and then evolves from $t = 0$ to $t = 6$ with time step $dt = 0.01$.

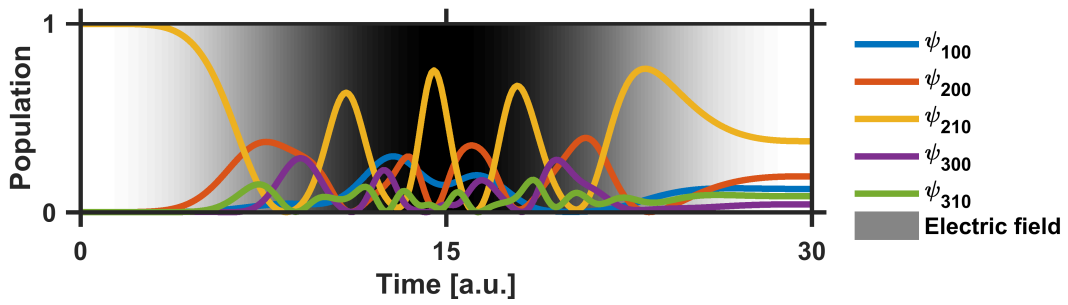


Figure 2.5. State populations of a hydrogen atom in the presence of a time dependent electric field. The electron is initially in the $2p$ -state ϕ_{210} . The gray background shows the magnitude of the laser pulse. States with non-zero magnetic quantum numbers don't couple to the $2p$ state (a direct consequence of the linear polarisation of the laser pulse along the z -axis).

2.5.3. Atomic hydrogen in a time dependent electric field

The third and last example is a hydrogen atom exposed to a time dependent electric field. The Hamiltonian of such a system is

$$H = \frac{1}{2}p^2 + V(r) + V_e(\mathbf{r}, t), \quad (2.49)$$

where p is the magnitude of the momentum operator \mathbf{p} , $V(r) = -1/r$ is the potential of the nucleus, and $V_e(\mathbf{r}, t) = zE(t)$ represents the electric potential. The electric field $E(t)$ is given by,

$$E(t) = \begin{cases} \mathcal{E}_0 \cos^2\left(\frac{\pi t}{2\tau}\right) & \text{if } |t| < \tau \\ 0 & \text{otherwise,} \end{cases} \quad (2.50)$$

where \mathcal{E}_0 represents the electric field amplitude, and τ is the full width at half maximum (FWHM) of the pulse. The values used here are $\mathcal{E}_0 = 0.3$ a.u. and $\tau = 15$ a.u.

The first five relevant states of the hydrogen atom are shown in fig. 2.5 (15 states were included in the simulation), which are obtained by directly diagonalising the McLachlian matrix. The norm of the wave function remains constant over time which is expected of any valid simulation.

These three examples should convince us that the ground state of the McLachlian matrix is indeed the solution to the TDSE.

2.6. Finding the ground state of the McLachlian matrix: FCIQMC

Having made sure in the last section that the ground state of the McLachlian matrix is the solution to the TDSE, we are in a position now to use the FCIQMC method to find this ground state, as the FCIQMC method is in fact a method for finding the ground state of a hermitian matrix [21, 22].

By taking a look at the two diffusion equations in table 2.1, it becomes clear that in order to use the FCIQMC method to find the ground state of the McLachlian matrix, the Hamiltonian in the FCIQMC method should be replaced with the McLachlian. Consequently, the McLachlian matrix elements should replace those of the Hamiltonian in all the probabilities appearing in eqs. (2.14).

After several months of numerous attempts to find the ground state of the McLachlian using the FCIQMC method, all our attempts failed one after another. Among the systems we considered were systems as simple as few-level systems all the way to systems as complex as the quantum harmonic oscillator, see fig. 2.6. After letting the simulations run for a very long time (of the order of 10^8 time steps), there were no sign of convergence in any of the systems. In order to see why the FCIQMC method had failed in all the systems mentioned, we had to take a deeper look at the FCIQMC method.

As it was briefly mentioned towards the end of section 2.2, diffusion methods have to run for a long time if the difference between the eigenstates of the ground state and the next state is small, i.e. the ground state is quasi-degenerate. In fig. 2.7 two systems are shown, one with a distinct ground state and one with a quasi-degenerate ground state; the eigenstates are drawn on a vertical axis to the right of each plot. FCIQMC was used in both cases with the same number of time steps and with the same number of walkers. It can be clearly seen that FCIQMC fails to find the ground state if the ground state is quasi-degenerate. Of course it might be possible to let the algorithm run for longer until convergence is achieved, but the question is how much longer. If the two quasi-degenerate are too close together, the convergence time can be so long that FCIQMC will be rendered practically ineffective.

Now let's have a look at the eigenvalues of the McLachlian. In fig. 2.8 the nine

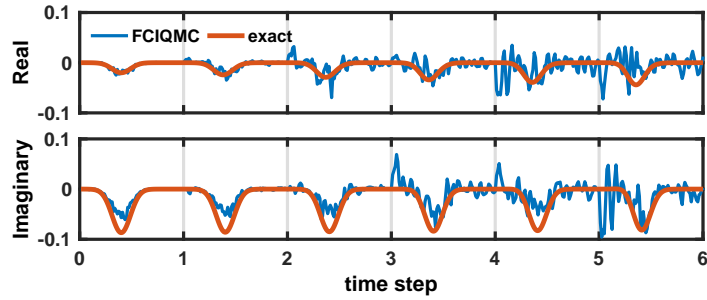


Figure 2.6. The quantum harmonic oscillator ($H = p^2/2 + 1/2x^2$) is solved for 6 time steps (i.e. 6 temporal basis functions) using the FCIQMC method and direct diagonalisation of the McLachlian matrix (eq. (2.43)). Each block represents the wave function at each of these 6 time steps. For instance, the block between 2 and 3 is the wave function at the 3rd time step. 5000 walkers were used in the FCIQMC method, and the algorithm was allowed to run for 10^8 time steps (these time steps shouldn't be confused with the 6 time steps mentioned earlier. These are the time steps used in the FCIQMC method for the evolution of the walkers), where $\Delta\tau = 5 \times 10^{-4}$. It is clearly visible that the solution has started to form from the first time step towards later time steps (the first time step is where the initial conditions are applied.). However, the convergence time appears to be unattainably long.

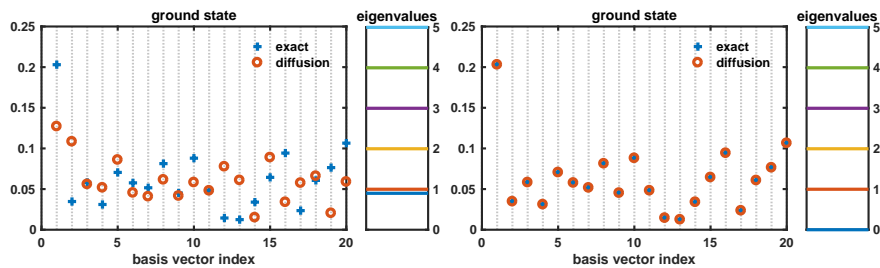


Figure 2.7. The FCIQMC method has been used to find the ground states of two random hermitian matrices. The eigenvalues of each matrix are shown on a vertical axis to the right of each plot. The FCIQMC method fails to find the ground state of the matrix on the left due to the presence of two quasi-degenerate ground states. As soon as the degeneracy is lifted, as in the plot on the right, the FCIQMC method successfully finds the ground state.

smallest eigenvalues of the McLachlian of the three-level system considered earlier are shown. In the four plots the time step is increased while the total time is kept constant. The three-fold degeneracy of the ground state is obvious in all of these plots. In fact the ground state prior to the application of the initial conditions is always M -fold degenerate where M is the size of the Hamiltonian. Each of these M -degenerate ground states is the solution to the TDSE for the initial condition that only one of the states is occupied. The solution for any arbitrary initial condition can be expressed as a superposition of these ground states.

All the eigenvectors of the McLachlian matrix have to be orthogonal due to the hermiticity of the McLachlian operator. If the time axis is discretised (as was done in sec. 2.4) into N sections, and the position space is M dimensional, the total number of eigenvectors is $M \times N$. In this space there are M orthogonal solutions corresponding to M orthogonal initial conditions. In order to see how the rest of these eigenvectors are constructed, we need to consider two different eigenvalues α_1^2 and α_2^2 . According to eq. (2.31) the corresponding eigenvectors are given by $\exp(-i\alpha_1 t) \exp(-iHt) \Psi_0(x)$ and $\exp(-i\alpha_2 t) \exp(-iHt) \Psi_0(x)$, assuming both of them satisfy the same set of initial conditions. In order for these two eigenvectors to be orthogonal, the two prefactors $\exp(-i\alpha_1 t)$ and $\exp(-i\alpha_2 t)$ should be orthogonal (since they are the only part distinguishing the two eigenvectors), and this only happens if $\alpha_1 \propto k_1/T$ and $\alpha_2 \propto k_2/T$ where N is the number of basis functions in time, T is the total duration of the simulation, $k_1, k_2 = 0, 1, \dots, N-1$ and $k_1 \neq k_2$ (remember $\exp(-i2\pi kt/T)$ with $k \in \mathbb{N}$ form an orthonormal Fourier basis over the interval of $[0, T]$). Therefore, in general we have

$$\alpha \propto \frac{k}{T} \quad \text{with} \quad k = 0, 1, \dots, N-1. \quad (2.51)$$

Looking again at eq. (2.31), we can see a general solution has two terms, one with eigenvalue α^2 , and the other with $(-\alpha)^2$. If the eigenvalues of the McLachlian are shown on a plot, there should be then two eigenvalues equal to α^2 for each of the M orthogonal initial conditions. Therefore, there should be M eigenvalues equal to 0, followed by $2M$ eigenvalues $\sim 1/T^2$, followed by $2M$ eigenvalues $\sim 2^2/T^2$, etc. This is indeed the pattern seen in fig. 2.8.

When the initial conditions are applied through the introduction of the penalty term, eq. (2.44), the degeneracy is lifted as seen in fig. 2.9. There is only one eigenvalue left unaltered, and that corresponds to the solution for the given initial condition. The difference between the lowest and second lowest eigenvalues in this

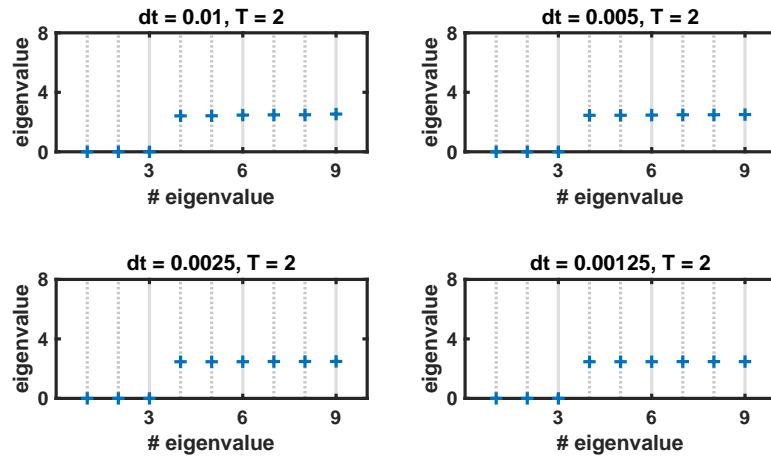


Figure 2.8. The eigenvalues of the McLachlian matrix of the three-level system considered in subsection 2.5.2 prior to applying the initial conditions, shown for various choices of dt while fixing the simulation length T . The 3-fold degeneracy of the ground state is clearly visible. The difference between the eigenvalues of the ground state and the next state scales as $1/T^2$ and therefore independent of the time step.

case is smaller than the difference between the two before the application of the initial conditions (the difference is $\sim 1/T^2$ according to eq.(2.51)). However, we can use the latter as an upper bound, and this upper bound shrinks as $\sim 1/T^2$. This is exactly the reason why the FCIQMC method doesn't work as expected when set to find the ground state of the McLachlian. The ground state is highly degenerate, and the degeneracy worsens as T increases, meaning for practical simulation lengths we have a multifold degeneracy unresolvable by any diffusion method.

Having realised the severity of the quasi-degeneracy of the McLachlian, one might be tempted to resort to one last option: increasing the weight of the penalty term,

$$p \rightarrow wp, \quad (2.52)$$

where p is the penalty term defined in eq. (2.44) and w is the weight. In fig. 2.11 the effect of four different weights on the eigenvalues are shown. It is seen that the difference in eigenvalue between the ground state and the next state does not grow beyond a threshold of the order of α^2 . A similar plot for a two-level system is drawn in fig. 2.12.

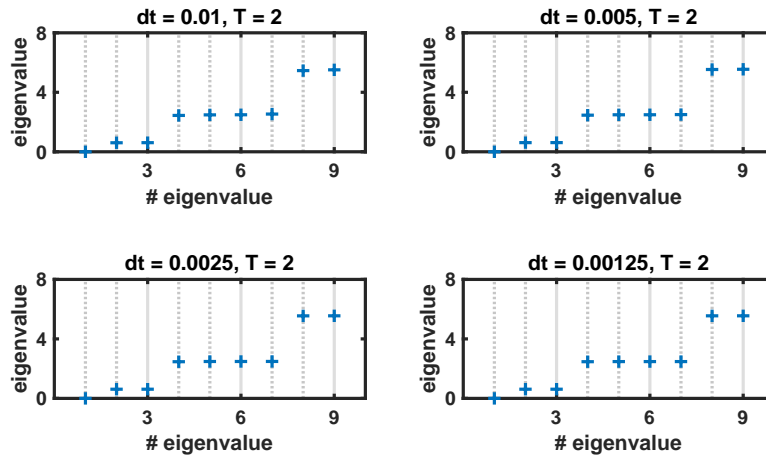


Figure 2.9. The same as fig. 2.8 except that the initial conditions are applied. It is seen that the degeneracy of the ground state is lifted, leaving only the one state that satisfies the initial conditions. Again it is seen that the difference between the eigenvalues of the ground state and the next state is determined by the simulation length T and is therefore independent of dt .

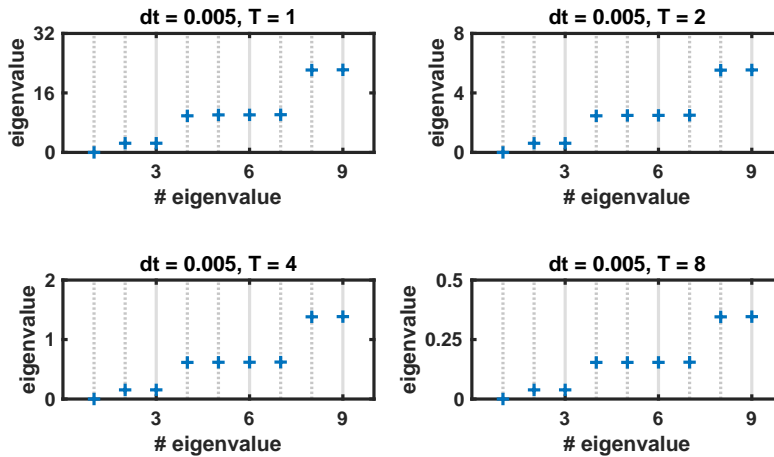


Figure 2.10. The eigenvalues of the McLachlan matrix of the three-level system in subsection 2.5.2 with the initial conditions applied. The different plots are drawn for different values of the simulation duration. It is clearly seen that the difference between the eigenvalues of the ground state and the next state scales as $1/T^2$ (note the limits of the vertical axes).

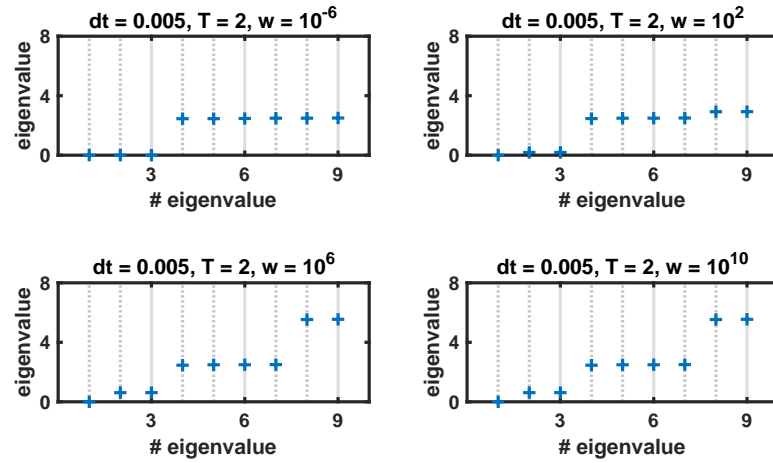


Figure 2.11. The eigenvalues of the McLachlian of the three-level system in subsection 2.5.2 with simulation length $T = 2$. The initial conditions are applied by introducing a penalty term. The weight of this penalty term is increased to investigate if the difference between the eigenvalues of the ground state and the next state increases substantially. It is unfortunately observed that the difference does not grow beyond a negligible value, rendering the FCIQMC method unusable.

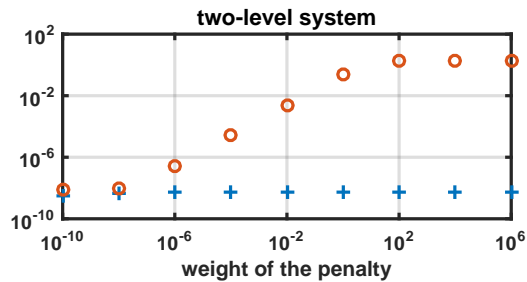


Figure 2.12. The eigenvalues of the ground state (shown by plus signs) and the next state (shown by circles) of the two-level system in subsection 2.5.1 for different weights of the penalty term. The difference reaches a threshold.

2.7. Analysis of the McLachlian eigenvalue problem

2.7.1. Complexity

Despite the failure faced while trying to efficiently solve the McLachlian eigenvalue problem, it is of academic interest to investigate how the complexity of solving this eigenvalue problem compares with that of textbook propagation methods. In the discussion to follow, N represents the number of time steps, and the Hamiltonian matrix is of size $M \times M$.

Using the big O notation [37], the computational complexity of extracting the ground state of a matrix is assumed to be $O(l^\alpha)$, where l is the size of the matrix. It is further assumed that the computational complexity of matrix-vector multiplication is $O(l^\beta)$, where l is the size of the matrix again.

First we will estimate the computational complexity of a typical time propagation scheme, and then compare it with the complexity of finding the ground state of the McLachlian. For the sake of argument, the Hamiltonian is considered to be represented by a full matrix, even though this may not always be the case; it is hoped that both methods, i.e. time propagation and finding the eigenstate of the McLachlian, could benefit from a computational speed-up in the case of a sparse Hamiltonian, and therefore the following argument should be relevant for all types of Hamiltonians.

Each time step of a usual time propagation method contains a vector-matrix multiplication¹³ which in total adds up to N multiplications, thus a complexity of $O(N \times M^\beta)$. On the other hand, the complexity of finding the ground state of the McLachlian is $O((N \times M)^\alpha) = O(N^\alpha \times M^\alpha)$ where the size of the McLachlian is $(N \times M) \times (N \times M)$.

In order for the McLachlian eigenvalue problem to be of a lower complexity, it is required that $N^\alpha M^\alpha < N M^\beta$. To put this inequality in perspective, we consider two different situations:

1. the size of the Hamiltonian remains fixed while the number of time steps

13. It is assumed that an explicit method is used, otherwise the complexity is slightly higher.

increases. In this case the required inequality reduces to

$$\alpha < \beta. \quad (2.53)$$

2. the size of the Hamiltonian increases while the number of time steps is kept fixed, yielding

$$\alpha < 1. \quad (2.54)$$

Finding an algorithm satisfying any of these two requirements would be a big revolution in computer science. Even in the unlikely event that such an algorithm is found, the memory requirements of the new formalism are much more demanding than those of a time propagation scheme. While it is required to save an M -dimensional vector in the case of a time propagation method, the McLachlian method requires a vector of length $N \times M$.

2.7.2. Parallelisability

It is claimed [5] that the McLachlian formalism of the TDSE as an eigenvalue problem has the potential to run on multiple cores in parallel, and therefore by increasing the number of cores it is possible to beat any textbook time propagation method. Unfortunately, it is easily seen in fig. 2.6 that once the initial conditions are applied, the solution starts forming at the neighbouring time step, and then the next time step, and so on. This is a clear manifestation of the fact that the McLachlian is a band matrix. To put it another way, once the initial conditions are applied the task of finding the dynamics of the system is an inherently sequential problem. Thus introducing extra cores will most likely not help as they will have to wait until the correct distribution of walkers reaches the corresponding time step.

2.8. McLachlian eigenvalue problem as a time propagation scheme and vice versa

As it turns out, the McLachlian eigenvalue problem can be reformulated as a time propagation scheme, and conversely all time propagation schemes can be rearranged into a McLachlian matrix.

The time dependent Schrödinger equation is a first order differential equation in time

$$\dot{\psi} = f(t, \psi) \text{ with } \psi(t_0) = \psi_0, \quad (2.55)$$

where $f(t, \psi) = -iH(t)\psi$ and $H(t)$ is an operator. The solution of such an equation is usually found using multistep methods. Multistep methods use the value of the solution from several time steps earlier to approximate the value of the solution at the current time step. The time propagation methods used in physics and engineering are an example of such methods. A linear multistep method uses a linear expression for that purpose. A general linear multistep method using the previous s steps can be written as [35]

$$\begin{aligned} a_0\psi_j + a_1\psi_{j-1} + a_2\psi_{j-2} + \cdots + a_s\psi_{j-s} \\ = \Delta t (b_0f(t_j, \psi_j) + b_1f(t_{j-1}, \psi_{j-1}) + \cdots + b_sf(t_{j-s}, \psi_{j-s})) , \end{aligned} \quad (2.56)$$

where a_i and b_i dictate the behaviour of the method. If $b_0 = 0$, the method is called explicit and can be directly solved. If $b_0 \neq 0$, the method is implicit meaning the value of the solution at the current time is needed to calculate the value of the solution at the current time. Using an implicit method is usually more involved than using an explicit method.

A multistep method is said to be consistent if the local error goes to zero faster than the step size Δt as $\Delta t \rightarrow 0$. This happens if and only if [35]

$$\sum_{k=0}^s a_k = 0, \quad (2.57)$$

and

$$\sum_{k=0}^s b_k + \sum_{k=0}^s ka_k = 0. \quad (2.58)$$

A linear multistep method has order p if the local error is of order $O(h^{p+1})$ as Δt goes to zero. This is equivalent to the following condition on the coefficients of the method [35]

$$q \sum_{k=0}^s k^{q-1} b_k + \sum_{k=0}^s k^q a_k = 0, \quad q = 1, \dots, p. \quad (2.59)$$

Now we can study a few of the usual methods and then go on to express the McLachlan matrix introduced in this chapter as a multistep scheme. The abbreviation $H_j = H(t_j)$ is used below.

2.9. Conclusion

The Full Configuration Interaction Quantum Monte Carlo method (FCIQMC) proves to be a method for finding the ground state of hermitian matrices. Motivated by this fact, we reformulate the time dependent Schrödinger equation (TDSE) as an eigenvalue problem. The operator involved in this problem, termed McLachlian, is shown to have a degenerate ground state which is only lifted when the initial conditions are applied. The initial conditions are applied through the inclusion of a penalty term, lifting all the ground states but the one satisfying the desired initial condition. In this case, by directly diagonalising the McLachlian, it is shown that the ground state indeed represents the solution of the TDSE.

Upon using the FCIQMC method to find the ground state of the McLachlian matrix, however, no convergence is reached. The main reason is revealed to be that the lifting of the degeneracy of the ground state by the penalty term is insufficient to be detectable by usual textbook diffusion methods in a reasonable time.

A comparison between the McLachlian eigenvalue problem and common textbook time propagation schemes shows that the complexity of the McLachlian eigenvalue problem, if to be solved using textbook eigensystem solvers, scales unfortunately very unfavourably with the number of time steps and the size of the system. This revelation along with the failure of the FCIQMC method in extracting the ground state of the McLachlian matrix, show that the new formalism will most likely be only of academic interest in the future.

It appears to be possible to transform any time propagation method into an eigenvalue problem, and conversely to rewrite the McLachlian matrix as a time propagation method. Even though it is a reasonable idea to assume that the time propagation version of the McLachlian matrix developed in this chapter may be used for tackling real problems in the future, in reality it is nothing but a special case of a wide variety of methods called multistep methods.

3. Optimisation of complex systems using Gaussian processes

The field of machine learning has been revolutionising almost every aspect of our lives over the past few years. So much so that it has become almost inevitable to come across a handful of its applications on any given day. Online shopping portals nowadays use sophisticated algorithms to provide their customers with better service. Or news websites use these algorithms to tailor each user's news feed to their interests. Similar algorithms have been successfully used to beat the human expert at a number of different games such as chess [38], and Go [39].

It is only natural that one might think these algorithms should also be given a chance in areas like physics. On the one hand, there is a fundamental difference between the examples mentioned above and physics. While there is no known mathematical formula governing the likes and dislikes of a user, the laws of physics are all formulated in a mathematically sound way. In the former, complex algorithms are used to model the behaviour of a mathematically inexpressible object such as the behaviour of a person, whereas in the latter these algorithms may be used to avoid time consuming simulations. On the other hand, what they both have in common is that the ultimate goal in either case is to have a simple description of the system. For instance, even though a physical system might be difficult and time consuming to simulate, it might be easy to predict its output given the input once a few simulations are done. This simplified description can then be used for other applications.

Machine learning has been used in physics to accelerate simulation of molecular dynamics [40], to improve experimental procedures [41], to extract quantum mechanical properties of materials [42–44], and to approximate complex potentials [45, 46]. There is even a quick review of machine learning from a physicist's perspective [47].

Another potential application is the optimisation of the output of a system. Optimisation algorithms developed in the field of machine learning can and should be used in other areas of science if they show better performance. Given the complexity of today's problems, it is very crucial that we should take advantage of the fastest algorithms available. Fortunately, over the past few years different groups have started the motion using different approaches; for example the particle swarm optimization algorithm has been used to automatically devise quantum measurement policies [48]; the differential evolution algorithm has helped physicists tackle the so-called oracle decision problem [49]; and in ref. [50] different clustering algorithms are used to optimise the assignment fidelity in a superconducting system.

Methods based on Bayes' theorem form a very important class of optimisation algorithms [51]; for a review of Bayesian inference in physics please refer to ref. [52]. Bayes' theorem turns out to be a natural pathway to learning from acquired data, and furthermore a recipe for correcting misjudgements [53, 54]. This theorem is used to develop algorithms which correct their own predictions as new evidence becomes available (loosely speaking, the process of repeatedly learning from new data is referred to as *active learning* [55]). These algorithms are in complete contrast to those which stop improving once the training phase is over; neural network based algorithms are a classic example [56] (this type of learning which completely ignores the effect of the newly acquired data is called *supervised learning* [55]). Historically, most neural networks, once trained, won't change their prediction no matter how many counter examples are found. For a textbook on machine learning please refer to ref. [57].

Bayesian active learning has a very special place in machine learning [58] and in particular in optimisation. A good optimisation unit should make use of all the available data to improve its perception of the latent function (the function to be optimised) as it progresses towards finding the optimum value. For instance, the Newton method [59] uses only the last examined point to suggest the next point, and the Nelder-Mead method [60] (also known as downhill simplex method) uses the last few examined points (the actual number of points depends on the dimension of the problem) to suggest the next point. There are other methods which need all the previously covered points to suggest the next point. Some of these methods describe the data using what is known as non-parametric models. While a parametric model has a few parameters which are set in the training phase once and for all, non-parametric models have no such parameters [57]; each

data point acts as a separate parameter without which the model cannot make predictions. Bayesian non-parametric models have the advantage that they make the most informed guesses about the next point given the available data [61]. But this comes at a price, i.e. all the previously covered points have to be kept, which may become a hindrance as the algorithm progresses and the number of covered points grows.

Complex systems form a main branch of contemporary physics. These are highly nonlinear systems which are usually very hard to tackle analytically [1]. More often than not, in such systems, it is necessary to find the optimum conditions for a particular phenomenon to occur. Thanks to the computational advancements in the past few decades, it is now possible to simulate these systems with relative ease. However, simulation runtimes might still be very long, making optimisation of complex systems demanding, particularly as the complexity of the system grows. This is where Bayesian optimisation algorithms can be of assistance.

In this chapter we will be mainly concerned with the optimisation of different atomic systems using a Bayesian optimisation algorithm based on Gaussian processes (a Gaussian process is an infinite dimensional Gaussian distribution). The Gaussian process is a non-parametric model and proves a very flexible and computationally cheap candidate for modelling generic data [62, 63]. Gaussian processes have found application in weather forecast [64, 65], time analysis [66], etc. Gaussian processes are particularly efficient when the number of data points n isn't too large, as the computational cost of making predictions scales (in the worst case) with n^3 [62] (as will be clear soon, matrix inversion is the bottleneck), which means Gaussian processes lose their appeal as this number grows. In other words, optimisation algorithms based on Gaussian processes may be efficiently used for systems of a relatively small number of variables. This number translates directly into how many data points are needed in order to find the optimum. As a rule of thumb, the more flexible the system, the more data points are needed; this is a manifestation of the so-called *curse of dimensionality* and the interested reader is referred to ref. [67] for a brief discussion.

This chapter is organised such that the reader will be first familiarised with the key concepts underlying Gaussian processes. That is, in section 3.1 a few basic statistical definitions are explained followed by a short discussion of Bayes' theorem. In section 3.2 we will see, through an example, how Bayes' theorem can be used as a guide to learning in the light of incoming data. This is then used to

introduce the concept of a fully Bayesian inference in section 3.3. A short primer on the Gaussian distribution is given in section 3.4 to make the transition to Gaussian processes smoother; the Gaussian process is defined in section 3.5. In section 3.6 the idea of using Gaussian processes as a Bayesian model is developed, which paves the way for Bayesian optimisation methods to be presented in section 3.7. A brief summary of the algorithm developed so far is reiterated in section 3.8. This method is then used to find the optimum settings in four different systems in sections 3.9, 3.10, 3.11, and 3.12. The chapter is concluded in section 3.13.

3.1. Probability: the basics

We will repeatedly refer to some basic statistical quantities, and to make sure there is no ambiguity, a list of them is given here. For the rest of this section, let y_1 , y_2 , etc. be random variables each of which can take on values A , B , etc.

Joint probability

The joint probability is the probability of two events happening at the same time. For instance, $p(y_1 = A, y_2 = B)$ is the probability that y_1 takes on value A , and *simultaneously* y_2 takes on value B .

Marginal probability

The marginal probability is the probability that one particular event happens regardless of the outcome of other events. That is, $p(y_1 = A)$ is the probability that y_1 takes on value A no matter what value y_2 might have. Mathematically speaking, this probability is calculated by integrating variable y_2 out of the joint probability

$$p(y_1) = \int p(y_1, y_2) dy_2. \quad (3.1)$$

Conditional probability

The conditional probability is the probability that one particular event happens given that some other event has happened already, i.e. $p(y_1 = A|y_2 = B)$ is the probability that y_1 is equal to A assuming it is already known that y_2 is equal to B . Note that $p(y_1 = A|y_2 = B)$ is mathematically and conceptually different from $p(y_2 = B|y_1 = A)$; these two are connected by Bayes' theorem as discussed next.

Bayes' theorem

Bayes' theorem is probably best understood through an example. Imagine that one of the symptoms of a particular type of allergy to peanuts is a rash. Without any knowledge of Bayes' theorem, knowing if a patient is allergic or not won't help us have a better assessment of the probability that they will get a rash after having peanuts. That is, our assessment would be the same regardless of the patient's medical history. This is indeed very unfortunate because even though we have access to a relevant piece of information regarding the status of the patient, we cannot incorporate this knowledge into our judgement. But fortunately this is exactly where Bayes' theorem comes to our assistance [68].

Bayes' theorem modifies the probability of an event in the light of newly gathered information. If $p(y_1 = A)$ is the probability of some event prior to any observation, the probability of the same event after observing $y_2 = B$ is given by [68]

$$p(y_1 = A|y_2 = B) = \frac{p(y_2 = B|y_1 = A)p(y_1 = A)}{p(y_2 = B)}. \quad (3.2)$$

where $p(y_1 = A)$ is called the prior, $p(y_2 = B|y_1 = A)$ is called the likelihood, and $p(y_2 = B)$ is called the evidence. The posterior $p(y_1 = A|y_2 = B)$ is in fact the prior improved in view of the observation.

Coming back to our example, suppose the probability that a person gets a rash after eating peanuts is $p(\text{rash}) = 0.05$, and the probability that a person is allergic to peanuts is $p(\text{allergy}) = 0.1$. Further assume that the medical statistics show that 80% of the people getting a rash have reported an allergy to peanuts in their medical history, i.e. $p(\text{allergy}|\text{rash}) = 0.8$. Without Bayes' theorem, all we could say at this point is "everyone is 5% likely to get a rash after eating peanuts". This judgement is certainly not wrong, but fails to take into account the fact that we

know that the patient has an allergy. A better judgement can be made using Bayes' theorem. We will have to consider two cases, 1. the patient is allergic to peanuts,

$$p(\text{rash}|\text{allergy}) = \frac{p(\text{allergy}|\text{rash})p(\text{rash})}{p(\text{allergy})} = \frac{0.8 \times 0.05}{0.1} = 0.4, \quad (3.3)$$

and 2. the patient is not allergic,

$$p(\text{rash}|\text{no allergy}) = \frac{p(\text{no allergy}|\text{rash})p(\text{rash})}{p(\text{no allergy})} = \frac{0.2 \times 0.05}{0.9} = 0.011. \quad (3.4)$$

The first one is the probability that a person with an allergy gets a rash, and the second one is the probability that a person without an allergy gets a rash. It is clear that knowing the medical history of a patient can make a huge difference in our assessment, and simply ignoring it may not be the best way forward.

Perhaps it is constructive to view this example in a slightly different way. At first, we didn't know anything about the patient, and therefore estimated that he would get a rash with a probability of 5%. However, it was then revealed that the patient had an allergy (no allergy) to peanuts. Based on this fact, we were able to have a better assessment of the status of the patient, i.e. he would get a rash with a probability of 40% (1.1%). If more information was made available to us about the patient, we could again update the probability accordingly. This process of incorporating the newly acquired data into our analysis is called learning, and is the topic of the next section.

3.2. Bayes' theorem as a guide to learning

Bayes' theorem lays the perfect mathematical foundation for learning from previous observations. If Bayes' theorem is used successively, every single new piece of information can be taken into account as it becomes available [54]. Imagine we are asked to investigate the fairness of a coin. We initially assume that the probability of a coin coming up heads is given by the Beta distribution¹

$$\text{prior}(p) = \text{Beta}(p; \alpha, \beta) = \frac{1}{\text{B}(\alpha, \beta)} p^{\alpha-1} (1-p)^{\beta-1}. \quad (3.5)$$

1. This choice of prior might seem rather unintuitive, but we'll soon see it will result in an analytic posterior. Even though in general it is by no means necessary to worry about the final form of the posterior, an analytic expression saves us from getting bogged down in the mathematical detail while learning the basics. This particular type of prior has merely been chosen to minimise the mathematical clutter. It should be emphasized that a good prior is always obtained by assuming no more or less than what we know, and it is therefore imperative

where $B(\alpha, \beta)$ is the beta function. In the case of no prior information whatsoever, it is best to use an uninformative prior [69], which is a uniform distribution in this case, and can be obtained by setting $\alpha = 1$ and $\beta = 1$ in the Beta distribution, i.e. $Beta(p; 1, 1)$.

After flipping the coin, Bayes' theorem enables us to learn about the probability that the coin comes up heads or tails next time, that is, it refines our belief. If the coin is tossed n -times, and the observed sequence is h, t, t, h, \dots consisting of H times heads and T times tails, the likelihood is given by the Bernoulli distribution

$$\text{likelihood}(h, t, t, h, \dots | p) = \text{Bernoulli}(p; H, T) = p^H (1 - p)^T, \quad (3.6)$$

in terms of which the evidence can be expressed

$$\begin{aligned} \text{evidence}(h, t, t, h, \dots) &= \int \text{likelihood}(h, t, t, h, \dots | p) \times \text{prior}(p) dp \\ &= \frac{B(\alpha + H, \beta + T)}{B(\alpha, \beta)}. \end{aligned} \quad (3.7)$$

Putting everything together, it becomes evident that the posterior is a beta distribution with modified parameters

$$\text{posterior}(p | h, t, t, h, \dots) = \text{Beta}(p; \alpha + H, \beta + T). \quad (3.8)$$

Using this posterior, it is easy to see what happens to our belief after a large number of coin flips. For instance, if after flipping a coin 10,000 times it turns out the coin has come up heads 4,899 times and has come up tails 5,101 times, a bookmaker is best off taking bets consistent with the updated probability distribution, i.e. $Beta(p; 1 + 4899, 1 + 5101)$.

The act of using Bayes' theorem is referred to as "inference" in the statistics community [70]. Contrary to the example above, it is not always straightforward to use Bayes' theorem. For instance, when the likelihood is not exactly known, or when the data is tainted with noise. In the next section we will see how such cases can be tackled in a fully Bayesian manner.

that we understand all the consequences a particular choice of prior function might have. For a more detailed discussion please refer to ref. [69].

3.3. Fully Bayesian inference

One of the cornerstones of Bayesian statistics is that the uncertainty stemming from our lack of knowledge is treated no differently from the uncertainty of the outcome of a random process. Each type of uncertainty is assigned a probability distribution and is treated as such [51]. In other words, the behaviour of an unknown function might be described by a *probabilistic model*, the same way the behaviour of a random variable would be. This probabilistic model may have several parameters which are referred to as hyperparameters and collectively denoted by θ ; the hyperparameters are just as important in making predictions as the model itself. Any attempt to make a prediction should take into account all sources of uncertainty by averaging over their corresponding probability distributions. For instance, if an exact knowledge of the values of the hyperparameters is not available, the following integral should be evaluated every time a prediction is made [51],

$$p(h) = \int p(h|\theta) p(\theta) d\theta, \quad (3.9)$$

where h is the quantity of interest, and $p(\theta)$ is the prior over hyperparameters containing our belief or expert knowledge over different choices of hyperparameters. $p(h|\theta)$ is provided by the model. There may be more than one model to describe the data, e.g. a mean-field model, a perturbative model, etc., each of which has its own hyperparameters. To account for all these models in a fully Bayesian manner, there should be a similar sum over different models [51],

$$p(h) = \sum_{\mathcal{M}} \int p(h|\theta, \mathcal{M}) p(\theta|\mathcal{M}) p(\mathcal{M}) d\theta, \quad (3.10)$$

where $p(\mathcal{M})$ is the prior and embodies our expert knowledge regarding the models; for instance, based on previous observations one user might give a certain model a higher probability. $p(\theta|\mathcal{M})$ is the prior on the hyperparameters of model \mathcal{M} , and $p(h|\theta, \mathcal{M})$ is the prediction made by the model given the hyperparameters. It is indeed this summation over all hyperparameters and models that prevents a fully Bayesian approach from under- or overfitting [53]. In theory, in a situation where there are several potential models, all of the models should be considered just as explained. In practice however, this proves very resource-intensive, and there are several ways to simplify this integral, the simplest one of which is to consider only the most probable model. There will be more on this in subsection 3.3.4.

Having a final probability distribution at hand, we can now go ahead and apply the Bayes' theorem every time a new observation/measurement is made,

$$p(h|\mathcal{D}) = \sum_{\mathcal{M}} \int p(h|\theta, \mathcal{M}, \mathcal{D}) p(\theta|\mathcal{M}, \mathcal{D}) p(\mathcal{M}|\mathcal{D}) d\theta, \quad (3.11)$$

where \mathcal{D} represents the acquired data. We will soon see how this expression can be evaluated.

A very important source of uncertainty in many areas of science is the measurement error. Measurement is inherently prone to error, and needs to be dealt with correctly in a Bayesian setting. An integral is needed over all the values of the ground truth \mathcal{D} that may have given rise to the “reading” on the measuring device,

$$p(h|\text{reading}) = \int p(h|\mathcal{D}) p(\mathcal{D}|\text{reading}) d\mathcal{D}, \quad (3.12)$$

where $p(\mathcal{D}|\text{reading})$ is related to the uncertainty of the measurement. If eq. (3.11) is plugged in this equation we get

$$p(h|\text{reading}) = \sum_{\mathcal{M}} \int p(h|\theta, \mathcal{M}, \mathcal{D}) p(\theta|\mathcal{M}, \mathcal{D}) p(\mathcal{M}|\mathcal{D}) p(\mathcal{D}|\text{reading}) d\theta d\mathcal{D}. \quad (3.13)$$

This is called the posterior predictive probability distribution, and is the probability of observing h given the outcome of the previous measurements. Before we can actually use this probability there are three terms that we need to evaluate, i.e. $p(\theta|\mathcal{M}, \mathcal{D})$, $p(\mathcal{M}|\mathcal{D})$, and $p(\mathcal{D}|\text{reading})$. These three terms give rise to several levels of Bayesian inference which we will deal with in the forthcoming subsections [62].

3.3.1. Inferring the influence of uncertainty in measurement

The outcome of any measuring device is best described through a probability distribution whose standard deviation represents the precision of the device. If \mathcal{D} is the ground truth, the outcome is given by conditional probability $p(\text{reading}|\mathcal{D})$, where “reading” is what appears on the device. The normal distribution is a very common choice for this distribution, i.e. $p(\text{reading}|\mathcal{D}) \sim \mathcal{N}(\mathcal{D}, \sigma^2)$ with σ being the precision of the measuring device.

The probability distribution $p(\mathcal{D}|\text{reading})$ appearing in eq. (3.13) is related to the aforementioned probability distribution through Bayes' theorem

$$p(\mathcal{D}|\text{reading}) = \frac{p(\text{reading}|\mathcal{D}) p(\mathcal{D})}{p(\text{reading})}, \quad (3.14)$$

where $p(\mathcal{D})$ is usually substituted for by a uniform distribution unless there is a preference for a particular value of the ground truth. The denominator is given by

$$p(\text{reading}) = \int p(\text{reading}|\mathcal{D}) p(\mathcal{D}) d\mathcal{D}. \quad (3.15)$$

This integral is almost never evaluated, as it only serves as a normalising factor. Instead, the nominator is normalised to one. Putting everything together gives

$$p(\mathcal{D}|\text{reading}) \sim \mathcal{N}(\mathcal{D}, \sigma^2). \quad (3.16)$$

It is worth reiterating once more that $p(\mathcal{D}|\text{reading})$ and $p(\text{reading}|\mathcal{D})$ are conceptually different quantities. $p(\mathcal{D}|\text{reading})$ is the probability that \mathcal{D} is the ground truth given the reading, while $p(\text{reading}|\mathcal{D})$ is the probability that the measuring device shows “reading” if the ground truth is \mathcal{D} . These two might not always be identical.

3.3.2. Inferring the hyperparameters

The second probability distribution that needs evaluating is $p(\theta|\mathcal{M}, \mathcal{D})$. This is the probability that θ is the correct choice of hyperparameters given the model and the ground truth. This distribution is connected to the prediction made by the model with the help of Bayes' theorem [62]

$$p(\theta|\mathcal{M}, \mathcal{D}) = \frac{p(\mathcal{D}|\mathcal{M}, \theta) p(\theta|\mathcal{M})}{p(\mathcal{D}|\mathcal{M})}, \quad (3.17)$$

where $p(\mathcal{D}|\mathcal{M}, \theta)$ is the prediction made by the model. Probabilistic models describe the observed data using probability distributions. Later in the chapter we will see how Gaussian processes can be used here as a model. $p(\theta|\mathcal{M})$ is usually chosen to be a uniform distribution prior to any observations unless there are some constraints the hyperparameters have to satisfy, e.g. being non-negative. As observations are made, the posterior from the last observation serves as the prior for the next observation. The denominator $p(\mathcal{D}|\mathcal{M})$ is obtained by an integral over

the hyperparameters

$$p(\mathcal{D}|\mathcal{M}) = \int p(\mathcal{D}|\mathcal{M}, \theta) p(\theta|\mathcal{M}) d\theta. \quad (3.18)$$

Unfortunately, this integral can almost never be exactly evaluated analytically or numerically. But fortunately, this integral behaves merely as a normalising factor, and there are several approximate methods in which this integral is not evaluated at all. Some of these approximate methods are discussed in section 3.6.2.

3.3.3. Inferring the covariance function

Lastly, the probability $p(\mathcal{M}|\mathcal{D})$ is evaluated by yet another Bayes' theorem [62]

$$p(\mathcal{M}|\mathcal{D}) = \frac{p(\mathcal{D}|\mathcal{M}) p(\mathcal{M})}{p(\mathcal{D})}, \quad (3.19)$$

where $p(\mathcal{D}|\mathcal{M})$ was evaluated in the previous subsection. In fact, $p(\mathcal{M}|\mathcal{D})$ is the probability that model \mathcal{M} describes the data. A uniform distribution is usually chosen for $p(\mathcal{M})$ if there is no preference for any model. After each observation, the posterior of the last observation is used as the prior of the next observation. And as before, the denominator is given by a summation

$$p(\mathcal{D}) = \sum_{\mathcal{M}} p(\mathcal{D}|\mathcal{M}) p(\mathcal{M}). \quad (3.20)$$

As it will be explained in the next section, this summation is rarely evaluated in practical applications. Usually only the most probable model is chosen.

3.3.4. Model selection

Model selection is the process of evaluating the probability that each model describes the data, as a result of which only one is selected [62]. Even though the name might not suggest directly, this process also covers the tuning of hyperparameters. Model selection only comes about because we cannot consider every possible model and therefore have to settle for the most promising one.

Evaluating the integral in eq. (3.13) is a very difficult task if done fully, as there may always be many promising models. For example, a linear model, a second-order model, and a third-order model might all seem reasonable for a given set

of data. Furthermore, if the number of data points is very large, evaluating three models prior to every prediction is hard work, not forgetting the integrals over their respective hyperparameters. For this reason, the most promising model is very often selected as the only model, making the summation over different models disappear.

To evaluate the probability of a given model describing the data, we will take advantage of a probability evaluated in the previous subsection, that is $p(\mathcal{M}|\mathcal{D})$. This is the probability that model \mathcal{M} is the correct model given the data, and using Bayes' theorem it can be written in terms of $p(\mathcal{D}|\mathcal{M})$, and $p(\mathcal{M})$. Assuming there is no preference for any particular model, $p(\mathcal{M})$ is a uniform distribution. Meaning, the maximum of $p(\mathcal{D}|\mathcal{M})$ coincides with the maximum of $p(\mathcal{M}|\mathcal{D})$. Hence it suffices to find the model that maximises $p(\mathcal{D}|\mathcal{M})$. Sometimes a generic model is chosen merely because of its simplicity and flexibility, in spite of its lower probability of describing the observations.

The same can be said about hyperparameters. It is virtually impossible to carry out an integral over all the hyperparameters if there are more than a handful of them. That said, merely choosing the most probable hyperparameters as a substitute for the integral can lead to pathologies [53, 70]; for example overfitting. Fortunately, there are some workarounds which we will discuss in a later section. Should there be any need for choosing one particular set of hyperparameters, the safest bet is to find the most probable hyperparameters given the data. Similarly to what was done earlier, we take advantage of the probability introduced in the last section, i.e. $p(\theta|\mathcal{M}, \mathcal{D})$. If there is no particular prior on hyperparameters, i.e. $p(\theta|\mathcal{M})$ is a uniform distribution, the maximum of $p(\mathcal{D}|\mathcal{M}, \theta)$ matches the maximum of $p(\theta|\mathcal{M}, \mathcal{D})$. Given the model, it should be fairly straightforward to evaluate $p(\mathcal{D}|\mathcal{M}, \theta)$ for a particular set of hyperparameters. Later in the chapter we will see how Gaussian processes can be used as a model, and how this quantity is evaluated.

3.4. Gaussian distribution

To see how the Gaussian process emerges naturally from the Gaussian distribution, we will first have a look at the d -dimensional Gaussian distribution. The Gaussian

distribution is fully determined by its mean vector $\vec{\mu}$, and covariance matrix Σ ,

$$\vec{y} \sim \mathcal{N}(\vec{\mu}, \Sigma) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp\left(-\frac{1}{2}(\vec{y} - \vec{\mu})^\top \Sigma^{-1}(\vec{y} - \vec{\mu})\right), \quad (3.21)$$

where d is the dimension of the distribution, Σ is the covariant matrix, and $|\dots|$ should be understood as the determinant of the matrix. \vec{y} is the outcome of the distribution when a draw is made. The prefactor makes sure the distribution is normalised.

3.4.1. Conditional and marginal distributions

The Gaussian distribution is in many ways a very convenient distribution to work with. In particular, we will take advantage of the fact that the conditional and marginal distributions of the Gaussian distribution are also Gaussian. In the discussion to follow in the subsequent sections, we will repeatedly make use of these two distributions.

If \vec{y} consists of two parts,

$$\vec{y} = \begin{bmatrix} \vec{y}_1 \\ \vec{y}_2 \end{bmatrix}, \quad (3.22)$$

and the value of one part, say \vec{y}_2 , is known, the distribution of the value of the other part is given by what is known as the conditional distribution,

$$\vec{y}_1 | \vec{y}_2 \sim \mathcal{N}(\vec{\mu}^*, \Sigma^*), \quad (3.23)$$

which is a Gaussian distribution with mean

$$\vec{\mu}^* = \vec{\mu}_1 + \Sigma_{12} \Sigma_{22}^{-1} (\vec{y}_2 - \vec{\mu}_2), \quad (3.24)$$

and covariance

$$\Sigma^* = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}. \quad (3.25)$$

$\vec{\mu}_1$ and $\vec{\mu}_2$ are the two constituent vectors of a similar partitioning of the total mean vector

$$\vec{\mu} = \begin{bmatrix} \vec{\mu}_1 \\ \vec{\mu}_2 \end{bmatrix}, \quad (3.26)$$

Σ_{ij} are the constituent matrices of the total covariance matrix

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \quad (3.27)$$

and Σ_{22}^{-1} is the inverse of Σ_{22} .

The distribution of \vec{y}_1 regardless of the value of \vec{y}_2 is called the marginal distribution, and is evaluated by integrating out \vec{y}_2 ,

$$\vec{y}_1 \sim \mathcal{N}(\vec{\mu}_1, \Sigma_{11}), \quad (3.28)$$

which is again a Gaussian distribution. $\vec{\mu}_1$ and Σ_{11} are defined as before.

3.5. Gaussian processes

Sometimes it is required to increase the dimension of the d -dimensional Gaussian distribution to infinity. A Gaussian process is nothing but an infinite-dimensional Gaussian distribution, meaning each draw from a Gaussian process is an infinite-dimensional vector, just as each draw from a d -dimensional Gaussian distribution is a d -dimensional vector [62]. Vectors of limited dimension can be fairly easily visualised, but to visualise an infinite-dimensional vector, we will have to seek the assistance of functions.

If each component of an infinite-dimensional vector is represented by a point on the x -axis, the entire vector can be depicted as a function. And conversely, any function defined along the x -axis will be equivalent to an infinite-dimensional vector. This is indeed how infinite dimensional vectors are represented in the discussion of Gaussian processes.

If the value of the function at each point is determined by a draw from a one-dimensional Gaussian distribution with some mean and covariance, every time that a new draw is made on the entire axis, a new function is born. Therefore, it is more intuitive to think of the Gaussian process as a device capable of generating random functions, just as the Gaussian distribution is a device capable of generating random numbers. It is further possible to tangle the mean and covariance of each point on the x -axis to those of other points by introducing a mean function $m(x)$ and a two-variable covariance function $k(x_1, x_2)$,

$$f(x) \sim \mathcal{GP}(m(x), k(x_1, x_2)), \quad (3.29)$$

where \mathcal{GP} stands for Gaussian Process. $f(x)$ is the infinite-dimensional vector output of the Gaussian process which is represented as a function here. The covariance

function determines how the value of the function at one point affects the distribution of the value of the function at other points. Different covariance functions have different properties and make different assumptions. For a list of different covariance functions and their underlying assumptions please refer to ref. [62]. The mean and covariance functions may have some parameters which are commonly referred to as hyperparameters in the literature.

Very similarly to the Gaussian distribution, it is possible to calculate the conditional and marginal distributions of the Gaussian process.

3.5.1. Conditional and marginal distributions

In the discussion to appear in the next section, we will need the conditional and marginal distributions of the Gaussian process. Fortunately, it is very easy to generalise these distributions from the Gaussian distribution introduced in the previous section to the case of the Gaussian process.

To calculate the conditional probability, we assume that some observations have been made and we are interested in the distribution of the unobserved points affected by the observations. The data gathered so far is collectively represented by \mathcal{D} , consisting of matrix \mathbf{X} , whose rows $\vec{x}_1, \vec{x}_2, \dots$ are the coordinates of the observed points, and vector \vec{Y} whose elements y_1, y_2, \dots are the corresponding observed values. The conditional probability is given by a new Gaussian process [62] (the arrow accents on x_i are omitted for better readability)

$$f^*(x)|\mathcal{D} \sim \mathcal{GP}(m^*(x), k^*(x_1, x_2)), \quad (3.30)$$

where

$$m^*(x) = m(x) + \vec{\Sigma}(\mathbf{X}, x)^\top \Sigma^{-1}(\vec{Y} - \vec{\mu}) \quad (3.31)$$

$$k^*(x_1, x_2) = k(x_1, x_2) - \vec{\Sigma}(\mathbf{X}, x_1)^\top \Sigma^{-1} \vec{\Sigma}(\mathbf{X}, x_2). \quad (3.32)$$

$\Sigma(\mathbf{X}, \mathbf{X})$ is called the covariance matrix and is given by the covariance function evaluated at the observed data points, i.e. $\Sigma_{ij} = k(x_i, x_j)$. $\vec{\Sigma}(\mathbf{X}, x_j)$ is the j -th column of the covariance matrix, and the inverse of the covariance matrix is represented by Σ^{-1} (for brevity, we use Σ^{-1} instead of $\Sigma^{-1}(\mathbf{X}, \mathbf{X})$). $\vec{\mu} = m(\mathbf{X})$ is called the mean vector and is given by the mean function evaluated at the observed data points.

The marginal probability at point x regardless of all other points is given by [62]

$$f \sim \mathcal{N}(m(x), k(x, x)), \quad (3.33)$$

where \mathcal{N} represents a *one-dimensional Gaussian distribution* as defined in the previous section. If the marginal probability is desired at two points simultaneously, a two-dimensional Gaussian distribution has to be evaluated,

$$\begin{bmatrix} f_1 \\ f_2 \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} m(x_1) \\ m(x_2) \end{bmatrix}, \begin{bmatrix} k(x_1, x_1), k(x_1, x_2) \\ k(x_2, x_1), k(x_2, x_2) \end{bmatrix} \right). \quad (3.34)$$

3.6. Gaussian processes as a Bayesian model

Gaussian processes are a powerful and flexible tool to model data. As was briefly pointed out before, the probability distribution $p(\mathcal{D}|\theta, \mathcal{M})$ is given by the model. And in the case of a model based on Gaussian processes [62], we have

$$p(\mathcal{D}|\theta, \mathcal{M}) = \frac{1}{\sqrt{(2\pi)^d |\Sigma(\mathbf{X}, \mathbf{X})|}} \exp \left(-\frac{1}{2} (\vec{Y} - \vec{\mu})^\top \Sigma^{-1} (\vec{Y} - \vec{\mu}) \right), \quad (3.35)$$

where all the quantities are as defined in the previous section. θ represents the parameters in the mean and covariance functions. The exact choice of the covariance and mean functions will be explained later. But before we can use Gaussian processes as a prospective model, there are a few steps that we will have to take. These steps are the three levels of inference explained in section 3.3. These are followed by different approximate methods that help evaluate the integral in eq. (3.13). The approximate methods are explained below under their corresponding level of inference.

It may sound unintuitive, but it turns out it is conceptually easier to start from the third inference and continue towards the other levels of inference.

3.6.1. Inferring the covariance function

Starting from the third inference, if we restrict ourselves to models based on Gaussian processes, different covariance functions represent different models. According to the rules of Bayesian statistics, these models have to be summed over. However, this proves a formidable (if not impossible) task in practice forcing one to include

only as many covariance functions as there are computational resources for. Quite often only one covariance function is considered and the summation over different models in eq. (3.13) is ignored. The choice of the covariance function depends to a large extent on the problem under study. For instance, if a periodic behaviour is expected, a covariance function with a periodic component may display better predicting power. In cases where there is no prior knowledge available about the underlying structure of the data, it is best to choose a generic covariance function. The most common generic covariance function in the literature is the so-called squared exponential covariance function [62]

$$k(\vec{x}, \vec{x}') = \sigma_f^2 \exp \left(-\frac{1}{2} \sum_i \frac{(x_i - x'_i)^2}{\lambda_i^2} \right), \quad (3.36)$$

where σ_f is an overall scaling prefactor, and λ_i are known as the covariance lengths. In fig. 3.1 the predictive probability is shown for two different choices of λ . A larger covariance length means the Gaussian process is more confident making predictions at points further away from the observed points compared to a smaller covariance length. The scaling prefactor along with the covariance lengths collectively form the hyperparameters of the squared exponential covariance function. The analytic simplicity and numerical flexibility of the squared exponential covariance function have made it far and away the most common choice in the literature. The squared exponential covariance function makes the assumption that the output of the Gaussian process can be differentiated an infinite number of times. Since this may not be a realistic assumption to make about a physical system, it is necessary to increase the uncertainty involved in the first level of inference to compensate for any artefact this assumption might have (there will be more on this in subsection 3.6.2).

3.6.2. Inferring the hyperparameters

Moving on to the second inference, an integral over all the hyperparameters is required according to eq. (3.13). This integral is virtually impossible to evaluate due to the large number of hyperparameters and the relatively large space they sweep. There are, however, approximate methods to deal with these integrals. All these approximations are, one way or another, reliant on the maximum of the posterior probability distribution of the hyperparameters given by eq. (3.17). For numerical stability, it is customary to work with the log of the posterior,

$$\log p(\theta|\mathcal{M}, \mathcal{D}) = \log p(\mathcal{D}|\theta, \mathcal{M}) + \log p(\theta|\mathcal{M}) - \log p(\mathcal{D}|\mathcal{M}), \quad (3.37)$$

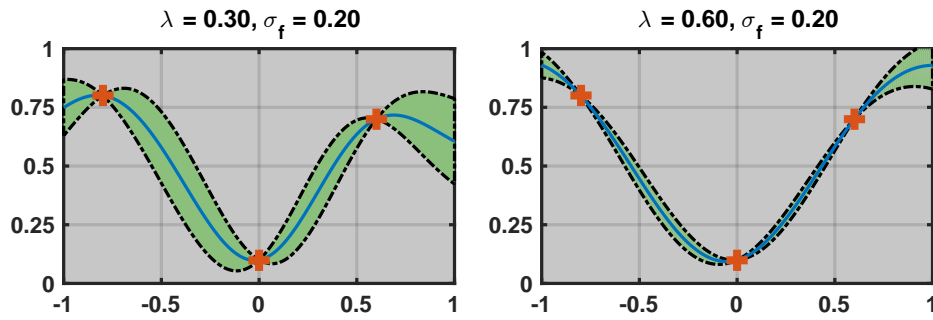


Figure 3.1. The effect of the covariance length on the prediction made by a Gaussian process. The covariance length used in the plot on the left hand side is smaller than that used in the plot on the right hand side. The larger the covariance length, the more confident the Gaussian process becomes about points further away from the data.

where $p(\theta|\mathcal{M})$ is the prior. The likelihood $p(\mathcal{D}|\theta, \mathcal{M})$ is a measure of how likely it is to observe \mathcal{D} given the model and its hyperparameters. Note that the likelihood is not a probability function of the hyperparameters, but the observed data points. The evidence $p(\mathcal{D}|\mathcal{M})$ is independent of the choice of hyperparameters and therefore a constant as long as the model and \mathcal{D} remain unchanged.

In the following, four different approximate methods are discussed, each of which is an attempt at evaluating the integration over the hyperparameters in the predictive probability distribution given by eq. (3.13).

Maximum Likelihood Estimate (MLE)

The simplest way to deal with the integral over hyperparameters is to replace the posterior with a delta function, i.e. $p(\theta|\mathcal{M}, \mathcal{D}) \rightarrow \delta(\theta - \theta_{MLE})$ [71] (we will soon see how θ_{MLE} is defined). This substitution is justified by the idea that the posterior will most likely become a delta function in the limit of big data. This approximation is easily prone to over-/underfitting and should therefore be used with care. If the prior is a uniform distribution, the only term in eq. (3.37) changing as a function of the hyperparameters is the likelihood,

$$\mathcal{L}(\theta, \mathcal{M}|\mathcal{D}) \equiv p(\mathcal{D}|\theta, \mathcal{M}). \quad (3.38)$$

In this case, to find the maximum of the posterior probability, it suffices to find the maximum of the likelihood. The maximum can be easily found using textbook

optimisation algorithms. According to eq. (3.35) the log-likelihood of the Gaussian process is

$$\begin{aligned} \log \mathcal{L}(\theta, \mathcal{M}|\mathcal{D}) &= -\frac{d}{2} \log 2\pi \\ &\quad -\frac{1}{2} \log |\Sigma(\mathbf{X}, \mathbf{X})| \\ &\quad -\frac{1}{2} (\vec{Y} - \vec{\mu})^\top \Sigma^{-1} (\vec{Y} - \vec{\mu}). \end{aligned} \quad (3.39)$$

The set of hyperparameters maximising this likelihood will be called θ_{MLE} (short for Maximum Likelihood Estimate), and is the point appearing in the delta function above. By inserting the delta function in the predictive probability integral in eq. (3.13) we get

$$\begin{aligned} p(h|\mathcal{M}, \mathcal{D}) &= \int p(h|\theta, \mathcal{M}, \mathcal{D}) p(\theta|\mathcal{M}, \mathcal{D}) d\theta \\ &\approx \int p(h|\theta, \mathcal{M}, \mathcal{D}) \delta(\theta - \theta_{\text{MLE}}) d\theta \\ &= p(h|\theta_{\text{MLE}}, \mathcal{M}, \mathcal{D}). \end{aligned} \quad (3.40)$$

And when a Gaussian process is used as the model

$$p(h|\theta_{\text{MLE}}, \mathcal{M}, \mathcal{D}) = \mathcal{GP}(m_{\text{MLE}}^*(x), k_{\text{MLE}}^*(x_1, x_2)), \quad (3.41)$$

where $m_{\text{MLE}}^*(x)$ and $k_{\text{MLE}}^*(x_1, x_2)$ are the mean and covariance functions (eqs. (3.31) and (3.32)) evaluated at $\theta = \theta_{\text{MLE}}$.

Maximum A posteriori Probability (MAP)

The previous approximation was based on the assumption that the prior is uniform. But this may not always be the case, for example sometimes the hyperparameters cannot be negative, hence a non-uniform prior. If the prior is non-uniform, the maximum of the posterior should be used (denoted hereafter by θ_{MAP}) [71],

$$\log p(\theta|\mathcal{M}, \mathcal{D}) = \log p(\mathcal{D}|\theta, \mathcal{M}) + \log p(\theta|\mathcal{M}) - \log p(\mathcal{D}|\mathcal{M}). \quad (3.42)$$

But since the last term on the right hand side has no dependence on θ , it suffices to find the maximum of $\log p(\mathcal{D}|\theta, \mathcal{M}) + \log p(\theta|\mathcal{M})$. Again the maximum can be found using common textbook optimisation algorithms.

Similarly, the posterior is replaced with a delta function in this approximation, i.e. $p(\theta|\mathcal{M}, \mathcal{D}) \rightarrow \delta(\theta - \theta_{\text{MAP}})$. Again by inserting the delta function in the predictive probability integral eq.(3.13) we get

$$\begin{aligned} p(h|\mathcal{M}, \mathcal{D}) &= \int p(h|\theta, \mathcal{M}, \mathcal{D}) p(\theta|\mathcal{M}, \mathcal{D}) d\theta \\ &\approx \int p(h|\theta, \mathcal{M}, \mathcal{D}) \delta(\theta - \theta_{\text{MAP}}) d\theta \\ &= p(h|\theta_{\text{MAP}}, \mathcal{M}, \mathcal{D}), \end{aligned} \quad (3.43)$$

which in the case of a Gaussian process yields

$$p(h|\theta_{\text{MAP}}, \mathcal{M}, \mathcal{D}) = \mathcal{GP}(m_{\text{MAP}}^*(x), k_{\text{MAP}}^*(x_1, x_2)), \quad (3.44)$$

where $m_{\text{MAP}}^*(x)$ and $k_{\text{MAP}}^*(x_1, x_2)$ are the mean and covariance functions, eqs. (3.31) and (3.32), evaluated at $\theta = \theta_{\text{MAP}}$.

Affine transformation: Linear Mean + Constant Covariance

As will be seen, it is in fact the analytic simplicity of this approximation that has made it particularly appealing. This approximation has two steps [72],

1. Instead of a delta function, the posterior is approximated by a normal distribution about θ_{MAP} , the covariance matrix of which is the Hessian of the log posterior with respect to the hyperparameters at θ_{MAP} ,

$$p(\theta|\mathcal{M}, \mathcal{D}) \approx \mathcal{N}(\theta; \theta_{\text{MAP}}; \Sigma_{\text{MAP}}), \quad (3.45)$$

where $\Sigma_{\text{MAP}}^{-1} = -\partial^2 \log p(\theta|\mathcal{M}, \mathcal{D}) / \partial \theta_i \partial \theta_j |_{\theta_{\text{MAP}}}$. In the case of a uniform prior on hyperparameters, we get $\theta_{\text{MAP}} \rightarrow \theta_{\text{MLE}}$ and $\Sigma_{\text{MAP}}^{-1} \rightarrow \Sigma_{\text{MLE}}^{-1} = -\partial^2 \log \mathcal{L}(\theta, \mathcal{M}|\mathcal{D}) / \partial \theta_i \partial \theta_j |_{\theta_{\text{MLE}}}$.

2. Furthermore, it is assumed that the mean function of the Gaussian process is linear in θ , and the covariance matrix is constant throughout the hyperparameter space,

$$\mathcal{GP}(m^*(x), k^*(x_1, x_2)) \approx \mathcal{GP}(a(x) \cdot \theta + b(x), k_{\text{MAP}}^*(x_1, x_2)), \quad (3.46)$$

where

$$a(x) = \left. \frac{\partial m^*(x)}{\partial \theta} \right|_{\theta=\theta_{\text{MAP}}} \quad (3.47)$$

$$b(x) = m_{\text{MAP}}^*(x) - \left. \frac{\partial m^*(x)}{\partial \theta} \right|_{\theta=\theta_{\text{MAP}}} \cdot \theta_{\text{MAP}}. \quad (3.48)$$

Since θ represents all the hyperparameters collectively, it should be understood as rather a vector here, and consequently \cdot should be understood as a dot product.

By inserting the above relations in the predictive probability integral in eq. (3.13) we get

$$\begin{aligned}
 p(h|\mathcal{M}, \mathcal{D}) &= \int p(h|\theta, \mathcal{M}, \mathcal{D}) p(\theta|\mathcal{M}, \mathcal{D}) d\theta \\
 &\approx \int \mathcal{GP}(a(x)\theta + b(x), k_{\text{MAP}}^*(x_1, x_2)) \mathcal{N}(\theta; \theta_{\text{MAP}}; \Sigma_{\text{MAP}}) d\theta \quad (3.49) \\
 &= \mathcal{GP}(a(x)\theta_{\text{MAP}} + b(x), k_{\text{MAP}}^*(x_1, x_2) + a(x)\Sigma_{\text{MAP}}a(x)) \\
 &= \mathcal{GP}(m_{\text{MAP}}^*(x), k_{\text{MAP}}^*(x_1, x_2) + a(x)\Sigma_{\text{MAP}}a(x)),
 \end{aligned}$$

where the new mean function is identical to that of the previous approximation, but the covariance function has an extra term. This extra term is always positive, meaning the uncertainty has increased compared to the MAP approximation.

As mentioned before, in Bayesian statistics whenever the integral over different models or different hyperparameters is approximated by a point estimate, there is a risk that the predictive distribution might behave unexpectedly in some situations. For instance, if the peak of the likelihood happens to be in a relatively flat region, Σ_{MAP} will be a matrix with very large elements. Inserting such a matrix in eq. (3.49) leads to huge uncertainties which can obviously endanger the stability of the method. However, if the training set is large enough, it is almost always guaranteed that the likelihood has a sharp peak. In fact, as the number of observations grows, this peak becomes even sharper. For this reason, we will safely use this approximation later in this chapter to evaluate the predictive probability integral for practical applications.

Taylor Expansion

In 2013 an extension of the previous approximation was put forward by R. Garnett *et al.* [73]. However, the derivation seems to lack a mathematically sound foundation. We present a similar but rigorous extension here which leads to a thorough final expression. Similarly to the previous approximation, the posterior $p(\theta|\mathcal{M}, \mathcal{D})$ is approximated by a Gaussian distribution centred at θ_{MAP} , with covariance matrix Σ_{MAP} calculated just as before. The difference, however, is that the Gaussian

process is Taylor expanded to third order about θ_{MAP} ,

$$\begin{aligned} \mathcal{GP}(m^*(x), k^*(x_1, x_2)) &\approx \mathcal{GP}(m_{\theta_{\text{MAP}}}^*(x), k_{\theta_{\text{MAP}}}^*(x_1, x_2)) \\ &\quad + f(\theta) (\theta - \theta_{\text{MAP}}) \\ &\quad + g(\theta) (\theta - \theta_{\text{MAP}})^2 \\ &\quad + h(\theta) (\theta - \theta_{\text{MAP}})^3 \\ &\quad + O(\theta^4), \end{aligned} \tag{3.50}$$

which is then inserted into the expression for the predictive probability, eq. (3.13). By defining a new Gaussian process whose mean and variance at each point are equal to the mean and variance of the integral (a procedure known as *moment matching* in the literature [74]) we obtain

$$\begin{aligned} p(y_*|x_*, \mathcal{M}, \mathcal{D}) &= \int p(y_*|x_*, \theta, \mathcal{M}, \mathcal{D}) p(\theta|\mathcal{M}, \mathcal{D}) d\theta \\ &\approx \int (\mathcal{GP}_{\theta_{\text{MAP}}} + f(\theta) (\theta - \theta_{\text{MAP}}) + \dots) \mathcal{N}(\theta; \theta_{\text{MAP}}; \Sigma_{\text{MAP}}) d\theta \\ &\approx \mathcal{GP}(m_T, k_T), \end{aligned} \tag{3.51}$$

where

$$\begin{aligned} m_T &= m_{\theta_{\text{MAP}}}^*(x) + \frac{1}{2} \text{Tr} \left(\Sigma_{\text{MAP}} m_{\theta_{\text{MAP}}}^{*''}(x) \right) \\ k_T &= k_{\theta_{\text{MAP}}}^*(x, x) + m_{\theta_{\text{MAP}}}^{*'}(x) \Sigma_{\text{MAP}} m_{\theta_{\text{MAP}}}^{*'}(x) \\ &\quad + \frac{1}{2} \text{Tr} \left(\Sigma_{\text{MAP}} k_{\theta_{\text{MAP}}}^{*''}(x, x) \right) - \frac{1}{4} \left(\text{Tr} \left(\Sigma_{\text{MAP}} m_{\theta_{\text{MAP}}}^{*''}(x) \right) \right)^2. \end{aligned} \tag{3.52}$$

$\text{Tr}(\dots)$ should be understood as the trace operator, and prime and double prime signs are the first and second derivatives respectively.

As can be seen, this approximation is much more involved than the previous one. It also proves to be more cumbersome to implement, especially when used for optimisation purposes. This is the main reason why we will use the previous approximation in the calculations appearing later in the chapter. This method was merely mentioned here to show the way for higher order approximations if need be.

3.6.3. Inferring the influence of uncertainty in measurement

Having done the other two inferences, it is time for the final stage of inference. In the discussion so far we have assumed that the data represented by \mathcal{D} is noise

free. In other words, the value read off the measuring device is interpreted as the ground truth. In reality, however, this is rarely the case.

Going back to subsection 3.3.1, we can evaluate the probability distribution of \mathcal{D} when the outcome of the measurement is “reading”,

$$p(\mathcal{D}|\text{reading}) = \frac{p(\text{reading}|\mathcal{D})p(\mathcal{D})}{p(\text{reading})}, \quad (3.53)$$

where $p(\text{reading}|\mathcal{D})$ represents the precision of the measuring device, which is usually modelled by a normal distribution $p(\text{reading}|\mathcal{D}) \sim \mathcal{N}(\mathcal{D}, \sigma_n^2)$. Assuming there is no prior, the posterior is

$$p(\mathcal{D}|\text{reading}) \sim \mathcal{N}(\mathcal{D}, \sigma_n^2). \quad (3.54)$$

In the case of a Gaussian process as the model, the only effect this inference has is that the covariance function $k(\vec{x}_i, \vec{x}_j)$ is replaced by $k(\vec{x}_i, \vec{x}_j) + \sigma_n^2 \delta_{ij}$ [62]; this can be confirmed by direct evaluation of the integral over \mathcal{D} in eq. (3.13). Equivalently, when written in matrix form, $\Sigma(\mathbf{X}, \mathbf{X})$ is replaced by $\Sigma(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \mathbf{1}$.

It was mentioned earlier that σ_n should be substituted for by the accuracy of the measuring device. This is a completely accurate statement if all the potential models are included in the predictive posterior distribution in eq. (3.13). But as was explained before, this is not possible due to the sheer number of different models one could come up with. In practice, it is very common to only pick a generic model flexible enough to describe the data. This generic model might make assumptions unwanted by the user. For instance, in fig. 3.2 some synthetic data (dashed line) is shown along with two fits (red and orange lines). Assume the precision of the data is given by the width of the line. First, σ_n is set to a much larger value than the measurement accuracy (shown in red) in fitting the model to the data. Obviously, this fit is capable of finding the period of the data successfully, and captures the overall features correctly. But it does not fall within measurement accuracy of the data. Now if the same model is forced to describe the data up to the measurement accuracy during the first cycle (shown in orange, and done by setting σ_n to the measurement precision), as shown by model 2, it is seen that the model completely fails to reproduce any of the characteristics of the data. The reason why this model never predicts the data “perfectly” has to do with the choice of model rather than the accuracy of the measurement. This model can never go through every single data point simply because it expects a different

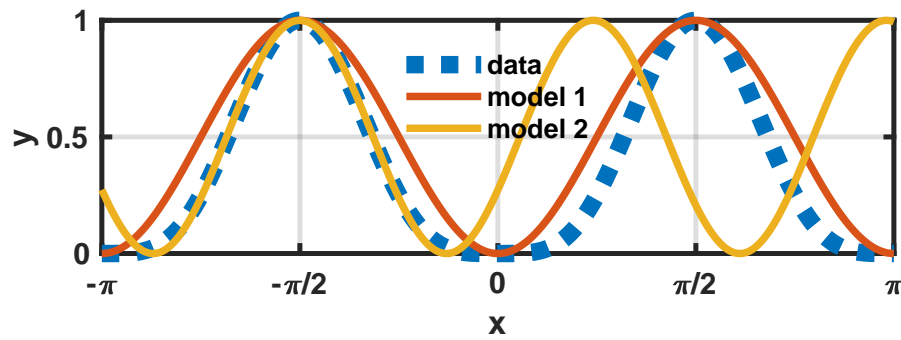


Figure 3.2. The synthetic data is shown by the blue dashed line, and the thickness represents the precision of the measurement. A hypothetical model has been used to describe the data using two different values of σ_n ; the red line uses a larger value than the yellow one. The yellow line fails to capture any of the large scale characteristics of the data, whereas the red line misses the details. Since we are usually after the general behaviour, if the model is not capable of reproducing the data *exactly*, it is always advisable to use a larger value of σ_n than the precision of the simulation or experiment.

behaviour from the data. More often than not, model 1 is preferred in practice as the large-scale behaviour (as opposed to the small-scale behaviour) is often what is desired. And since it is almost never possible to come up with a model capable of describing the data perfectly, it is advisable to consider a larger value for σ_n in practical applications than the measurement accuracy [75, 76]. It is hard to speculate what a good value is without knowing the problem, but our experience shows a value of a few percent of the range of the data almost always works fine. This is indeed why we use $\sigma_n \approx 0.1$ later in the chapter when $-5 < y < 5$, even though the uncertainty of the data points dictated by the numerical simulation is orders of magnitude smaller.

3.7. Heuristic methods of optimisation

If the function producing the outcome is referred to as the *latent function*, the objective of an optimisation algorithm is to find the minimum of this function. If the latent function is fairly cheap to evaluate, any textbook optimisation method could

be used. But in the limit of expensive functions these methods tend to become less and less efficient. The ultimate goal of an efficient optimization algorithm is to find the global minimum² by making as few latent function calls as possible.

Having learnt about Bayesian statistics, we have at our disposal all the necessary tools to estimate uncertainties and making predictions. According to the maximum entropy theorem [77], only the Gaussian distribution makes as few assumptions about unobserved features as possible. Consequently, Gaussian processes prove a promising option for what is known as the surrogate model [78], which is a mathematical model used to mimic the behaviour of the latent function.

In this section, a few of the most popular optimisation algorithms based on Gaussian processes will be introduced. The list is by no means exhaustive, as optimisation is still an active field of research [79–85]. They all share the same procedure in finding the global minimum,

1. The latent function is evaluated for a finite number of random points in the parameter space; these values are called target values. The parameters along with their corresponding target values form the training set.
2. A Gaussian process is trained using the training set. The training involves tuning the hyperparameters following any of the methods introduced in the previous section. For instance, if the MAP method is used, training is equivalent to finding θ_{MAP} .
3. A cost function is defined based on the predictive mean and variance of the Gaussian process. The minimum of this cost function is the point to consider in the next iteration. Note that this minimum is not necessarily the minimum of the latent function. That is, some steps are merely taken to explore the less explored areas of the parameter space.
4. The latent function is evaluated at the point suggested by the cost function. The results are added to the training set.

Then the algorithm goes back to stage 2. This loop continues until a satisfactory result is obtained. All the different methods appearing in this section differ only

2. Without loss of generality it is assumed that we are after the global minimum rather than maximum. A maximization problem can be easily converted into a minimization problem by multiplying the target function by -1 .

in their cost function.

One of the important factors regarding global optimum finders is the balance between exploitation and exploration maintained by the algorithm.³ Exploration is the process in which the algorithm picks points in the unexplored regions of the parameter space. And exploitation, on the other hand, is the process in which the algorithm picks points in the neighbourhood of the current minima. A good algorithm should not lose sight of any of the two. Unfortunately, there is no fundamental law setting the balancing point between the two, and hence trial and error seems to be the only way forward.

There is no immediate way of determining which one of the methods below performs best. However, there are studies comparing these methods and many other methods under different conditions [86]. Based on these studies, we have decided to pick the one-step lookahead method for our optimisation tasks.

3.7.1. Upper Confidence Bound (UCB) [87]

We assume here that we are interested in the maximum of the latent function (this is the convention used in ref. [87]); this can be easily converted into a minimisation problem by multiplying the latent function with -1 . If the Gaussian process is used to model the data, a greedy (completely exploitative) optimisation algorithm would suggest the maximum of $m^*(x)$ as the next point. Such an algorithm does not take advantage of the uncertainty provided by the Gaussian process. The Upper Confidence Bound method tries to alleviate this problem by incorporating the uncertainty in the following way,

$$u_{\text{UCB}}(x) \equiv m^*(x) + \beta_t^{1/2} \sqrt{k^*(x, x)}, \quad (3.55)$$

where β_t is the parameter that determines the level of exploration; the value of β_t is determined below. A confidence interval is any interval claiming to contain the true value of a random variable. For instance, an interval centred at the mean $m^*(x)$ extending as far as the uncertainty $\beta_t^{1/2} \sqrt{k^*(x, x)}$ on each side is a confidence interval. The upper bound of this confidence interval is believed to represent the maximum value that the latent function can have at that point with a high probability. Therefore, the maximum of this cost function over the entire

3. Local optimisation algorithms are always completely exploitative.

search space might in fact be where the maximum of the target function is. Thus, it is chosen as the next point. It is shown in ref. [87] that the following choice of β_t holds good convergence properties,

$$\beta_t = 2 \log \left(\frac{d\pi^2}{6\delta} t^2 \right), \quad \text{with } \delta \in [0, 1], \quad (3.56)$$

where d is the dimension of the space, and t is the lap number, i.e. $t = 1, 2, \dots$

The ultimate aim of this cost function is to minimise the uncertainty over the entire search space. To see that, notice that as the algorithm progresses (β_t increases), the next point tends to be solely dictated by the uncertainty. That is, the algorithm becomes completely exploratory in the limit of $t \rightarrow \infty$, something that might sound a good strategy in the case of unlimited resources. However, our resources are almost always limited in reality, and therefore a resource-friendly exploration-exploitation trade-off is needed.

3.7.2. Machine Learning Online Optimisation (MLOO) [41]

In contrast to the previous subsection, we are interested in the minimum of the latent function here. This method is very similar to UCB, except the prefactor b below changes periodically in Q steps from 0 to 1, e.g. if $Q = 3$ then $b = 0, 1/3, 2/3, 1, 0, \dots$ [41], rendering different degrees of exploitation-exploration trade-off,

$$u_{\text{MLOO}}(x) \equiv b m^*(x) - (1 - b) \sqrt{k^*(x, x)}, \quad (3.57)$$

where the term containing $m^*(x)$ represents exploitation and the term containing $k^*(x, x)$ represents exploration. The value of Q is chosen by the user. The minimum of this function is picked as the next point. The algorithm is completely exploitative when b is equal to 1, and completely exploratory when b is equal to 0.

One major downside of this scheme is that if the mean function varies over a wider range than the uncertainty (something that will eventually happen as the algorithm progresses), there is not much to be gained from the exploratory phase of the algorithm. Furthermore, since some of the steps are merely taken to reduce uncertainty, an improvement on the global minimum should not be expected more than once every Q steps. If the latent function is expensive to evaluate, this might enforce a huge overhead on the available computational resources.

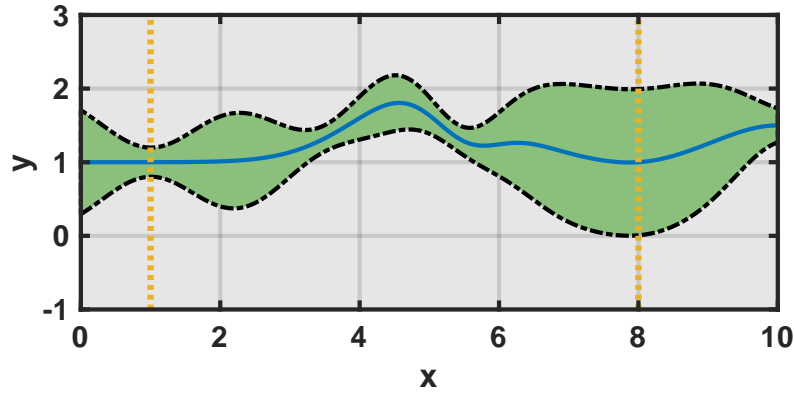


Figure 3.3. The prediction mean (shown by blue line) and uncertainty (shown by the green region) of a synthetic Gaussian process are shown here. The two points marked by the vertical dotted lines have the same mean and probability of improvement. However, the one on the right seems to be a better choice for the next lap.

3.7.3. Probability of Improvement (PI) [88]

Again we are interested in the global minimum here. This method evaluates the *probability* that the latent function at a given point has a lower value than the current minimum η [88], and the maximum of this probability is suggested as the next point,

$$\begin{aligned}
 u_{\text{PI}}(x) &\equiv \mathbb{E}[\theta(\eta - y)] = \\
 &= \int_{-\infty}^{\infty} \theta(\eta - y) \mathcal{N}(y; m^*(x), k^*(x, x)) dy \\
 &= \frac{1}{2} \operatorname{erfc} \left(\frac{m^*(x) - \eta}{\sqrt{2k^*(x, x)}} \right),
 \end{aligned} \tag{3.58}$$

where $\mathcal{N}(y; m^*(x), k^*(x, x))$ is the output of the Gaussian process at point x , and $\theta(\eta - y)$ is the Heaviside step function. The symbol \mathbb{E} means the average of its argument weighted by the output of the Gaussian process. This cost function hopes to find the global minimum merely by evaluating the probability that the latent function has a lower value at any given point. Even though convincing at first glance, this idea completely fails to distinguish between two points having identical probabilities of improvement, but different uncertainties, see fig. 3.3.

3.7.4. Expected Improvement (EI) [89]

Assuming we are again interested in the global minimum, the expected improvement scheme seems to have resolved the problem depicted in fig. 3.3. The cost function here is different from the previous one in that it is weighted by the difference between the value of the target function and the current minimum. That is, the expected improvement is defined as the average of $\max\{0, \eta - y\}$ weighted by the outcome of the Gaussian process at point x [89]

$$\begin{aligned}
 u_{\text{EI}}(x) &\equiv \mathbb{E}[\max\{0, \eta - y\}] \\
 &= \int_{-\infty}^{\eta} (\eta - y) \mathcal{N}(y; m^*(x), k^*(x, x)) dy \\
 &= -\frac{1}{2} \operatorname{erfc} \left(\frac{m^*(x) - \eta}{\sqrt{2k^*(x, x)}} \right) (m^*(x) - \eta) \\
 &\quad + \frac{1}{\sqrt{2\pi k^*(x, x)}} \exp \left(-\frac{(m^*(x) - \eta)^2}{2k^*(x, x)} \right) k^*(x, x),
 \end{aligned} \tag{3.59}$$

where again η is the current minimum. The next point is given by the maximum of $u_{\text{EI}}(x)$. The first term of the final expression is in charge of exploitation while the second term is in charge of exploration. This algorithm is by far the most common choice in the literature.

One piece of information unused by any of the cost functions introduced so far is the number of laps left. If it was known that the optimisation algorithm would be terminated in k steps, the optimisation algorithm could take advantage of this fact by taking a few exploratory shots first before trying to find the minimum. This is indeed the main drawback of the EI scheme. The next cost function is an attempt at easing this.

3.7.5. One-Step Lookahead [86]

The main idea behind this algorithm is that there are only enough resources for one more try, and after that the minimum will have to be reported [86]. If the value y of the last latent function call is lower than the current minimum η , the new value y will be reported, otherwise η . Mathematically speaking, $\min\{\eta, y\}$ will be reported. Therefore, the original minimisation problem is mapped into the

optimisation problem of finding the minimum of

$$\begin{aligned}
\Lambda_1(x) &\equiv \mathbb{E}[\min\{\eta, y\}] \\
&= \int_{-\infty}^{\eta} y \mathcal{N}(y; m^*(x), k^*(x, x)) dy + \int_{\eta}^{\infty} \eta \mathcal{N}(y; m^*(x), k^*(x, x)) dy \\
&= \eta \\
&\quad + \frac{1}{2} \operatorname{erfc} \left(\frac{m^*(x) - \eta}{\sqrt{2k^*(x, x)}} \right) (m^*(x) - \eta) \\
&\quad - \frac{1}{\sqrt{2\pi k^*(x, x)}} \exp \left(-\frac{(m^*(x) - \eta)^2}{2k^*(x, x)} \right) k^*(x, x),
\end{aligned} \tag{3.60}$$

where subscript 1 shows that there's only one try left. The second term of the final expression represents exploitation while the last term represents exploration. This method is very similar (if not identical) to the EI method introduced above. However, it sets a clear path towards generalisation of $\Lambda_1(x)$ to the multi-step case [86]. But unfortunately the final expression for the multi-step version tends to become very complicated very quickly, leading to multivariate integrals very expensive to evaluate [86].

We will repeatedly make use of the one-step version of this algorithm (or equivalently the EI method) in the rest of the chapter. To find the minimum of this cost function, a variety of quasi-Newton methods are used, the specific choice of which is irrelevant to our discussion. These methods take advantage of the derivatives of $\Lambda_1(x)$ in order to find the minimum efficiently. The derivatives are derived in appendix C.

3.8. General procedure of optimisation

A brief rundown of the optimisation method is given here to summarise the points discussed so far. The method will be used in the subsequent sections.

1. A number of random points are selected in the parameter space, for each of which a simulation is run and the target value is evaluated; this set is called the training set. An appropriate size for the training set is dictated by the dimension and extend of the parameter space, e.g. the higher the number of parameters, the larger the training set. Care has to be taken to make sure the random points satisfy the constraints on the parameters. For example,

in the first application to appear in the next section, the norm of the vector consisting of the optimisation parameters has to remain equal to one.

2. It is very important to normalise the acquired data before feeding it to the Gaussian process. Assume that in a two-dimensional parameter space the first parameter changes between 0 and 1, and the second parameter changes between 0 and 10^3 . Then the correlation length along the first axis might be of the order of 1 while it could be as large as 10^3 along the second axis. This difference in magnitude endangers the numerical stability of the method. Therefore, we will always normalise the data to fall in the interval $[-5, 5]$ along each dimension.
3. A mean function and a covariance function are chosen for the Gaussian process. The mean function is usually set to zero (as done in the rest of this chapter) if the data is normalised to be centred at 0. We will always choose the squared exponential covariance function.
4. A Gaussian process is trained with the training set. The training phase entails finding the set of hyperparameters either maximising the likelihood when using the MLE approximation, or maximising the posterior when using the MAP approximation; for better stability, the log likelihood/posterior is maximised. To find the maximum, a handful of random numbers are generated in the hyperparameter space and used as the initial seed to the Newton method of optimisation.
5. After the Gaussian process is trained, the one-step lookahead cost function given in eq. (3.60) is used to investigate what points to consider next. The minimum of this surface is found using ad hoc quasi-Newton methods which are fashioned for each specific problem. For instance, in the first application below, this quasi-Newton method is restricted to find the minimum of the one-step lookahead cost function on the surface of a sphere.
6. The coordinates of the lowest n local minima of the cost function are passed on to n nodes, on each of which a simulation is run for the given values of the optimisation parameters.
7. The new results are first normalised and then added to the training set. The algorithm goes back to step 4. This loop continues until the termination criterion is met. For the rest of this chapter a bound on the total number of

simulations will be used as the termination criterion.

An illustration of the method is shown in fig. 3.4. There, the global minimum of the one-dimensional Ackley function is found; the Ackley function is given by [90]

$$f(\vec{x}) = -a \exp \left(-b \sqrt{\frac{1}{d} \sum_{i=1}^d x_i^2} \right) - \exp \left(\frac{1}{d} \sum_{i=1}^d \cos(cx_i) \right) + a + \exp(1), \quad (3.61)$$

where d represents the dimension of the vector \vec{x} . An initial training set of 3 points is used. To train the Gaussian process, the set of hyperparameters maximising the likelihood should be found. In the case of the squared exponential covariance function the hyperparameters are λ and σ_f , see eq. 3.36. The likelihood function at each step is shown on the right; the brighter colours represent higher values. After finding the maximum, the prediction of the Gaussian process using the ‘‘Affine Transformation’’ approximation is shown on the left. The blue line is the mean, and the green area is the uncertainty at each point ($\pm\sqrt{k^*(x, x)}$ from the mean). The latent function (which is the Ackley function here) is shown by the black line. The points evaluated so far are marked by the plus signs. The smaller plot below this plot in each row shows the one-step lookahead cost function (for better visibility $\exp(\Lambda_1(x))$ is shown here), and the point to be considered next is marked by a circle. Between each two rows 3 laps of the aforementioned loop are carried out, hence a jump of 3 in the number of points.

The plausibility of pointwise estimates such as the MLE and MAP estimates is based on the conjecture that in the limit of a large number of samples the distribution of probable hyperparameter values will converge to a delta function. This behaviour can be seen in the plots on the right hand side of fig. 3.4 where the size of the peak in the hyperparameter space shrinks as more observations are made.

The hyperparameter space could in principle have more than one local maximum, and therefore it is always necessary to explore this space thoroughly before reporting θ_{MAP} .

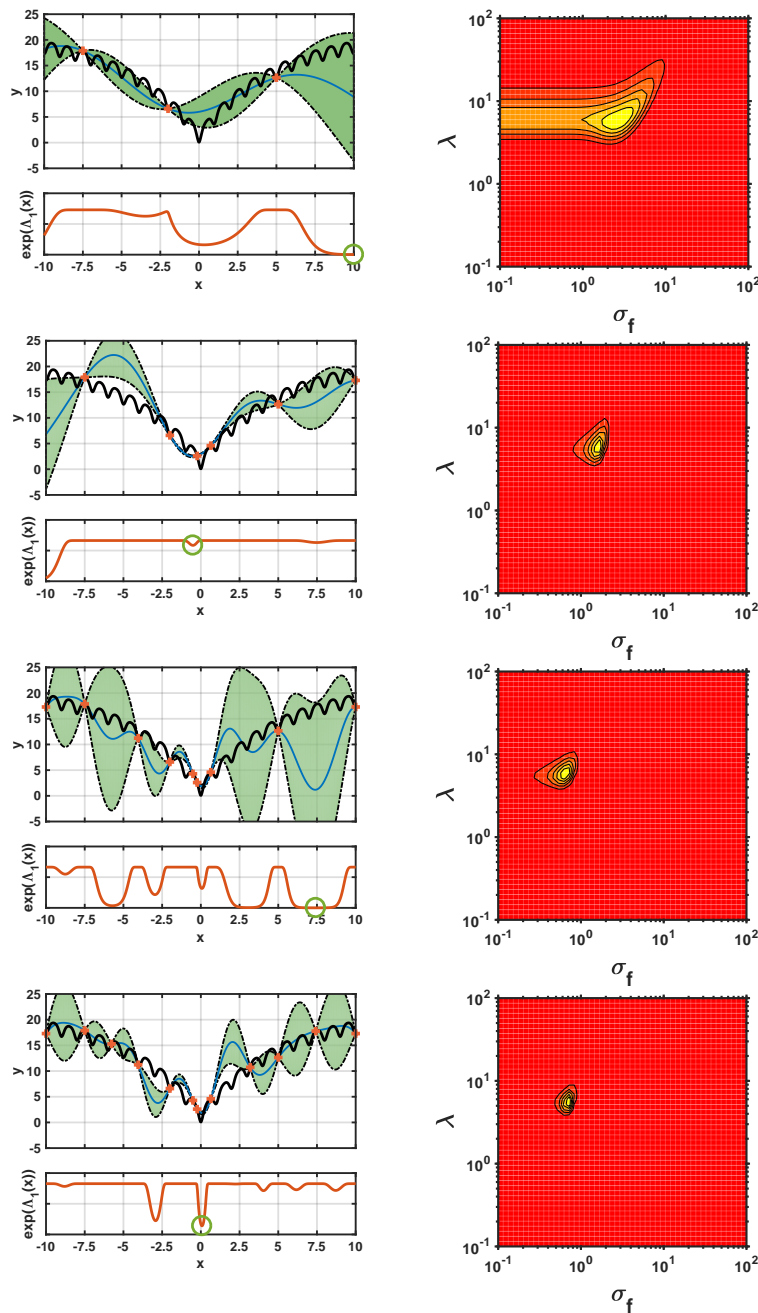


Figure 3.4. Finding the minimum of the 1-dimensional Ackley function (black solid line) with parameters $a = 20$, $b = 0.2$, $c = 2\pi$, and $d = 1$. A training set of three points is used. Left: the predicted mean by the Gaussian process in blue, and the uncertainty in green. The smaller plots are the one-step lookahead cost function, the global minimum of which (marked by a circle) is the next point. The plots share the same x-axis. Right: the likelihood as a function of the hyperparameters; the brighter areas show higher values. Between each two rows three latent function calls have been made. See eq. 3.36 for the definitions of λ and σ_f .

3.9. Maximising the electron yield

As the first application, we will use the optimization method introduced so far to maximize the electron yield of a hydrogen atom when exposed to a laser pulse. The envelope of the pulse is modified in search of the optimal pulse shape while keeping the pulse energy fixed.

In a regime where the first order perturbation theory successfully describes the behaviour of the system, the final population of the continuum is given by the Fourier transform of the laser envelope (see section 4.2)

$$a_\epsilon(t) = -id_{1\epsilon} \int_{-\infty}^t E(t') e^{-i(E_1 - \epsilon)t'} dt', \quad (3.62)$$

where ϵ is the energy of the continuum state, $d_{1\epsilon}$ is the dipole matrix element coupling the ground state to the continuum, and E_1 is the energy of the ground state. $E(t) = \mathcal{E}(t) \sin(\omega t)$ is the electric component of the laser pulse, and $\mathcal{E}(t)$ is the envelope. The total ionisation of the atom is the sum of the final populations of all the continuum states,

$$Q = \int_0^\infty |a_\epsilon(\infty)|^2 d\epsilon, \quad (3.63)$$

which according to Parseval's theorem is equal to (assuming $d_{1\epsilon}$ only changes negligibly over the populated region of the energy spectrum)

$$Q \propto \int_{-\infty}^\infty |\mathcal{E}(t')|^2 dt', \quad (3.64)$$

which is the area under the square of the envelope. This means so long as the area is unchanged, the amount of ionisation remains unaffected. Therefore, the total amount of ionisation is completely determined by the area under the square of the laser pulse.

The first order perturbation regime should be contrasted with the second order perturbation regime in which the calculation of the final continuum population is more involved (this expression along with its detailed derivation are given in section 4.6, which is the central equation of chapter 4)

$$a_\epsilon(t) = -id_{1\epsilon} \int_0^t dt' e^{-i(\delta-1)J(t')} E(t') e^{-i(E_1 - \epsilon)t'} e^{-\gamma J(t')/2}, \quad (3.65)$$

where $J(t) = \frac{1}{\omega^2} \int_0^t \mathcal{E}^2(t') dt'$, and $\mathcal{E}(t')$ is the envelope of the electric field. The Stark shift of the ground is determined by δ , and γ is the depletion rate of the

ground state. The highly non-linear dependence of the above equation on the envelope allows us to affect the total ionisation by changing the envelope.

In this section, first the pulse envelope will be expanded in terms of a finite number of the quantum harmonic oscillator eigenstates. Thanks to the exponentially decaying nature of these eigenstates, any envelope represented by such a finite expansion is a realistic choice of envelope (see subsection E.1 for more details). Then it will be seen how to engineer the pulse shape in order to have pulses of different shapes but the same energy. Even though there is no fundamental law requiring us to use pulses of identical energy, it is certainly more meaningful to talk about the optimum pulse shape out of a pool of pulses sharing the same energy. The final goal is to find the pulse maximising the ionisation. The simulations are done using the SCID package [91] (see appendix F).

3.9.1. Pulse envelope

An expansion in terms of the eigenstates of a quantum harmonic oscillator (QHO) is used to express the envelope of the laser pulse⁴

$$\mathcal{A}(t) = \mathcal{A}_0 \sum_{i=1}^d a_i \pi^{1/4} \phi_i(2\sqrt{2 \ln 2} \frac{t}{\tau}), \quad (3.66)$$

where $\phi_i(t)$ are QHO eigenstates and a_i are their corresponding weights.

Each QHO eigenstate consists of a hermite polynomial $H_n(x)$ multiplied by an exponential term [92],

$$\phi_i(x) = \frac{\pi^{-1/4}}{\sqrt{2^{i-1} (i-1)!}} e^{-\frac{x^2}{2}} H_{i-1}(x), \quad i = 1, 2, 3, \dots \quad (3.67)$$

The exponential term vanishes in the limit of $x \rightarrow \pm\infty$; when these eigenstates are used to express the electric field, this exponential decay makes sure the electric field is zero before and after the lifetime of the laser pulse. Furthermore, by including higher eigenstates in the expansion (which is equivalent to including higher hermite polynomials), it is possible to systematically increase the flexibility of the laser pulse while avoiding cusps in the electric field. Obviously we have to make sure

4. It is easy to show that a laser pulse expressed in the time domain using such an expansion (assuming the laser pulse is real-valued), can be expressed in the Fourier domain using the exact same quantum harmonic oscillator eigenstates with the exact same weights.

the value of the highest included eigenstate does not undergo considerable change over one period of the pulse, otherwise the separation of the pulse into a product of a carrier wave and an envelope might not be very meaningful.

To see why the prefactors $\pi^{1/4}$ and $2\sqrt{2\ln 2}$ are introduced, let's consider the simplest choice of envelope, that is, the QHO ground state,

$$\mathcal{A}(t) = \mathcal{A}_0 \pi^{1/4} \phi_1(2\sqrt{2\ln 2} \frac{t}{\tau}) = \mathcal{A}_0 \exp(-4 \ln 2 \frac{t^2}{\tau^2}), \quad (3.68)$$

where it is seen that $2\sqrt{2\ln 2}$ is introduced to have a full width at half maximum (FWHM) of τ , and $\pi^{1/4}$ is introduced to have an amplitude of \mathcal{A}_0 .

3.9.2. Pulse energy

If the pulse envelope changes slowly with time compared to the carrier frequency (that is, eq. (E.2) from appendix E is satisfied), the energy of the pulse can be directly calculated from the vector potential (eq. (E.20) has to be inserted in eq. (E.22))

$$U \approx \frac{c\epsilon_0}{2} \omega^2 \int_{-\infty}^{+\infty} \mathcal{A}^2(t) dt. \quad (3.69)$$

By replacing the vector potential with its expansion from eq. (3.66), we get

$$U \approx \frac{c\epsilon_0}{2} \mathcal{A}_0^2 \omega^2 \sum_i a_i^2 \int_{-\infty}^{+\infty} \phi_i^2(2\sqrt{2\ln 2} \frac{t}{\tau}) dt, \quad (3.70)$$

where the cross terms vanish due to the orthogonality of QHO states. The QHO eigenstates are also normalised, meaning the integral is the same for all ϕ_i , and therefore it suffices to only evaluate it for one, say $i = 0$,

$$\int_{-\infty}^{+\infty} \exp^2\left(-4 \ln 2 \frac{t^2}{\tau^2}\right) dt = \sqrt{\frac{\pi}{\ln 4}} \frac{\tau}{2}, \quad (3.71)$$

which when replaced in the expression above yields

$$U \approx \sqrt{\frac{\pi}{\ln 4}} \frac{c\epsilon_0}{4} \mathcal{A}_0^2 \omega^2 \tau \sum_i a_i^2. \quad (3.72)$$

If the laser parameters (i.e. \mathcal{A}_0 , ω and τ) are kept constant, a variety of pulses of different envelopes but equal energy can be generated if a_i are chosen such that $\sum_i a_i^2 = 1$.

3.9.3. Target function

The function that is minimised here is the remaining bound charge in the hydrogen atom targeted by the laser pulse. This is equivalent to maximising the electron yield. The bound charge is calculated by adding up all the bound state populations

$$y = \sum_i |\langle \phi_i | \psi(t_f) \rangle|^2, \quad (3.73)$$

where i runs over all bound states, $|\phi_i\rangle$ are the eigenstates of the hydrogen atom, and $|\psi(t_f)\rangle$ is the final wave function.

3.9.4. Procedure & Results

As discussed earlier, fixing the energy of the pulse is achieved by making sure $\sum_i a_i^2 = 1$. Therefore, when a_i lie on a d -dimensional sphere of radius one, the pulse energy is completely determined by the overall amplitude, frequency, and width of the pulse. Therefore, the amplitude, frequency, and width of the pulse are kept fixed in this optimisation task, while different points on a d -dimensional sphere are used to represent new envelopes.

First, a training set has to be set up. For this purpose, a few random points are generated on a d -dimensional sphere, whose coordinates are then used as the pulse shape expansion coefficients a_i . For each envelope a simulation is run and the target function is evaluated. Each simulation runs from $t_i = -2.5\tau$ to $t_f = 2.5\tau$ where τ is the FWHM of the laser pulse. The envelopes and their corresponding target values form the training set used to train the Gaussian process.

The one-step lookahead scheme is used as the optimisation algorithm. During each lap of the optimisation algorithm the next four points⁵ are found by finding the four lowest local minima of the cost function defined in eq. (3.60). These four points represent four new envelopes. A simulation is run for each of the new envelopes giving rise to four new target values. The four envelopes and their target values are added to the training set before proceeding to the next lap of the loop. This loop continues until the termination criterion is fulfilled, which in this task is merely an upper bound on the total number of simulations.

5. The reason why four points are chosen is that there were four cores available on the computer used for this project

The optimisation procedure has been done three times with increasing flexibility of the pulse. The first one is a pulse made up of the first three QHO eigenstates (referred to as $3-d$ below), the second one of the first 5 states (referred to as $5-d$ below), and the last one of the first 7 states (referred to as $7-d$ below). The vector potential amplitude is set to $\mathcal{A}_0 = 3$ a.u. for all the cases. This corresponds to an intensity of 1.3×10^{18} W/cm². The frequency of the carrier wave is $\omega = 2$ a.u., equivalent to a wavelength of $\lambda = 23$ nm. And the FWHM is $\tau = 25$ a.u., equivalent to 0.6 fs. These parameter values are chosen such that the first order perturbation theory is inadequate to describe the behaviour of the system, and at the same time the ground state is not completely depleted within a fraction of the pulse.

For the $3-d$ case, a training set of 40 points is used. The optimisation algorithm is then allowed to run for 20 laps. These 120 points are shown on a 3-dimensional sphere in fig. 3.5, where the plus signs are the points for which a simulation is run, and the colour coding represents the prediction by the Gaussian process. The fact that there are points all around the sphere is a sign of exploration, whereas the concentration of points around the minimum is a sign of exploitation. The optimal pulse is shown in fig. 3.6 along with its decomposition in terms of QHO eigenstates. This pulse achieves an electron yield of 0.9522. As a side note, it is worth noting that even though the symmetry of the problem with respect to the overall sign of a_i wasn't specifically taken into account, its presence is clearly visible in the prediction by the Gaussian process. However, for larger numbers of optimisation parameters, it is certainly worthwhile to take advantage of the known symmetries of the system to reduce the parameter space and therefore enhance the performance of the optimisation algorithm. As a second note, even though the minimum of the target function is what is found here by the algorithm, the prediction by the Gaussian process makes it very easy to find the maximum almost immediately without any need for further simulations.

For the $5-d$ case, a training set of 60 points is used, followed by 20 laps. In total 140 points are considered. The optimal pulse is shown in fig. 3.7. This pulse achieves an electron yield of 0.9629.

For the $7-d$ case, the training set has 100 points. The algorithm runs for 60 laps, hence a total of 340 points are considered. This pulse achieves an electron yield of 0.9643. The pulse and its decomposition are shown in fig. 3.8.

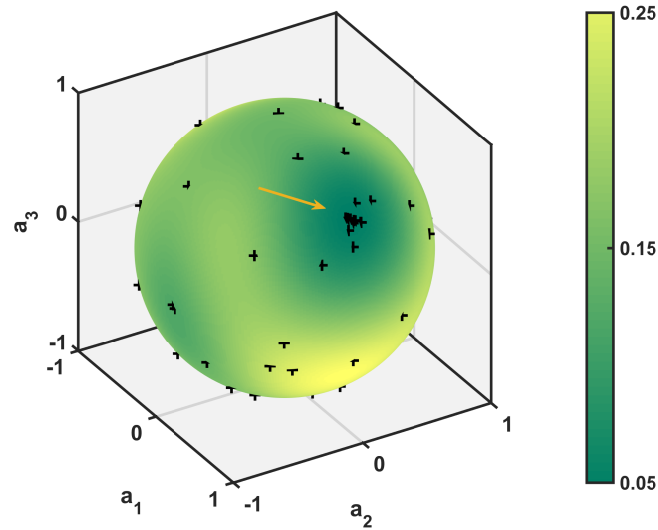


Figure 3.5. The Gaussian process used to find the optimal pulse shape made up of three components ($d = 3$ in eq. 3.66). The colour coding shows the prediction by the Gaussian process of the remaining charge in the atom after the pulse is gone. The points for which a function call has been made are shown by crosses. The minimum is marked by an arrow.

3.9.5. Analysis

In all of the systems above two things can be seen, 1. it seems that the optimisation algorithm is trying to create a pulse with a relatively constant envelope everywhere apart from the rising and falling parts, 2. the optimal pulses have no component along the even components.

The first observation can be justified by noting that if the intensity is too high at any given point during the pulse, the atom might enter the stabilisation regime around that point (in the stabilisation regime the ionisation rate decreases significantly [93]). On the other hand, if the intensity is too low the ionisation decreases according to the perturbation theory. Therefore it is best if the envelope stays somewhere between these two limits at all times.

As to why the even components do not appear, if an even component is used to increase (decrease) the intensity during the first half of the laser pulse, the intensity

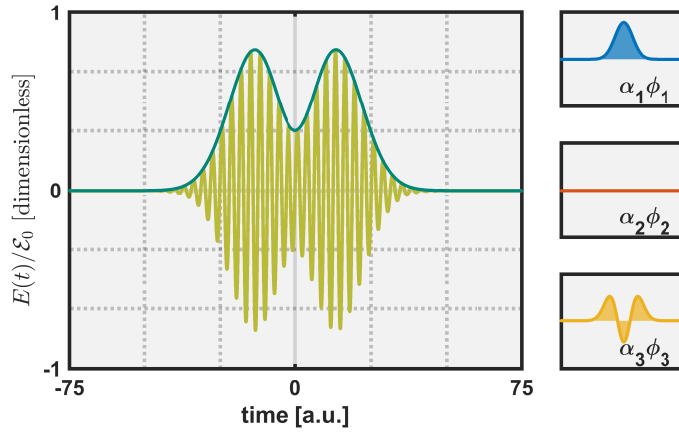


Figure 3.6. The optimal pulse and its decomposition in terms of QHO eigenstates. Note that $\mathcal{E}_0 = \mathcal{A}_0\omega$. This pulse is made up of three components ($d = 3$ in eq. 3.66), which are shown on the right hand side. The scale of the smaller plots is the same as that of the main one.

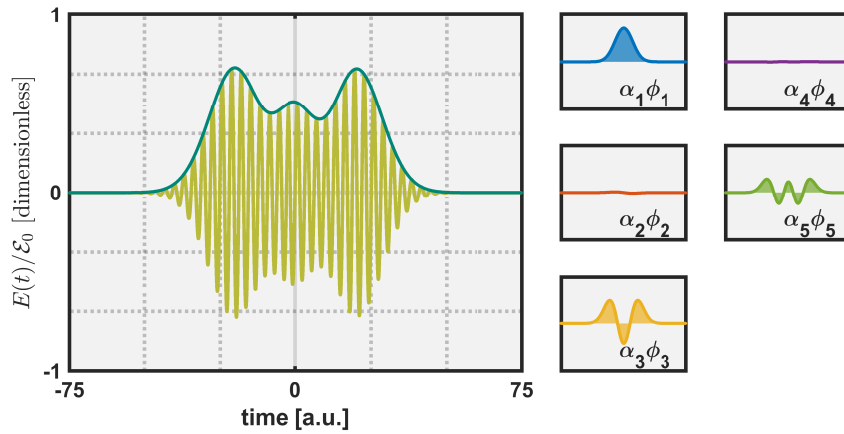


Figure 3.7. The optimal pulse and its decomposition in terms of QHO eigenstates. Note that $\mathcal{E}_0 = \mathcal{A}_0\omega$. This pulse is made up of five components ($d = 5$ in eq. 3.66), which are shown on the right hand side. The scale of the smaller plots is the same as that of the main one.

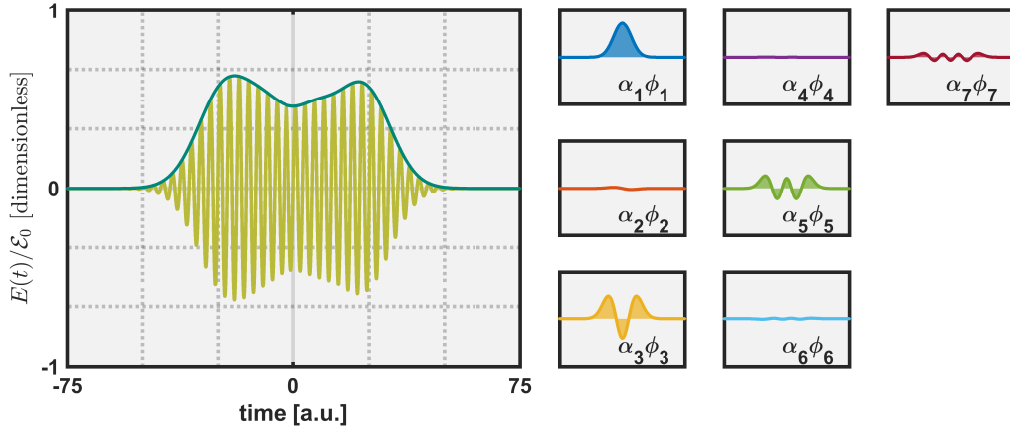


Figure 3.8. The optimal pulse and its decomposition in terms of QHO eigenstates. Note that $\mathcal{E}_0 = \mathcal{A}_0\omega$. This pulse is made up of seven components ($d = 7$ in eq. 3.66), which are shown on the right hand side. The scale of the smaller plots is the same as that of the main one.

necessarily decreases (increases) during the second half due to the antisymmetry of the even components. This antisymmetry prevents the envelope from building up a plateau in the central section of the pulse.

The slight asymmetry in the pulse of the 7- d example is a direct consequence of the fact that the difference in target value between the current global minimum and the true global minimum is very small (based on the previous paragraph, we expect the true global minimum to be completely symmetric), and for the optimisation algorithm to find the true minimiser, a lot more function calls would have been needed.

3.10. Maximising the asymmetry in the photo-electron angular distribution

The photo-electron angular distribution (PAD) is calculated by integrating the norm of the wave function over the radius and the azimuthal angle,

$$S(\theta) = \int_0^{2\pi} d\varphi \int_0^\infty dr |\psi_{\text{con}}(\mathbf{r})|^2 r^2 \quad (3.74)$$

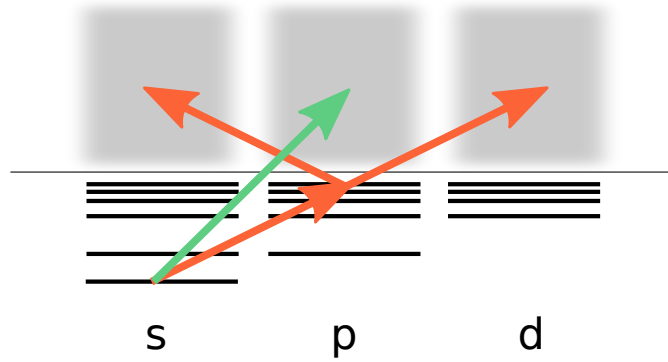


Figure 3.9. The energy structure of the hydrogen atom. The red arrows show photons of the first laser, and the green arrow shows a photon of the second laser. The shaded regions represent the continuum. The usual selection rules apply as a direct consequence of the linear polarisation of the two laser pulses.

where θ is the polar angle, φ is the azimuthal angle, \mathbf{r} is the position vector with r representing its magnitude, and $\psi_{\text{con}}(\mathbf{r})$ is the ionised part of the wave function, i.e. the part that is in the continuum.

Now consider the two ionisation processes below,

1. One-photon above-threshold ionisation using a linearly polarised laser pulse, where the electron always ends up in the p-continuum if it is initially in the ground state.
2. Two-photon below-threshold ionisation using a linearly polarised laser pulse, where the final state of the electron is a mixture of s- and d-continua.

In each of these scenarios the final continuum population is symmetric with respect to the polar angle (assuming the lasers are polarised along the z -axis), see fig. 3.9. In other words, the PAD is identical in the northern and southern hemispheres. However, if these two laser pulses are used simultaneously, we will have access to both of these ionisation processes, i.e. one-photon and two-photon, and consequently access to populated s-, p-, and d-continua. As a result, the PAD might not be symmetric any more.

The coupling between the bound p-states and the s-continuum is much weaker than that between the bound p-states and the d-continuum, meaning the s-continuum

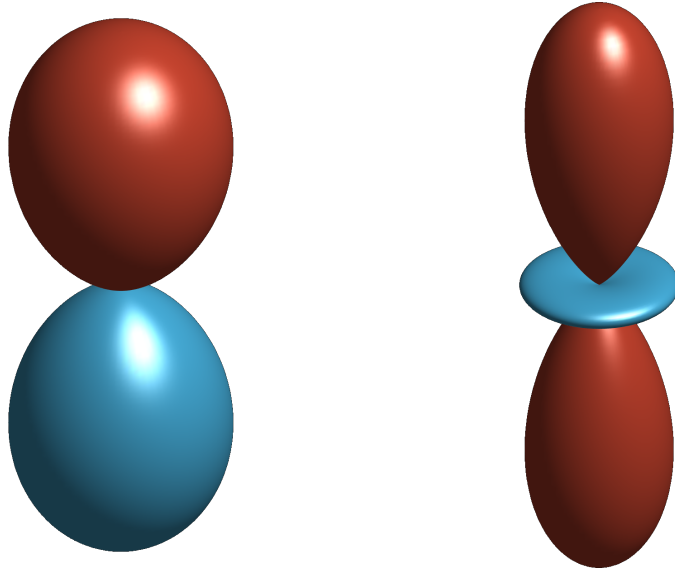


Figure 3.10. Spherical harmonics $Y_{l,m}$; $Y_{1,0}$ is shown on the left and $Y_{2,0}$ on the right. Red portions represent regions where the spherical harmonic is positive, and the blue portions where it is negative. The one-photon ionisation process gives rise to $Y_{1,0}$, whereas the two-photon ionisation process gives rise to $Y_{2,0}$. By combining the two ionisation processes it is possible to generate asymmetry along the polar angle.

does not play much of a role in this scenario. The other two competing channels, i.e. p- and d-channels, are shown in fig. 3.10. The interplay of these two channels results in an asymmetry in the photo-electron spectrum with respect to the polar angle. And the goal of the current optimisation task is to maximise this asymmetry by changing the laser parameters.

The vector potential consisting of two laser pulses sharing the same envelope $\mathcal{A}(t)$ is given by

$$A(t) = \mathcal{A}(t) (\mathcal{A}_1 \cos(\omega_1 t) + \mathcal{A}_2 \cos(\omega_2 t + \delta)) , \quad (3.75)$$

where \mathcal{A}_1 and \mathcal{A}_2 are respectively the amplitudes of the two components. The frequencies are denoted by ω_1 and ω_2 . And δ is the phase difference between the two carrier waves. The presence of this phase difference is crucial to the success of this optimisation task, as a change in this phase leads directly to a change in the relative phase of the two channels. And the relative phase of the two channels in turn governs the degree of asymmetry in the photo-electron angular distribution by

dictating how the two competing channels interfere. The optimisation algorithm will vary the phase difference as one of its parameters in an attempt to maximise the asymmetry.

The second important parameter here is the intensity. If the intensity of one of the two pulses is much higher than the other, the continuum population will be dominated by the corresponding channel, again resulting in a symmetric photo-electron angular distribution. Therefore, we will keep the intensity of the first pulse fixed, while the intensity of the second pulse is used as the second optimisation parameter by the optimisation algorithm to maximise the asymmetry.

An identical envelope is chosen for both of the pulses to maximise the combined action of the two lasers. However, in a more complicated optimisation task, the envelopes of the two pulses may be separately parametrised in a similar fashion to sec. 3.9.1 and used as extra optimisation parameters. But we will just use a simple envelope in this section.

In sum, there are two variables that the optimisation algorithm is allowed to change, the phase difference δ and the amplitude of the second pulse \mathcal{A}_2 .

3.10.1. Target function

The target function is defined as the difference between the continuum populations of the northern and southern hemispheres,

$$y = a_N - a_S, \quad (3.76)$$

where a_N is the population of the continuum states in the northern hemisphere, and a_S is the population of the continuum states in the southern hemisphere. By finding the laser parameters which minimise this function, it is possible to get a visibly asymmetric wave function with respect to the polar angle [8]. The population of the continuum states in the northern hemisphere is obtained by integrating the photo-electron distribution of the continuum states from $\theta = 0$ to $\theta = \pi/2$. And similarly the population of the continuum states in the southern hemisphere is obtained by integrating the photo-electron distribution of the continuum states from $\theta = \pi/2$ to $\theta = \pi$. The reason why only the continuum states are considered is that in experiment it is only possible to detect the the continuum part of the wave function.

3.10.2. Procedure & Results

An exponential envelope $\mathcal{A}(t) = \exp(-4 \ln 2 t^2 / \tau^2)$ is used where τ is the FWHM of the pulse. The frequencies are $\omega_1 = 24.5$ eV and $\omega_2 = \omega_1/2$; the frequency of the first laser is chosen such that the frequency of the second laser is just below the threshold of atomic hydrogen (the threshold is at 13.6 eV). The FWHM of the envelope is $\tau = 1.16$ fs, and the intensity of the first laser is $I_1 = 1.0 \times 10^{15}$ W/cm². The regime defined by these parameter values can be achieved nowadays thanks to free electron lasers [94, 95].

The training set consists of 20 randomly chosen points in an area defined by $0 < \delta < 2\pi$ and $10^{14} < I_2 < 10^{16}$ W/cm², which are then passed on to the SCID package [91] (see appendix F). The simulations run from $t = -2.5\tau$ to $t = 2.5\tau$. The target values are calculated according to eq. (3.76). These results collectively form the training set used for training the Gaussian process.

The one-step lookahead scheme is the method of choice here, and is allowed to run for 20 laps. In every lap the four most promising points are chosen by finding the four lowest local minima of the one-step lookahead cost function defined in eq. (3.60) (we have used a generic quasi-Newton method to find the local minima of the cost function, but any other textbook gradient-based optimisation algorithm would equally work). These four points are then given to the SCID package. The target values and the points are added to the training set before the next lap starts.

The parameter space consisting of the two optimisation parameters, i.e. the intensity of the second pulse and the phase difference between the two carrier waves, is shown in fig. 3.11. The crosses show the points for which a simulation has been run. The colour coding shows the prediction by the Gaussian process. The fact that there are crosses throughout the entire region shows the exploratory aspect of the optimisation algorithm, and the fact that the promising regions are more densely sampled shows the exploitative aspect. The optimum points are shown by two arrows on the plot. The photo-electron angular distribution obtained using the optimal pulse is depicted in fig. 3.12, where the asymmetry is clearly visible.

Similarly to the previous example, even though the goal was to find the maximum asymmetry, it is possible to find the point for which the minimum asymmetry occurs almost instantly just by looking at the prediction made by the Gaussian process.

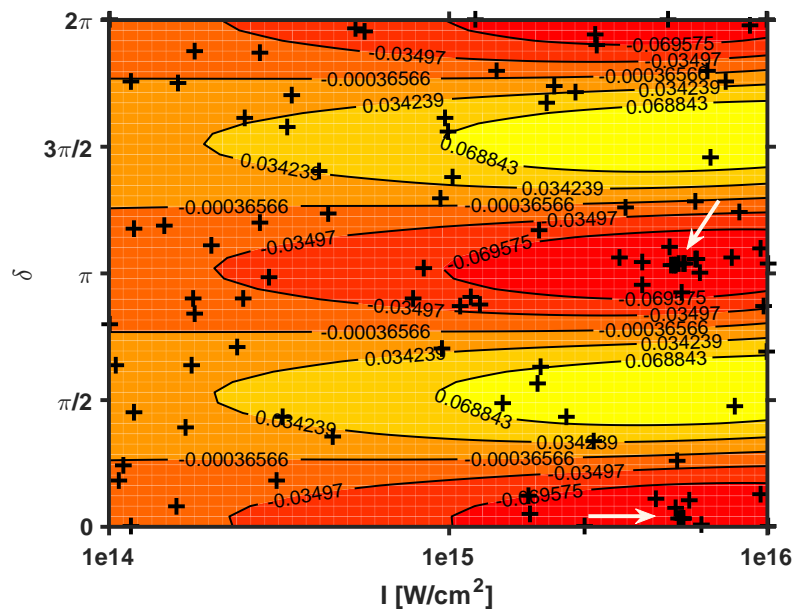


Figure 3.11. The difference between ionisation into the northern and southern hemispheres is depicted here as a function of the intensity of the second laser pulse and the phase difference between the two laser pulses. The arrows show the optimum regions.

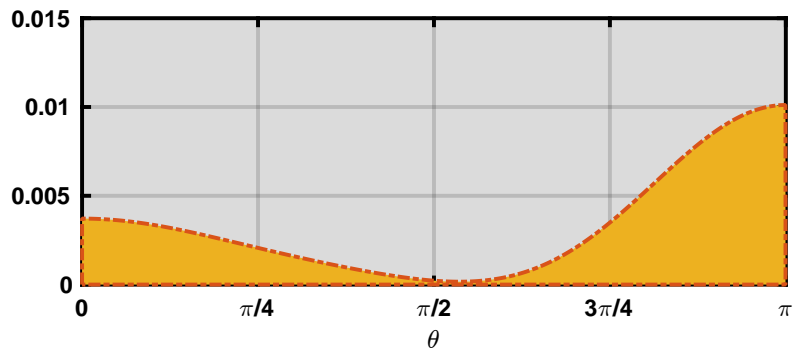


Figure 3.12. The photo-electron angular distribution having the highest level of asymmetry. The horizontal axis is the polar angle.

3.10.3. Analysis

It should be fairly straightforward to justify the obtained optimum values of the optimisation variables. As to the phase difference, when δ is 0 ($\pi/2$) the upper (lower) halves of spherical harmonics $Y_{1,0}$ and $Y_{2,0}$ add up, while the lower (upper) halves cancel out, see fig. 3.10. Either of these cases results in maximum contrast between the upper and lower parts.

Regarding the intensity of the second pulse, if the difference between the two intensities is high, as it was mentioned earlier, only one process dominates, leading to a significant reduction in the asymmetry. Therefore, the two intensities are expected to be of the same order when the asymmetry is to be maximised. The exact prediction of the optimum value is, however, only possible by running the optimisation algorithm.

3.11. Maximising the HHG yield by incorporating an XUV laser pulse

A typical radiation spectrum obtained from a high harmonic generator contains a region where all the radiated harmonics are of equal intensity, see figure 3.13. This region, called the plateau, extends up to a certain energy called the cut-off energy. The cut-off energy (in atomic units) is given by [96, 97]

$$\omega_{\text{cut}} = I_p + 3.17 U_p, \quad (3.77)$$

where I_p is the ionization potential, and U_p is the ponderomotive energy. The ponderomotive energy (in atomic units) of an electron in a laser field of amplitude \mathcal{E}_0 and frequency ω reads

$$U_p = \frac{\mathcal{E}_0^2}{4\omega^2}. \quad (3.78)$$

If harmonics beyond the cut-off energy are desired, the cut-off energy has to be somehow extended. This can naively be done in two ways, increasing the intensity of the laser, or decreasing the frequency of the laser. The intensity can only be increased so much, as at some intensity all the bound charge will be ionised leaving nothing behind to radiate. As to the frequency, as the wavelength becomes larger (lower frequencies), so does the quiver amplitude of the ionised electron wave

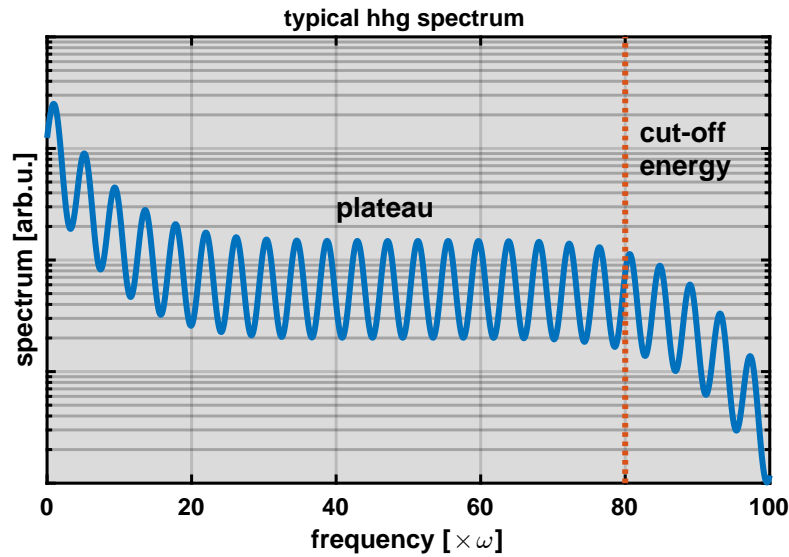


Figure 3.13. A (synthetic) typical spectrum obtained from a HHG experiment. The spectrum has three regions, the perturbative region on the far left, the plateau in the middle, and the cut-off region on the far right. The red dotted line marks the cut-off energy. The peaks are at odd multiples of the frequency of the incident laser.

packet. A larger quiver amplitude means the wave packet has to travel a larger distance before returning to the nucleus, and a longer trip in turn means the wave packet has more time to spread over space. That means, a smaller fraction of the wave packet has the chance to collide with the nucleus on the return journey, hence a decline in the efficiency. In other words, even though it is possible to increase the cut-off energy by decreasing the frequency, the efficiency of such a setup is unfavourable. But the story isn't over yet. The envelope of the pulse can be modified to enhance one particular harmonic beyond the original cut-off energy of the $1d$ hydrogen atom [98]. The quantum optimal control theory can also be used to find the optimum electric field at any instance of time (as opposed to only the envelope) maximising the radiation within a certain energy range of the $1d$ hydrogen atom [99, 100]. One further possible way to push the cut-off energy towards higher energies is to incorporate a second laser [96, 101]. Assuming the first laser operates in the IR regime,⁶ the second one can be an XUV laser. This is the setup we are concerned with in this section, and the goal is to maximise the radiation within a

⁶ This is a justified assumption as 800 nm lasers are the most common lasers in HHG experiments.

certain range of frequencies by changing the parameters of the second laser.

The vector potential of two simultaneous laser pulses is given by

$$A(t) = \mathcal{A}(t) (\mathcal{A}_1 \cos(\omega_1 t) + \mathcal{A}_2 \cos(\omega_2 t)) , \quad (3.79)$$

where \mathcal{A}_1 and \mathcal{A}_2 are the amplitudes, ω_1 and ω_2 are the frequencies, and $\mathcal{A}(t)$ is the envelope shared by both pulses (the reason why the two pulses share the same envelope is explained in sec. 3.10). If the intensity of one of the lasers is much higher than the other, the spectrum will resemble that of a single laser setup. We will therefore fix the intensity of the first laser while the intensity of the second laser is left to the optimisation algorithm.

It turns out, the frequency of the second laser plays a crucial role, and will consequently be varied by the optimisation algorithm in search of optimum settings. The phase of the carrier wave of the second laser is irrelevant due to its high frequency.

In sum, there are two parameters the optimisation algorithm will be in control of, the intensity and frequency of the second laser.

3.11.1. Target function

The total radiation between frequencies ω_l and ω_h is calculated by integrating the radiation spectrum $S(\omega)$ (please refer to appendix E.5 to see how the radiation spectrum is calculated) over the respective range,

$$y = \int_{\omega_l}^{\omega_h} S(\omega) d\omega , \quad (3.80)$$

where ω_h and ω_l are the limits of the window of interest. This quantity will be maximised by changing the laser parameters.⁷ The two limits of the target function are $\omega_l = 30\omega_{\text{IR}}$ and $\omega_h = 50\omega_{\text{IR}}$. These two frequencies define a window beyond the cut-off region of the IR pulse given the parameter values used in this task (see the next section).

⁷ In fact $-\log y$ will be minimised. When working with quantities that can vary over multiple orders of magnitude, it's best to work with their logarithm for better numerical stability.

3.11.2. Procedure & Results

Both pulses share the same exponential envelope, $\mathcal{A}(t) = \exp(-t^2/\tau^2)$, where $\tau = 2.9$ fs. An 800 nm laser is chosen as the first laser (equivalent to $\omega_{\text{IR}} = 0.057$ a.u.) with intensity 1.0×10^{14} W/cm². These are standard laser parameters used in HHG experiments [102].

The training set contains 20 randomly chosen points in the region defined by $1 < \omega_{\text{XUV}} < 3$ a.u. and $10^{10} < I_{\text{XUV}} < 10^{12}$ W/cm². The intensity of the second pulse is chosen low enough to make sure it acts as a small perturbation to the first laser pulse. A simulation from $t = -2.5\tau$ to $t = 2.5\tau$ is run for each of the points and the respective target values are evaluated. The SCID package is used for all the simulations [91] (see appendix F).

One-step lookahead procedure runs for 5 laps on four CPUs. In each lap the next four points are chosen by finding the four lowest minima of the cost function defined in eq. (3.60). A simulation is run for each of the four points, and the target values are evaluated as explained before. The results are added to the training set before proceeding to the next lap. By the end of the life of the optimisation algorithm a total of 40 points are considered.

In fig. 3.14 the prediction of the Gaussian process is shown as a contour plot, and the points for which a simulation is done are shown as crosses. As before, the degree to which the crosses are scattered all over the parameter space is a sign of exploration, whereas the fact that there are more crosses around the optimum value is a sign of exploitation. In this particular case we can see that there is a flat region, shown by the white arrows, where there are plenty of points. It turns out these points do not differ much in target value. The spectra from two of these points (the top most and bottom most arrows) are shown in fig. 3.15.

3.11.3. Analysis

To justify the optimum frequency found by the optimisation algorithm, an analytic analysis of the system is required. Fortunately, as becomes clear towards the end of this subsection, a crude analysis succeeds in describing the findings successfully. The analysis appearing here closely follows that of ref. [103]. The strong field approximation (SFA) is assumed throughout [104]. That is, the effect of the laser

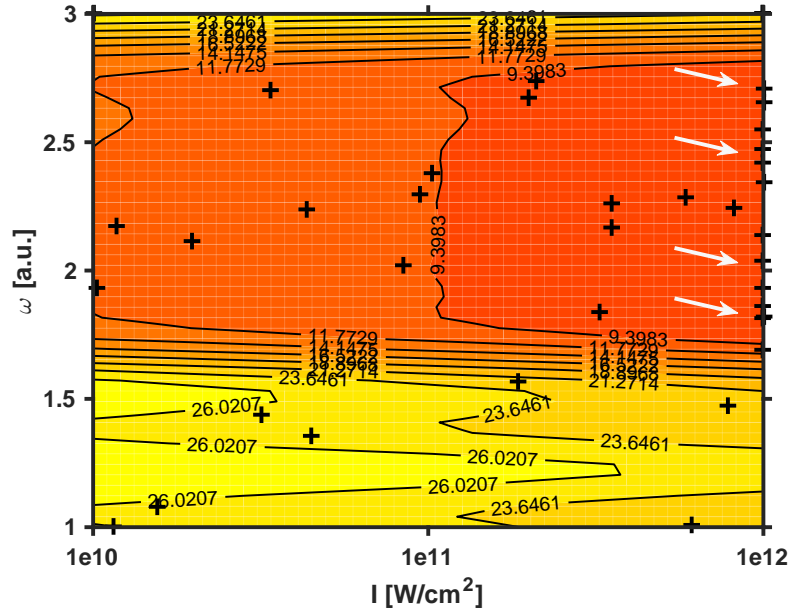


Figure 3.14. The region of the parameter space that the optimisation algorithm had access to. The horizontal axis represents the intensity of the XUV laser, while the vertical axis represents its frequency. The optimum region found by the algorithm is marked by the white arrows. The spectra obtained using the uppermost and the lowermost ones are depicted in fig. 3.15.

field on the bound states is neglected, and so is the effect of the atomic potential on the continuum states.

The evolution of a field-free hydrogen atom is considered first. The IR field is then added to the analysis, and at the very end the XUV pulse is considered as a perturbation.

First, the wave function of the hydrogen atom at the time of recombination t is decomposed into two parts, the bound part $\psi_b(\mathbf{r}, t)$ and the continuum part $\psi_c(\mathbf{r}, t)$,

$$\psi(\mathbf{r}, t) = \psi_b(\mathbf{r}, t) + \psi_c(\mathbf{r}, t). \quad (3.81)$$

The bound part evolves according to the field-free Hamiltonian $H_{\text{atom}} = p^2/2 + V(r)$ where $V(r) = -1/r$ and $p = |\mathbf{p}|$, while the continuum part evolves under the influence of only the external field, $H_{\text{ex}} = p^2/2 + \mathbf{A}_{\text{IR}}(t) \cdot \mathbf{p} + \mathbf{A}_{\text{XUV}}(t) \cdot \mathbf{p}$, where $\mathbf{A}_{\text{IR}}(t)$ is the vector potential of the IR field and $\mathbf{A}_{\text{XUV}}(t)$ is the vector potential

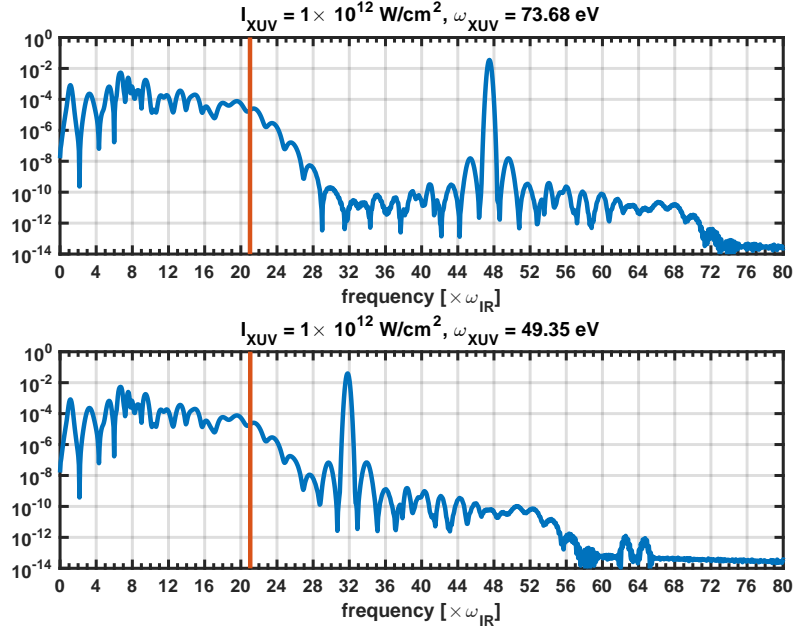


Figure 3.15. The spectra obtained from using the optimum values corresponding to the uppermost and lowermost arrows appearing in fig. 3.14. The upper plot corresponds to the uppermost arrow, and the lower plot to the lowermost arrow. The limits of the integral in the target function are $\omega_l = 30\omega_{\text{IR}}$ and $\omega_h = 50\omega_{\text{IR}}$. The red lines are the cut-off energy.

of the XUV field.

The evolution of the bound part is given by

$$\psi_b(\mathbf{r}, t) = e^{-iE_0 t} \phi_0(r), \quad (3.82)$$

where $\phi_0(r)$ is the ground state of the field-free Hamiltonian, and E_0 is its corresponding energy. The absence of any trace of the applied fields on this state is a direct consequence of the SFA approximation.

We will now consider the effect of the IR laser only, and then add the XUV pulse as a perturbation. The continuum states are given by the Volkov solution of the Schrödinger equation [104],

$$\phi_v^{\text{IR}}(\mathbf{r}, t) = \exp\left(i\mathbf{p} \cdot \mathbf{r} - i\frac{p^2}{2}t - i\mathbf{p} \cdot \mathbf{J}_{\text{IR}}(t)\right), \quad (3.83)$$

where $\mathbf{J}_{\text{IR}}(t) = \int_{t_0}^t \mathbf{A}_{\text{IR}}(t') dt'$. And if the effect of the XUV laser is included,

$$\begin{aligned} \phi_v(r, t) &= \exp\left(i\mathbf{p} \cdot \mathbf{r} - i\frac{p^2}{2}t - i\mathbf{p} \cdot (\mathbf{J}_{\text{IR}}(t) + \mathbf{J}_{\text{XUV}}(t))\right) \\ &= \phi_v^{\text{IR}}(\mathbf{r} - \mathbf{J}_{\text{XUV}}(t), t) \\ &\approx \phi_v^{\text{IR}}(\mathbf{r}, t) - \mathbf{J}_{\text{XUV}}(t) \cdot \nabla \phi_v^{\text{IR}}(\mathbf{r}, t), \end{aligned} \quad (3.84)$$

where $\mathbf{J}_{\text{XUV}}(t) = \int_{t_0}^t \mathbf{A}_{\text{XUV}}(t') dt'$, and it is assumed the intensity of the XUV pulse is low.

The continuum wave function at the time of recombination can be expanded in terms of the Volkov states

$$\begin{aligned} \psi_c(\mathbf{r}, t) &= \int \Omega(\mathbf{p}) \phi_v(\mathbf{r}, t) d^3p \\ &= \int \Omega(\mathbf{p}) \phi_v^{\text{IR}}(\mathbf{r}, t) d^3p - \mathbf{J}_{\text{XUV}}(t) \cdot \int \Omega(\mathbf{p}) \nabla \phi_v^{\text{IR}}(\mathbf{r}, t) d^3p \\ &= \psi_c^{\text{IR}}(\mathbf{r}, t) - \mathbf{J}_{\text{XUV}}(t) \cdot \nabla \psi_c^{\text{IR}}(\mathbf{r}, t), \end{aligned} \quad (3.85)$$

where $\Omega(\mathbf{p})$ is a function representing the expansion weights for different momenta, and eq. (3.84) is used in going from the first line to the second, and

$$\psi_c^{\text{IR}}(\mathbf{r}, t) \equiv \int \Omega(\mathbf{p}) \phi_v^{\text{IR}}(\mathbf{r}, t) d^3p, \quad (3.86)$$

which is the expansion of the wave function in terms of the Volkov states if only the IR laser is present (this can be confirmed by letting $\mathbf{J}_{\text{XUV}}(t) \rightarrow 0$ in eq. (3.85)).

The field radiated by the atom is given by (see appendix E.5)

$$\mathbf{a}(t) = \langle \psi(t) | -\nabla V(r) | \psi(t) \rangle, \quad (3.87)$$

where $\psi(t)$ is given by eq. (3.81),

$$\begin{aligned} \mathbf{a}(t) &= (\langle \psi_b(t) | + \langle \psi_c(t) |) - \nabla V(r) (|\psi_b(t)\rangle + |\psi_c(t)\rangle) \\ &= \langle \psi_b(t) | - \nabla V(r) | \psi_b(t) \rangle + \langle \psi_b(t) | - \nabla V(r) | \psi_c(t) \rangle \\ &\quad + \langle \psi_c(t) | - \nabla V(r) | \psi_b(t) \rangle + \langle \psi_c(t) | - \nabla V(r) | \psi_c(t) \rangle. \end{aligned} \quad (3.88)$$

If it is assumed most of the wave function stays largely in the ground state, then $\langle \psi_c(t) | - \nabla V(r) | \psi_c(t) \rangle \approx 0$. By using eq. (3.82) it is easily seen that $\langle \psi_b(t) | - \nabla V(r) | \psi_b(t) \rangle = 0$. The two remaining terms are complex conjugates and we will only consider one of them, with the presence of the other implied. If both laser

pulses are polarised along the z-axis, only the z-component of the acceleration vector is non-zero. Then, if the two parts of the wave function are substituted by eqs. (3.82) and (3.86), we get

$$\begin{aligned} a_z(t) &= \langle \psi_b(t) | -\partial_z V(r) | \psi_c^{\text{IR}}(t) \rangle - J_{\text{XUV}}(t) \langle \psi_b(t) | -\partial_z V(r) \partial_z | \psi_c^{\text{IR}}(t) \rangle \\ &= a_z^{\text{IR}}(t) - J_{\text{XUV}}(t) a_z^{\text{XUV}}(t), \end{aligned} \quad (3.89)$$

where in this case $\mathbf{J}_{\text{XUV}}(t) = J_{\text{XUV}}(t) \hat{z}$,

$$a_z^{\text{IR}}(t) \equiv \langle \psi_b(t) | -\partial_z V(r) | \psi_c^{\text{IR}}(t) \rangle, \quad (3.90)$$

and

$$a_z^{\text{XUV}}(t) \equiv \langle \psi_b(t) | -\partial_z V(r) \partial_z | \psi_c^{\text{IR}}(t) \rangle. \quad (3.91)$$

The HHG spectrum is given by the Fourier transform of $a_z(t)$

$$\mathcal{FT} \{a_z(t)\} = \mathcal{FT} \{a_z^{\text{IR}}(t)\} - \mathcal{FT} \{J_{\text{XUV}}(t)\} \star \mathcal{FT} \{a_z^{\text{XUV}}(t)\}, \quad (3.92)$$

where \star represents the convolution operation. The first term is simply the HHG spectrum generated by the IR pulse. To decipher the second term, let's assume the width of the envelope of the XUV pulse is much larger than its period, and therefore the Fourier transform of $J_{\text{XUV}}(t)$ can be approximated by a delta function,

$$\mathcal{FT} \{J_{\text{XUV}}(t)\} \approx \delta(\omega - \omega^{\text{XUV}}). \quad (3.93)$$

Convolution with a delta function is nothing but a shift in the frequency domain, i.e. if $g(y) = \delta(y - y_0)$

$$\begin{aligned} (f \star g)(y) &= \int f(x)g(y-x) dx \\ &= \int f(x)\delta(y-x-y_0) dx = f(y-y_0). \end{aligned} \quad (3.94)$$

Therefore, what the convolution appearing in eq. (3.92) does is shift $\mathcal{FT} \{a_z^{\text{XUV}}(t)\}$ by ω^{XUV} . This means in addition to the usual HHG spectrum generated by the first term, the second term gives rise to a structure around ω^{XUV} .

To analyse the symmetries of the radiated field, suppose two consecutive bursts of radiation are generated at t and $t + T/2$, where T is the period of the IR laser. And further assume $\psi_c^{\text{IR}}(\mathbf{r}, t)$ was born at some initial time t_i from $\psi_b(\mathbf{r}, t_i) = \phi_0(r)e^{-iE_0 t_i}$. Then $\psi_c^{\text{IR}}(\mathbf{r}, t + T/2)$ must have been born at $t_i + T/2$ from $\psi_b(\mathbf{r}, t_i +$

$T/2) = \psi_b(\mathbf{r}, t_i)e^{-iE_0T/2} = \phi_0(r)e^{-iE_0(t_i+T/2)}$ (the bound state has accumulated a phase of $E_0T/2$ over $T/2$). Furthermore, the direction of the laser pulse at $t + T/2$ is opposite that at t , and consequently the two continuum wave functions are released in opposite directions, i.e. $\psi_c^{\text{IR}}(\mathbf{r}, t + T/2) = \psi_c^{\text{IR}}(-\mathbf{r}, t)e^{-iE_0T/2}$. Plugging the relations for $\psi_b(\mathbf{r}, t_i + T/2)$ and $\psi_c^{\text{IR}}(\mathbf{r}, t + T/2)$ in eq. (3.90) yields

$$a_z^{\text{IR}}(t + T/2) = -a_z^{\text{IR}}(t), \quad (3.95)$$

where we have used the following symmetries

$$\phi_0(-r) = \phi_0(r) \quad (3.96)$$

$$V(-r) = V(r) \quad (3.97)$$

$$\partial_z V(-r) = -\partial_z V(r). \quad (3.98)$$

The only non-zero components of the Fourier transform of an odd function are the odd harmonics, $\pm\omega, \pm3\omega, \dots$, where $\omega = 2\pi/T$ is the frequency of the IR laser. Therefore, it is seen the field radiated by the IR laser has components only along the odd harmonics.

A similar analysis can be done for $a_z^{\text{XUV}}(t)$, yielding

$$a_z^{\text{XUV}}(t + T/2) = a_z^{\text{XUV}}(t), \quad (3.99)$$

where $\partial_z \psi_c^{\text{IR}}(\mathbf{r}, t + T/2) = -\partial_z \psi_c^{\text{IR}}(-\mathbf{r}, t)e^{-iE_0T/2}$ has been used. The Fourier transform of an even function only consists of even harmonics, $0, \pm2\omega, \dots$. Therefore, the spectrum from the XUV pulse consists of components at $\omega_{\text{XUV}}, \omega_{\text{XUV}} \pm 2\omega, \dots$

The total spectrum should consequently consist of the spectrum from the IR field, plus a structure consisting of even harmonics of the IR frequency centred at the XUV frequency.

Given that ω_l is larger than the cut-off energy of the IR laser, and the radiation due to $a_z^{\text{XUV}}(t)$ is symmetric about ω^{XUV} , one may correctly expect to have ω^{XUV} anywhere between ω_l and ω_h as the optimum frequency found by the optimisation algorithm, because at these frequencies the overlap of the radiation due to $a_z^{\text{XUV}}(t)$ with the target window is maximum. That is indeed the frequency range suggested by the optimisation algorithm.

As to the optimum intensity, the probability that the electron absorbs an XUV photon is proportional to the intensity of the XUV laser. Therefore, the HHG yield

within the window of interest increases monotonically as the intensity of the XUV laser increases.

The findings of the current optimisation task might seem trivial at first glance: if maximum radiation within a certain frequency window is needed, the frequency of the XUV laser should lie in that window. However, this task is the first step towards a more interesting optimisation task, i.e. when the XUV laser pulse is allowed to go beyond a simple perturbation. In such a regime, the behaviour of the system most likely cannot be easily described using the analysis presented here, and perhaps higher order terms in eq. (3.84) should be considered.

3.12. Generating short pulses via combining the higher harmonics generated by multicoloured pulses

The high harmonic generation (HHG) is a highly non-linear process, and if the incident laser pulse consists of multiple frequencies (known as colours), it may not be straightforward to predict the outcome merely based on the outcomes of the individual pulses. However, it is known that the outcome of such a setup can be engineered for different purposes [105]. For instance, by delicately shaping the incident laser pulse, it is possible to single out one of the higher harmonics while suppressing the others. Such degree of manipulation, even though possible, is very hard in reality. To facilitate future experiments, we restrict ourselves to a very simple setup here, that is, a multicoloured pulse consisting of three components at 800 nm, 400 nm, and 200 nm. The three components are given different weights, governing their relative shares of the pulse. As with section 3.9, it is more relevant to speak of pulses of identical energy. An optimisation algorithm will be used to investigate the possibility of creating short bursts of radiation by illuminating a hydrogen atom with the aforementioned multicoloured pulse.

3.12.1. Pulse composition

The vector potential of a multicoloured laser pulse is made up of identically shaped pulses⁸ with different frequencies

$$A(t) = \mathcal{A}(t) \sum_{i=1} a_i \cos(\omega_i t + \varphi_i), \quad (3.100)$$

where $\mathcal{A}(t)$ is the envelope, ω_i are the frequencies, and φ_i are the initial carrier wave phases. A Gaussian envelope is used here, $\mathcal{A}(t) = \exp(-t^2/\tau^2)$, where τ determines the width of the pulse.

3.12.2. Pulse energy

All the pulses used in the current optimisation task will have the same energy. And in order to come up with a constraint enforcing that, first the square of the time derivative of the vector potential is needed. The time derivative of eq. (3.100) in the slowly-varying envelope regime (see appendix E) is given by

$$\frac{dA(t)}{dt} \approx -\mathcal{A}(t) \sum_i a_i \omega_i \sin(\omega_i t + \varphi_i), \quad (3.101)$$

which when raised to the power of two yields

$$\left(\frac{dA(t)}{dt}\right)^2 \approx \mathcal{A}^2(t) \sum_{i,j} a_i a_j \omega_i \omega_j \sin(\omega_i t + \varphi_i) \sin(\omega_j t + \varphi_j). \quad (3.102)$$

Since we are only interested in time scales larger than the period of any of the constituent components, we take the average of the above expression over a few cycles,

$$\left\langle \left(\frac{dA(t)}{dt}\right)^2 \right\rangle \approx \frac{1}{2} \mathcal{A}^2(t) \sum_i a_i^2 \omega_i^2, \quad (3.103)$$

where we have taken advantage of $\langle \sin(\omega_i t) \sin(\omega_j t) \rangle = \frac{1}{2} \delta_{ij}$. By inserting this in $\langle I(t) \rangle = c\epsilon_0 \left\langle \left(\dot{A}(t)\right)^2 \right\rangle$ (see eq. (E.16)), we obtain the mean intensity

$$\langle I(t) \rangle \approx \frac{c\epsilon_0}{2} \mathcal{A}^2(t) \sum_i a_i^2 \omega_i^2. \quad (3.104)$$

8. This restriction is applied to make it easier for the experimentalist to implement such a pulse. In experiment, the 800 nm laser pulse is directed into a non-linear crystal to generate the other two components, i.e. 400 nm and 200 nm. Obviously, they will all share the same envelope.

In order to calculate the energy of a multicoloured pulse, we only need to integrate the mean intensity over time (see eq. (E.22) in appendix E)

$$U = \int_{-\infty}^{+\infty} I(t) dt = \frac{c\epsilon_0}{2} \left(\int_{-\infty}^{+\infty} \mathcal{A}^2(t) dt \right) \sum_i a_i^2 \omega_i^2, \quad (3.105)$$

which means in order to keep the energy of the pulse fixed, one only needs to keep the sum $\sum_i a_i^2 \omega_i^2$ fixed, in which case the only factor affecting the pulse energy is the pulse envelope, and the simplest way to change the energy of the pulse is to multiply the envelope by a prefactor, which is equivalent to increasing the peak intensity of the pulse. For a general discussion of how a constraint can be imposed during an optimisation task please refer to [6, 7].

3.12.3. Target function

The field emitted by the atom is given by the second derivative of the dipole operator with respect to time, i.e. $\langle \ddot{\mathbf{r}} \rangle$ (see appendix E.5). It is our goal to minimise this radiation everywhere but over a very short space of time. This is done by first multiplying the radiation field by a window function which is positive over the time period of interest and negative everywhere else, and second integrating the product over time,

$$y_{t_0} = \int_{-\infty}^{\infty} |\langle \ddot{\mathbf{r}} \rangle| w_{t_0}(t) dt, \quad (3.106)$$

where $w_{t_0}(t)$ is the window function. Please note that the absolute value of $\langle \ddot{\mathbf{r}} \rangle$ is used. In order to see what t_0 is, let's first consider a potential window function,

$$w_{t_0}(t) = \exp\left(-\frac{(t-t_0)^2}{\tau_1^2}\right) - \alpha \exp\left(-\frac{(t-t_0)^2}{\tau_2^2}\right), \quad (3.107)$$

which consists of two Gaussian functions centred at t_0 , with the first Gaussian being much narrower than the second one. The parameters used in this section are $\tau_1 = 200$, $\tau_2 = 20$, and $\alpha = 8$. This window function is depicted in fig. 3.16. The first term encourages the presence of the burst at t_0 while the second one discourages the fluctuations/noise around the burst. Therefore, t_0 represents the time when the burst is emitted. The weight α is chosen to balance the two competing factors; if it is too large, the emphasis will be on not getting any fluctuations/noise no matter how weak the final burst, and if it is too small, the emphasis will be on the burst no matter the level of fluctuations/noise.

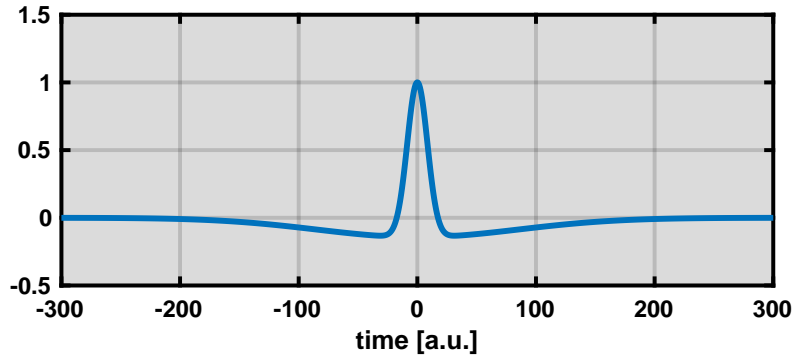


Figure 3.16. The window function used to separate out one single burst of radiation. The middle peak shows where the prospective burst is going to be. The negative areas around the peak make sure the final burst is as isolated as possible.

For this window function to work as hoped, the burst should be emitted at around t_0 . But this may not always be the case, and the window function might miss the exact time of the burst. For this reason, a better target function is achieved if t_0 is let to sweep the entire time axis and then the maximum is taken,

$$y = \max\{y_{t_0}\} \quad \text{for } -\infty < t_0 < \infty. \quad (3.108)$$

One very important point to bear in mind is that in our experience the choice of target function plays a very crucial role in the success of the optimisation task. For instance, we first used y_{t_0} as the target function with t_0 being the middle of the pulse, and were disappointed to discover that the global minimum did not meet our expectations.

3.12.4. Procedure & Results

The system under study is a hydrogen atom in the presence of a multicoloured pulse. The envelope is given by $\exp(-t^2/\tau^2)$, where $\tau = 2.9$ fs. The peak intensity of the pulse is 9.75×10^{11} W/cm²; these values are chosen such that there is enough ionisation while there is no saturation. The pulse consists of three components, $\omega_1 = 0.057$ a.u. ($\lambda = 800$ nm), $\omega_2 = 2\omega_1$ ($\lambda = 400$ nm), and $\omega_3 = 4\omega_1$ ($\lambda = 200$ nm). The amplitudes a_i are changed under the constraint that the pulse energy is fixed, i.e. $\sum_i a_i \omega_i$ remains unchanged. The phases are allowed to change freely between

0 and 2π .

A specific random number generator is devised to make sure every random point satisfies the constraint of constant energy. This random number generator produces a vector of length 6, the first three elements of which are the components of a point on the surface of a sphere. These three elements are used as the values of the products $a_i\omega_i$. The other three elements of the random vector are the components of a point picked unconditionally in a 3- d box extending from 0 to 2π in each dimension. Each of these three elements represents one of the phases mentioned above.

A training set of size 40 is prepared. Then the optimisation algorithm is allowed to run for 40 laps. In each lap, the next four points are suggested by the minima of the one-step lookahead method (eq. (3.60)), and then four simulations are run on four different nodes. A total of 200 simulations are run by the end of the life of the optimisation algorithm.

The vector potential of the optimum pulse is shown in the upper plot of fig. 3.17, which gives rise to the radiation shown in the lower plot. The amplitudes are $a_1 = 9.597$, $a_2 = 5.443$, and $a_3 = 2.465$ a.u., and the phases are $\varphi_1 = 0.245$, $\varphi_2 = 5.456$, and $\varphi_3 = 2.716$. The FWHM of the peak marked by the arrows is 3.3 a.u. (or 80 as).

3.12.5. Analysis

As it was mentioned earlier, the radiation by an electron in the presence of an electric field is proportional to the acceleration vector of the electron $\mathbf{a}(t) = \langle \psi(t) | \ddot{\mathbf{r}} | \psi(t) \rangle$, where \mathbf{r} is the position operator, and $|\psi(t)\rangle$ is the wave function of the electron. The acceleration vector is proportional to the exerted force, that is the electric field. The electric field is the time derivative of the vector potential. Therefore, in order to get a burst of radiation at a particular time, the time derivative of the vector potential needs to be large at that point in time. In other words, in order to have a burst of radiation, we need to have an abrupt change in the magnitude of the vector potential. That is indeed what is seen here. The radiation shown in fig. 3.17 contains a high peak marked by two arrows; this peak corresponds to the time instance pointed to in the upper plot where the slope is the greatest.

During the first half of the pulse there is no radiation merely due to the fact that not much of the initial wave function is ionised. In the second half, however, a train of bursts can be seen. Each of these bursts corresponds to a point in time when the slope of the vector potential is large.

3.13. Conclusion

Most traditional optimisation algorithms fail to incorporate all the available information about the latent function into their analysis (for instance there is no *uncertainty* consideration in the gradient descent method), which makes them less efficient in terms of the number of function evaluations required [106–108]. This weakness might not be a big problem when optimising cheap functions, but it will be a main issue when an expensive function is considered [109]. Bayesian methods, on the other hand, make sure the best decision is made in the light of the acquired data [61]. Note that the “best decision” of one user might be completely different from that of another user simply due to the fact that they don’t have access to the same information. In other words, in Bayesian statistics the concept of probability is a subjective quantity [68].

It was explained in this chapter how Bayes’ theorem lays the groundwork for learning algorithms, and how we can improve our beliefs in the light of new data. In sum, Bayesian statistics has three main cornerstones,

1. Each and every type of uncertainty has to be treated the same way. That is, the uncertainty stemming from our lack of knowledge is mathematically equivalent to the outcome of a random process.
2. Bayes’ theorem should be used to infer the posterior probability.
3. The posterior predictive distribution should be used when talking about unobserved events.

In optimisation, the first item incorporates in our analysis the uncertainty about the structure of the latent function where it hasn’t been thoroughly explored. The second item helps us improve our perception of the behaviour of the latent function as evidence becomes available. And the third item makes it possible to make predictions in view of the observed data and their corresponding uncertainties.

Later in the chapter, a Gaussian process was used as a potential Bayesian model, on top of which a heuristic optimisation algorithm was introduced. It was then demonstrated how this generic optimisation method can be helpful in not only finding the optimum parameter values, but also analysing the system under study. That is, by looking at the predictions made by the Gaussian process, it was possible to gain insight into the details of the system. For instance in the first application presented in section 3.9, even though the maximum ionisation was sought, it was possible to find the minimum without any need for extra simulations. Of course this cannot always be done, especially if the parameter space hasn't been thoroughly covered by the previous simulations. But this has to be contrasted with textbook optimisation methods in which there is usually no such privilege no matter how well the parameter space is sampled.

Each case studied here was accompanied by an analysis justifying the optimum with the help of a simplified picture of the system. Obviously such an analysis may not always be possible, especially as the complexity of the system grows. For instance, in the third application presented in this chapter, if higher intensities had been used for the two laser pulses, this perturbative approach would not have been enough, or at least higher orders would have been required.

There are several steps in any optimisation task that need very close attention. Firstly, it is very important that we understand what exactly is being optimised, i.e. what the target function is actually targeting. Although in certain situations there might not be many different choices, there are circumstances where the target function is just as important (if not more) as the optimisation algorithm itself. If the target function is not fine tuned, the global minimum may not necessarily represent the desired behaviour.

Secondly, the optimisation algorithm shouldn't be given too much/inadequate freedom. For instance, in the first example in section 3.9, the optimisation algorithm was only allowed to pick envelopes from a limited pool. On the one hand, such a pool definitely does not provide ultimate flexibility of the laser pulse. On the other, unconditionally flexible pulses are not what can be experimentally realised. If instead of just the envelope, the value of the vector potential at each time step had been put up for modification, the optimum pulse might not necessarily have represented a physically legitimate pulse. Another peril of using too much freedom is that as the number of parameters the optimisation algorithm has access to grows, the probability that the optimisation algorithm will actually find

the *global* optimum shrinks. Similar conclusions have been drawn before in other communities such as the quantum optimal control community [110, 111].

3.14. Outlook

In this chapter we only used the square exponential covariance function, even for latent functions of periodic behaviour. We expect that using periodic covariance functions will reduce the amount of required function calls considerably. Furthermore, in cases where one simple covariance function may not be sophisticated enough to capture the nature of the latent function, it is possible to design compound covariance functions simply by adding/multiplying two simple covariance functions [62]. This procedure can be automated such that the computer itself decides what combination of the available simple covariance functions performs best; the computer's judgement will be based on $p(\mathcal{M}|\mathcal{D})$ defined in subsection 3.3.1. This procedure, if not done very carefully, might result in overfitting. Such a procedure has been successfully used in the field of machine learning before [112].

As the number of optimisation parameters increases, so does the number of required function calls. It was seen earlier that matrix inversion is required in evaluating the prediction of the Gaussian process, which becomes the numerical bottleneck as the number of available data points increases. Fortunately, there are various approximate methods which can be useful on this front [62]. The feasibility of such approximations in physical settings should be studied. There might also be ways to customise the currently available approximate methods to systems prevalent in physics, such as periodic systems, systems with short-range interactions, etc.

The possibility of tailoring general-purpose covariance functions to specific applications is certainly worth considering. One possible way would be to incorporate the known symmetries of the system into the covariance function to the best of our ability. For instance, a covariance function defined only on the surface of a sphere might show better performance in section 3.9.

Tackling more complicated systems should definitely be on the agenda for future work. For instance, highly nonlinear systems are a very interesting candidate. In such systems, the output may not be as predictable as the relatively simple systems

considered in this chapter.

The only type of latent function used in this chapter was the output from simulation. But the output of experiment can also act as a latent function. This has been recently done [41] to reduce the preparation time of a Bose-Einstein condensate. Similar procedures can be used to find the optimum experimental conditions for a particular outcome.

A range of different techniques can be used to reduce the dimensions of high-dimensional problems, a procedure known as dimensionality reduction in the literature [57]. In effect, this procedure helps reduce the number of required data points, and therefore increase the efficiency of the Gaussian process.

Dimensionality reduction can also be used for visualisation of high-dimensional data. For instance, the method of Principal Component Analysis (PCA) uses a unitary transformation to convert a set of possibly correlated optimisation parameters into a set of linearly uncorrelated parameters [57]. A few of the uncorrelated variables can then be visualised using ordinary visualisation methods. Such techniques can provide insight into patterns hidden in the data, or help us have a better understanding of the inner workings of the system.

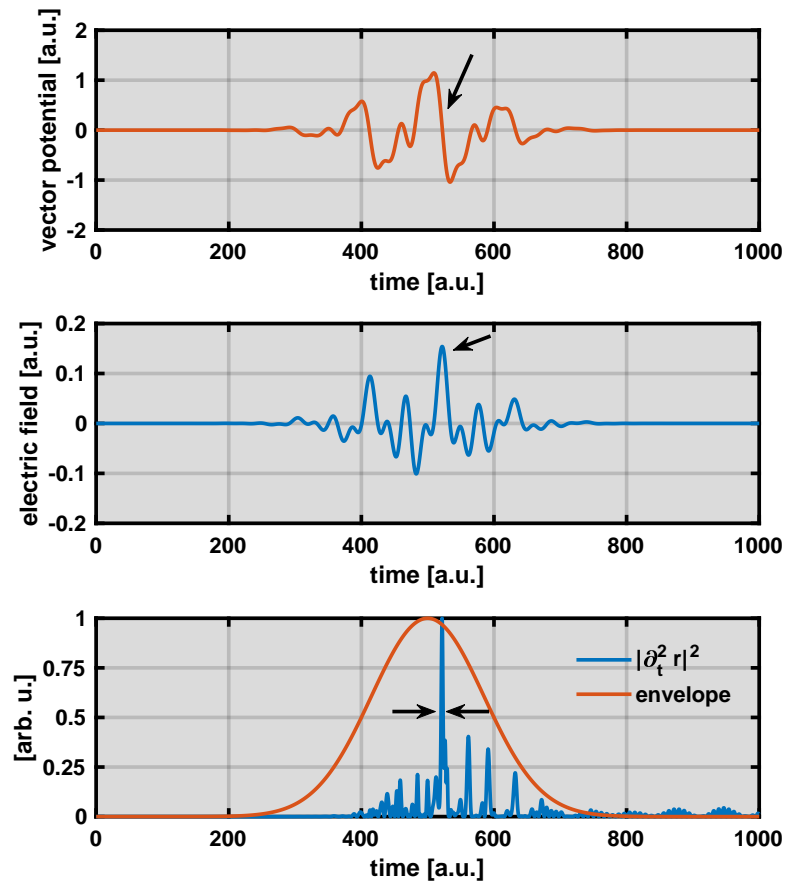


Figure 3.17. Upper plot: the vector potential of the optimum pulse obtained by the optimisation algorithm. The section marked by an arrow is when the burst is radiated. Middle plot: the electric field of the optimum pulse. The arrow marks the corresponding section of the electric field responsible for the burst. Lower plot: the radiation obtained by using such a pulse. The FWHM of the burst marked by the arrows is 3.3 a.u. (or 80 as).

4. Dynamic interference

The interaction of a laser pulse with a simple atom such as hydrogen can be expressed through the standard minimal coupling Hamiltonian. As uncomplicated as this Hamiltonian may look, it turns out its solution at one frequency or intensity cannot be easily generalised to other frequencies or intensities. This simple equation happens to entail rich physics.

The energy diagram of a hydrogen atom interacting with photons of different frequencies is depicted on the left hand side of fig. 4.1. When this atom is set to interact with a laser pulse, depending on the frequency and intensity of the laser, its behaviour varies considerably. At frequencies below the threshold, fig. 4.1a, the electron can only be ionised by absorbing multiple photons. The corresponding photo-electron spectrum of the atom shows a prominent peak at an energy equal to a multiple of the photon energy. As the intensity increases the above-threshold ionisation (ATI) becomes possible, fig. 4.1b. The ATI is a process in which ionised electrons absorb further photons (for a review of the ATI please refer to ref. [113]). The photo-electron spectrum of such a process shows a set of consecutive peaks separated by the photon energy. If the intensity is increased even further, one of two things will happen, either the lowest-energy peak in the continuum disappears, or an interference pattern will be modulated on the spectrum of that peak; both as a result of what is known as the dynamic Stark effect (DSE). The DSE is a process in which the energy levels of an atom change in the presence of a laser field [34]. The DSE will be thoroughly explained later in the chapter. The interference pattern just mentioned can be understood as a double-slit scenario in the time domain, that is, a wave packet ionised during the rising part of the laser pulse interferes with a wave packet released during the falling part of the laser pulse. This process is known as dynamic interference and is the topic of this chapter.

At higher frequencies of the laser pulse one might naively expect to see a similar

transition from the multi-photon regime (fig. 4.1c) to the ATI regime (fig. 4.1d).¹ However, due to the small continuum-continuum coupling, the dynamic Stark shift takes effect before any ATI is observed, again giving rise to a photo-electron spectrum unveiling a unique interference pattern, i.e. dynamic interference.

It is natural to expect to observe dynamic interference at above-threshold frequencies during the transition from the multi-photon regime (fig. 4.1e) to the ATI regime (fig. 4.1f). But we will be surprised to find out that there is no dynamic interference in this particular case despite the presence of the dynamic Stark shift. As will be clear later in the chapter, dynamic interference needs a very delicate balance between the strength of the dynamic Stark shift and the rate at which the initial state is depopulated. This balance turns out to be disturbed at above-threshold frequencies.

The main tool used to study the aforementioned phenomena is the perturbation theory, in which more and more terms are required to describe the behaviour of the system as the intensity of the laser increases [34]. Over the years the perturbation theory has been successfully applied to different systems of various complexity.² In the case of an atom interacting with a laser pulse, the first order perturbation theory describes the effect of the applied perturbation on the particle under study while completely ignoring the interaction of the perturbation with the energy levels of the potential the particle is in. The second order perturbation theory on the other hand, takes care of any change the energy levels may undergo. We will repeatedly take advantage of the second order perturbation theory in this chapter to discuss the phenomenon of dynamic interference.

This chapter is organised such that the reader will be first familiarised with the essential tools needed to describe dynamic interference. In section 4.1 the interaction picture is introduced, which is then used to develop the perturbation theory in section 4.2; even though the discussion of interaction picture and perturbation theory can be found in almost any textbook on quantum mechanics (for example ref. [34]), we will reiterate some of the key concepts that we will use later in the chapter. In multi-level systems it may not always be insightful or even possible to include all the states in the calculation. The unwanted states can be omitted using the method of essential states presented in section 4.3. This method is used

1. We are only considering an off-resonant case here.

2. The breakdown of the perturbation theory will not be discussed in this chapter and the interested reader is referred to ref. [114].

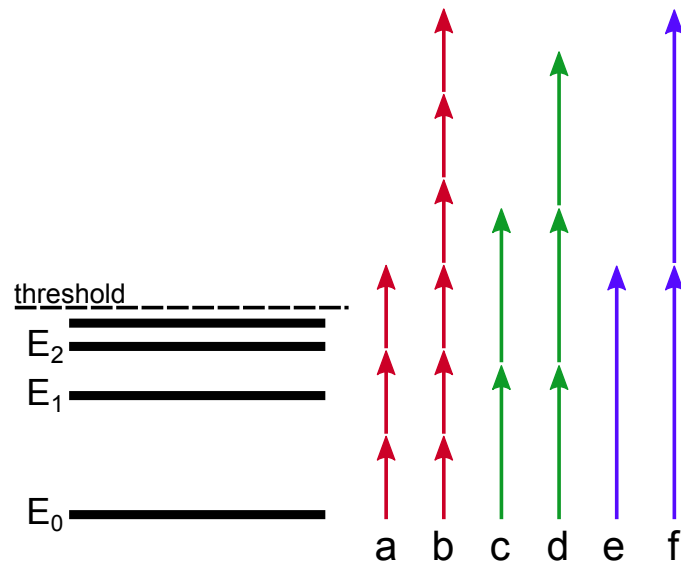


Figure 4.1. The atomic structure and the threshold of atomic hydrogen are shown on the left. Depending on the frequency and the intensity of the laser pulse, the response of the atom varies considerably. (a) low-frequency low-intensity multi-photon ionisation, (b) low-frequency high-intensity multi-photon ionisation leading to above-threshold ionisation, (c) high-frequency low-intensity multi-photon ionisation, (d) high-frequency high-intensity multi-photon ionisation leading to above-threshold ionisation, (e) low-intensity single-photon ionisation, (f) high-intensity single-photon ionisation giving rise to above-threshold ionisation

in section 4.4 to derive the Stark shift of the ground and continuum states of the hydrogen atom in the presence of a laser pulse. Different gauges used in the atomic physics community are given in section 4.5, where it is shown how Stark shift is a gauge-dependent quantity. The notion of interference in the time domain assisted by the dynamic Stark shift is discussed in section 4.6, where it is explained how some of the previous *analytical* studies went wrong. This is followed in section 4.7 by a deeper discussion of the aforementioned gauges, where it will be seen why some of the previous *numerical* studies falsely reported dynamic interference. The required conditions for dynamic interference are derived in section 4.8. Section 4.9 deals with a handful of different settings in which interference in the photo-electron spectrum is possible in a general sense. The chapter is concluded in section 4.10.

4.1. Interaction picture

Consider a Hamiltonian consisting of two parts $H = H_0 + V(t)$, where only the eigenfunctions of H_0 are known, $H_0 |\phi_n\rangle = E_n |\phi_n\rangle$. The wave function of the system $|\psi(t)\rangle$ evolves according to the Schrödinger equation, and it is customary to express this evolution in terms of the eigenfunctions of H_0 ,

$$|\psi(t)\rangle = \sum_n a_n(t) |\phi_n(t)\rangle, \quad (4.1)$$

where $|\phi_n(t)\rangle = U_0(t, t_0) |\phi_n\rangle$, and $U_0(t, t_0) = e^{-iH_0(t-t_0)}$ is the time evolution operator associated with H_0 . The basis defined by $|\phi_n(t)\rangle$ forms the interaction picture. Substituting the above expansion in the Schrödinger equation for H and using the Schrödinger equation for H_0 to simplify the obtained expression yields

$$i \sum_n \dot{a}_n(t) |\phi_n(t)\rangle = \sum_n a_n(t) V(t) |\phi_n(t)\rangle, \quad (4.2)$$

multiplying both sides by $\langle\phi_m(t)|$ from the right to derive the equation of motion of $a_n(t)$

$$i\dot{a}_m(t) = \sum_n a_n(t) \langle\phi_m(t)| V(t) |\phi_n(t)\rangle, \quad (4.3)$$

where it can be seen that the evolution of the wave function in the interaction picture is completely determined by the perturbation. This differential equation can be written as an integral equation

$$a_m(t) = a_m(t_0) - i \sum_n \int_{t_0}^t dt' \langle\phi_m(t')| V(t') |\phi_n(t')\rangle a_n(t'), \quad (4.4)$$

where $a_m(t_0)$ are the initial conditions. This integral equation is usually solved perturbatively in atomic physics [34]. In the next section the first and second order versions of the perturbation theory are presented.

4.2. Perturbation theory

The integral equation introduced in the previous section, eq. (4.4), can be solved iteratively by substituting the $a_n(t)$ under the integral with the entirety of eq. (4.4)

over and over again, yielding the following series [34]

$$\begin{aligned}
a_m(t) &= a_m(t_0) \\
&\quad - i \sum_n \int_{t_0}^t dt' \langle \phi_m(t') | V(t') | \phi_n(t') \rangle a_n(t_0) \\
&\quad - i \sum_n \int_{t_0}^t dt' \langle \phi_m(t') | V(t') | \phi_n(t') \rangle \times \\
&\quad \quad \left(-i \sum_k \int_{t_0}^{t'} dt'' \langle \phi_n(t'') | V(t'') | \phi_k(t'') \rangle a_k(t_0) \right) \\
&\quad + \dots .
\end{aligned} \tag{4.5}$$

If the series is truncated such that only the first and second terms on the right hand side are left (as a result of which the right hand side will be linear in $V(t)$), the result is called the first order perturbation theory. If the first, second, and third terms are kept (as a result of which the right hand side will be quadratic in $V(t)$), the result is called the second-order perturbation theory. Similarly, higher order versions are constructed by keeping higher numbers of terms on the right hand side. For small perturbations the series is expected to converge relatively quickly. In practice, it turns out there is rarely any need for the 3rd order (or any higher order) perturbation theory; for higher orders please refer to ref. [114]. This series can be written in a compact form after introducing some new notations,

$$a_m(t) = a_m^{(0)}(t) + a_m^{(1)}(t) + a_m^{(2)}(t) + \dots , \tag{4.6}$$

where

$$a_m^{(0)}(t) = a_m(t_0) \tag{4.7}$$

$$a_m^{(1)}(t) = -i \sum_n \int_{t_0}^t dt' \langle \phi_m(t') | V(t') | \phi_n(t') \rangle a_n^{(0)}(t') \tag{4.8}$$

$$a_m^{(2)}(t) = -i \sum_n \int_{t_0}^t dt' \langle \phi_m(t') | V(t') | \phi_n(t') \rangle a_n^{(1)}(t') , \tag{4.9}$$

or in general

$$a_m^{(j+1)}(t) = -i \sum_n \int_{t_0}^t dt' \langle \phi_m(t') | V(t') | \phi_n(t') \rangle a_n^{(j)}(t') , \tag{4.10}$$

where $a_m^{(1)}(t)$ is the first order correction (known as the first order perturbation theory), $a_m^{(2)}(t)$ is the second order correction (known as the second order perturbation

theory), etc. More often than not, the wave function is initially in the ground state, i.e. $a_m(t_0) = \delta_{mg}$, where g is the index of the ground state. In this case, $a_m^{(i)}(t)$ reduce to

$$a_m^{(0)}(t) = \delta_{mg} \quad (4.11)$$

$$a_m^{(1)}(t) = -i \int_{t_0}^t dt' \langle \phi_m(t') | V(t') | \phi_g(t') \rangle \quad (4.12)$$

$$a_m^{(2)}(t) = -i \sum_n \int_{t_0}^t dt' \langle \phi_m(t') | V(t') | \phi_n(t') \rangle a_n^{(1)}(t'). \quad (4.13)$$

The sum in the third line runs over all the states available in the system. Sometimes, however, the dynamics of only a few states is of interest (these states will be referred to as essential states in the discussion to follow), meaning while the other states (hereafter referred to as non-essential states) are just as important in the evolution of the system, an accurate knowledge of their evolution is not desired. In such a scenario, it is possible to reduce the Hamiltonian to an effective Hamiltonian which describes the essential states fully, while the non-essential states only enter through their effect on the essential states. This method is called the method of essential states [115, 116] and will be discussed in more detail in the next section.

4.3. Method of essential states

The evolution of a system under the influence of a perturbation can be expressed in terms of the eigenstates of the unperturbed system as described before. But it is not always desirable (or even feasible) to keep track of the population of every single state. For this reason, the states are often split into two groups, essential and non-essential states. The essential states are the ones we are interested in monitoring, while the non-essential states are only felt through their influence on the essential states. While $V(t)$ is defined on the entire space swept by $|\phi_n\rangle$, the dynamics of a reduced system consisting only of the essential states is governed by $V(t) + V^{\text{DSE}}(t)$, which is only defined in the space swept by $|\phi_j\rangle$ where j only runs over the essential states. The additional potential $V^{\text{DSE}}(t)$ entails the effect of all the non-essential states on the dynamics of the essential states. This additional potential gives rise to the dynamic Stark effect (see section 4.4). It is worth emphasising that there is no rule or precondition as to how to label each state essential or non-essential, see fig. 4.2. The same state might be labelled differently in different situations.

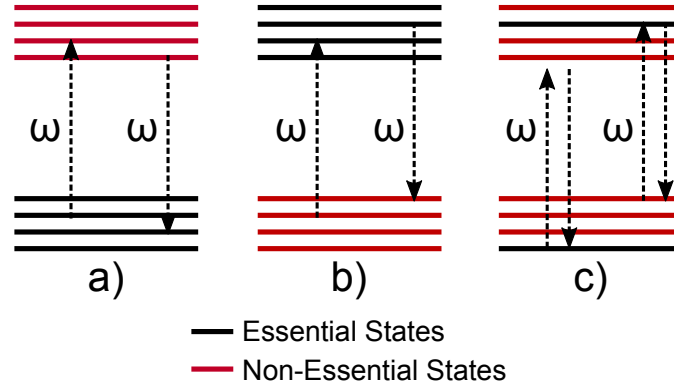


Figure 4.2. The classification of the states of a system into two groups of essential and non-essential states. The way the classification is done depends on the details of the problem. For instance, assume in the pictures above the two bands represent two electronic bands. The states are labelled essential or non-essential, because we are only interested in the dynamics of a) the first electronic band, b) the second electronic band, c) one particular state from the first band along with some other state from the second band.

In the discussion to follow, $|\phi_n\rangle$ are the eigenstates of the unperturbed system, and E_n are their corresponding eigenenergies. Subscripts j and k run over the essential states, and subscripts p and l run over the non-essential states. Subscripts n and m run over all the states (a convention similar to that of ref. [116] is used here). According to eq. (4.3) the dynamics of an essential state is given by

$$i\dot{a}_k(t) = \sum_j a_j(t) \langle \phi_k(t) | V(t) | \phi_j(t) \rangle + \sum_p a_p(t) \langle \phi_k(t) | V(t) | \phi_p(t) \rangle, \quad (4.14)$$

where the original sum has been split into two; one over the essential states and one over the non-essential states. Before we can remove the explicit dependence of the above expression on the non-essential states, we have to consider the dynamics of a non-essential state

$$i\dot{a}_p(t) = \sum_j a_j(t) \langle \phi_p(t) | V(t) | \phi_j(t) \rangle + \sum_l a_l(t) \langle \phi_p(t) | V(t) | \phi_l(t) \rangle, \quad (4.15)$$

where similarly the original sum has been divided into two over essential and non-essential states. If no transition occurs between any two non-essential states,³

3. Here we are assuming that all the non-essential states play a secondary role. That is, they are only here to facilitate the transition of population from one essential state to another. As the intensity increases, the validity of this assumption will eventually break down.

the second sum disappears immediately. Furthermore, if none of the non-essential states are initially populated, the solution of the above expression is given by

$$a_p(t) = -i \sum_j \int_{t_0}^t a_j(t') \langle \phi_p(t') | V(t') | \phi_j(t') \rangle dt'. \quad (4.16)$$

The procedure of writing the population of a non-essential state purely in terms of the essential states is called the adiabatic elimination. For further details regarding this topic please refer to ref. [117]. If $V(t) = \mathbf{r} \cdot \mathbf{E}(t)$, that is, the interaction in the length gauge of a dipole with a laser pulse defined by $\mathbf{E}(t) = \mathcal{E}(t) \cos(\omega t)$, the solution of this integral equation is given in the case of a slowly-varying envelope $\mathcal{E}(t)$ by [116] (see appendix E for a discussion of the slowly-varying envelopes)

$$a_p(t) = \frac{1}{2} \sum_j a_j(t) \left(\frac{e^{-i(\omega_{jp}-\omega)t}}{\omega_{jp}-\omega} + \frac{e^{-i(\omega_{jp}+\omega)t}}{\omega_{jp}+\omega} \right) \langle \phi_p | \mathbf{r} \cdot \mathcal{E}(t) | \phi_j \rangle, \quad (4.17)$$

where $\omega_{jp} = E_j - E_p$. Upon substituting this expression in eq. (4.14), and applying the rotating wave approximation to remove the fast oscillating terms we get

$$i\dot{a}_k(t) = \sum_j a_j(t) e^{-i\omega_{jk}t} V_{kj}^{\text{eff}}(t), \quad (4.18)$$

where the effective potential $V_{kj}^{\text{eff}}(t) = V_{kj}(t) + V_{kj}^{\text{DSE}}(t)$ consists of two terms, the dipole interaction

$$V_{kj}(t) = \langle \phi_k | \mathbf{r} \cdot \mathbf{E}(t) | \phi_j \rangle, \quad (4.19)$$

and the dynamic Stark effect

$$V_{kj}^{\text{DSE}}(t) = \frac{1}{4} \sum_p \left(\frac{1}{\omega_{jp}-\omega} + \frac{1}{\omega_{jp}+\omega} \right) \langle \phi_k | \mathbf{r} \cdot \mathcal{E}(t) | \phi_p \rangle \langle \phi_p | \mathbf{r} \cdot \mathcal{E}(t) | \phi_j \rangle. \quad (4.20)$$

The diagonal elements of this matrix correspond to the so-called dynamic Stark shift, whereas the off-diagonal elements describe non-resonant two-photon absorption/emission. This additional potential is only felt by the essential states as a direct consequence of neglecting the non-essential states. Any failure to include this term jeopardises the validity of the final results. As a side note, if the non-essential states had been labelled essential, there would be no need for this additional term, and a similar expression in the dynamics of the states would have emerged from the second order perturbation theory.

If the system has continuum as well as discrete states, the sum in eq. (4.20) should be understood as a sum over the discrete states and an integral over the

continuum. In this case, the denominator of at least one of the two fractions on the right hand side vanishes at $\omega_{jp} = \pm\omega$ if $E_p = E_j \mp \omega$ happens to lie in the continuum. The integral can then be evaluated by resorting to complex analysis,

$$V_{kj}^{\text{DSE}}(t) = \frac{1}{4} \text{p.v.} \sum_p \left(\frac{1}{\omega_{jp} - \omega} + \frac{1}{\omega_{jp} + \omega} \right) \langle \phi_k | \mathbf{r} \cdot \mathcal{E}(t) | \phi_p \rangle \langle \phi_p | \mathbf{r} \cdot \mathcal{E}(t) | \phi_j \rangle - i\pi \frac{1}{4} \int_0^\infty dE_p (\delta(\omega_{jp} - \omega) + \delta(\omega_{jp} + \omega)) \langle \phi_k | \mathbf{r} \cdot \mathcal{E}(t) | \phi_p \rangle \langle \phi_p | \mathbf{r} \cdot \mathcal{E}(t) | \phi_j \rangle, \quad (4.21)$$

where p.v. indicates the principle value, and the imaginary part is the residue of the integral. The real part of this potential induces a shift in the energy of the essential states, and the imaginary part causes the population of the essential states as a whole to increase/decrease over time. This expression has been derived before [34].

4.4. Stark shift of the ground and continuum states

If the ground state is the only state labelled essential in a system, all the other states affect the dynamics of the ground state through the effective potential. We will soon see what the effective potential looks like in this case if the ground state is the only state initially populated.

As explained in the last section, the effective potential has two terms, the dipole interaction and the dynamic Stark effect. The dipole interaction $V_{gg}(t)$ is zero in the case of a Hydrogen atom in a laser field due to the selection rules governing the transitions (the laser pulse cannot couple the ground state to itself). The dynamic Stark effect resulting from the coupling of the ground state to all the non-essential states is given by

$$V_{gg}^{\text{DSE}}(t) = \frac{1}{4} \text{p.v.} \sum_p \left(\frac{1}{\omega_{gp} - \omega} + \frac{1}{\omega_{gp} + \omega} \right) |\langle \phi_p | \mathbf{r} \cdot \mathcal{E}(t) | \phi_g \rangle|^2 - i\pi \frac{1}{4} \int_0^\infty dE_p \delta(\omega_{gp} + \omega) |\langle \phi_p | \mathbf{r} \cdot \mathcal{E}(t) | \phi_g \rangle|^2, \quad (4.22)$$

where the term containing $\delta(\omega_{gp} - \omega)$ is neglected because it is zero throughout the entire region defined by the limits of the integral. Substituting this expression in eq. (4.18) yields the equation of motion for the ground state

$$i\dot{a}_g(t) = \frac{\mathcal{E}^2(t)}{4\omega^2} (\delta_g^{\text{len}} - i\gamma_g^{\text{len}}/2) a_g(t), \quad (4.23)$$

where

$$\delta_g^{\text{len}} = \omega^2 \text{p.v.} \sum_p \left(\frac{1}{\omega_{gp} - \omega} + \frac{1}{\omega_{gp} + \omega} \right) |\langle \phi_p | \mathbf{r} \cdot \hat{e} | \phi_g \rangle|^2 \quad (4.24)$$

$$\gamma_g^{\text{len}} = 2\pi\omega^2 \int_0^\infty dE_p \delta(\omega_{gp} + \omega) |\langle \phi_p | \mathbf{r} \cdot \hat{e} | \phi_g \rangle|^2, \quad (4.25)$$

and the unit vector \hat{e} is the direction of the electric field. The superscript “len” reminds us that these quantities are evaluated in the length gauge (in a later section we will see how these quantities are related to their counterparts from the velocity gauge). The solution to this differential equation is

$$a_g(t) = \exp \left(-i(\delta_g^{\text{len}} - i\gamma_g^{\text{len}}/2) \frac{1}{4\omega^2} \int^t \mathcal{E}^2(t') dt' \right), \quad (4.26)$$

where it is clearly seen that δ_g^{len} represents a shift in the energy experienced by the ground state (the so-called dynamic Stark shift), and γ_g^{len} is responsible for the depletion of the ground state.

Naturally we would expect a similar expression for the evolution of the population of the continuum states

$$a_\epsilon(t) = \exp \left(-i(\delta_\epsilon^{\text{len,vel}} - i\gamma_\epsilon^{\text{len,vel}}/2) \frac{1}{4\omega^2} \int^t \mathcal{E}^2(t') dt' \right), \quad (4.27)$$

where the appropriate expressions for $\delta_\epsilon^{\text{len,vel}}$ and $\gamma_\epsilon^{\text{len,vel}}$ should in principle result from a similar analysis for the effective potential affecting a continuum state in the absence of any other state. However, this procedure proves more tedious in practice than expected. And therefore we will take a somewhat different route which turns out to be much easier in the particular case of the continuum states. As part of this procedure, we will have to go to the velocity gauge.

The high energy continuum states of an atom can be safely approximated by free electron states, which means in the case of an atom in the presence of an oscillating electric field, the core potential can be neglected, leaving behind the Hamiltonian of a free electron in an electric field $H = (\mathbf{p} + \mathbf{A}(t))^2/2$, where \mathbf{p} is the momentum of the electron, and $\mathbf{A}(t) = \mathcal{A}(t) \sin(\omega t)$ is the vector potential representing the oscillating electric field. The eigenstates of such a Hamiltonian are the so-called Volkov states [104],

$$\mathcal{V}_\epsilon(t) = (2\pi)^{-3/2} \exp \left(-i \int \frac{1}{2} (\mathbf{p} + \mathbf{A}(t'))^2 dt' \right), \quad (4.28)$$

where ϵ is the energy of the electron and is related to its momentum by $\epsilon = p^2/2$. The normalising factor makes sure $\langle \mathcal{V}_\epsilon(t) | \mathcal{V}_{\epsilon'}(t) \rangle = \delta(\mathbf{p} - \mathbf{p}')$. If the envelope of the laser pulse varies only slowly on a time scale of the order of the period of the carrier wave, the Volkov solutions can be simplified to

$$\mathcal{V}_\epsilon(t) = (2\pi)^{-3/2} \exp\left(-i\frac{p^2}{2}t - i\frac{1}{4\omega^2} \int^t \mathcal{E}(t')^2 dt'\right), \quad (4.29)$$

where $\mathcal{E}(t) = \omega\mathcal{A}(t)$, and the effect of $\mathbf{p} \cdot \mathbf{A}(t)$ is averaged out. By going to the interaction picture, the trivial temporal phase $p^2t/2$ caused by the unperturbed Hamiltonian can be removed, leaving us with

$$a_\epsilon(t) = (2\pi)^{-3/2} \exp\left(-i\frac{1}{4\omega^2} \int^t \mathcal{E}(t')^2 dt'\right). \quad (4.30)$$

In other words, the only term from the Hamiltonian affecting the field-free state is $A^2/2$. By comparing this expression with eq. (4.27), it is clear that $\delta_\epsilon^{\text{vel}} = 1$ and there is no depletion of the continuum states, i.e. $\gamma_\epsilon^{\text{vel}} = 0$. Then, the equation of motion for this state becomes

$$i\dot{a}_\epsilon(t) = \frac{\mathcal{E}^2(t)}{4\omega^2} \delta_\epsilon^{\text{vel}} a_\epsilon(t), \quad (4.31)$$

from which the corresponding dynamic Stark effect potential can be derived

$$V_{\epsilon\epsilon}^{\text{DSE}}(t) = \frac{\mathcal{E}^2(t)}{4\omega^2} \delta_\epsilon^{\text{vel}}. \quad (4.32)$$

It cannot be stressed enough that the effective potentials derived in this section are only valid if all the other states are labelled unessential. In a situation where there are other states labelled essential, the effective potentials have to be modified accordingly.

It is worth reiterating that the Stark shift of the ground state was evaluated in the length gauge, while the Stark shift of the continuum was derived in the velocity gauge. In the next section different gauges will be studied, and it will be shown how quantities from different gauges can be converted to each other.

4.5. Choice of gauge in atomic physics

The choice of gauge in quantum mechanics should be irrelevant as long as a complete basis is used. That being said, it is almost impossible to use a complete basis

in a numerical simulation. While one gauge might adequately describe a given problem using a limited basis, another gauge might completely fail to encompass any of the necessary details of the dynamics using the same basis [118].

There are mainly two gauges used in atomic physics, the length gauge and the velocity gauge. The interaction of an electron with an electric field in the length gauge is introduced by adding $\mathbf{r} \cdot \mathbf{E}$ to the field-free Hamiltonian, and the velocity gauge is obtained by making the substitution $\mathbf{p} \rightarrow \mathbf{p} + \mathbf{A}$ in the field-free Hamiltonian. The vector potential and consequently the electric field should be generally understood as time dependent, and it is only for brevity that their time dependence is not explicitly written here.

If the Hamiltonian in the velocity gauge is written out, $H = p^2/2 + \mathbf{p} \cdot \mathbf{A} + A^2/2 + V(r)$, it is seen that there is a spatially independent additive term $A^2/2$ which introduces a trivial phase factor to all states; this term can be easily included in the simulation merely by adding a phase to the wave function every time step. In the case of a continuum state, this phase comprises almost the entire Stark shift (refer to the last section), whereas the Stark shift of any bound state consists of two contributions, one resulting from $\mathbf{p} \cdot \mathbf{A}$ and one caused by the additive term $A^2/2$.

This additive term can be easily transformed away by the unitary transformation $T(t) = \exp\left(\frac{i}{2} \int^t A^2(t') dt'\right)$ [119], giving rise to a new gauge that we shall refer to as the *reduced gauge*. The Hamiltonian in this gauge reads $H = p^2/2 + \mathbf{p} \cdot \mathbf{A} + V(r)$. If $|\psi^{\text{vel}}(t)\rangle$ and $|\psi^{\text{red}}(t)\rangle$ represent the wave function in the velocity and reduced gauges respectively, the two are related according to

$$|\psi^{\text{red}}(t)\rangle = T(t) |\psi^{\text{vel}}(t)\rangle . \quad (4.33)$$

In the reduced gauge the continuum states do not undergo much Stark shift (since $A^2(t)/2$ is the only term in the Hamiltonian affecting the field-free states; refer to the last section), i.e. $\delta_\epsilon^{\text{red}} \approx 0$, as a result of which we infer $\delta_g^{\text{red}} = \delta_g^{\text{vel}} - 1$ (remember that $\delta_g^{\text{vel}} = 1$). And the only term introducing a Stark shift to the bound states is $\mathbf{p} \cdot \mathbf{A}$.

By contrast, in the length gauge the entire coupling is encapsulated by the position-dependent term $\mathbf{r} \cdot \mathbf{E}$. If the solution to the Schrödinger equation in the length and velocity gauges are respectively denoted by $|\psi^{\text{len}}(t)\rangle$ and $|\psi^{\text{vel}}(t)\rangle$, the

two are related by [119]

$$|\psi^{\text{len}}(t)\rangle = e^{i\mathbf{r}\cdot\mathbf{A}} |\psi^{\text{vel}}(t)\rangle, \quad (4.34)$$

where $e^{i\mathbf{r}\cdot\mathbf{A}}$ is referred to as the gauge operator hereafter. The Stark shift of any state in the length gauge is identical to its counterpart from the velocity gauge as a direct consequence of gauge invariance of the time dependent Schrödinger equation, that is $\delta_g^{\text{len}} = \delta_g^{\text{vel}}$ and $\delta_\epsilon^{\text{len}} = \delta_\epsilon^{\text{vel}}$. One main difference between the length and velocity gauges is that the coupling in the length gauge cannot be broken up similarly to what was done in the velocity gauge. This means, the term as a whole has to be covered by any convergent simulation; the implications of this fact will be discussed in section 4.7.

It can be confirmed by direct substitution that the Stark shift of any state relative to another state (e.g. $\delta_g \equiv \delta_g^{\text{len,vel,red}} - \delta_\epsilon^{\text{len,vel,red}}$) remains unaltered in all of the above gauges. This difference will be referred to as the *effective Stark shift* in later sections. It is seen that the effective Stark shift between a bound state and a continuum state in the length and velocity gauges is almost identical to the Stark shift of the bound state in the reduced gauge due to the fact that $\delta_\epsilon^{\text{red}} = 0$. The depletion rate introduced in eq. (4.25) is identical in all of the gauges defined above ($\gamma_g \equiv \gamma_g^{\text{len}} = \gamma_g^{\text{vel}} = \gamma_g^{\text{red}}$).

The three gauges are compared in fig. 4.3 where the energies of the ground state and a continuum state of a hydrogen atom in the presence of a laser pulse are shown. Recall that the strength of the Stark shift follows the envelope of the pulse, see eq. (4.20). The Stark shift in the reduced gauge is made up of two parts, one of which is transformed away to obtain the reduced gauge.

4.6. Notion of dynamic interference: analytics

To reiterate the upshot of the discussion so far, the ground and continuum states of an atomic system undergo Stark shift to different degrees in the presence of a laser pulse. As illustrated in fig. 4.3, the effective Stark shift follows the pulse envelope. Therefore, if a Gaussian envelope is used, the carrier frequency matches the energy difference between the ground state and each continuum state once in the rising half of the envelope and once in the falling half. This very observation lead Demekhin and Cederbaum in 2012 [9] to propose a double-slit setup in the

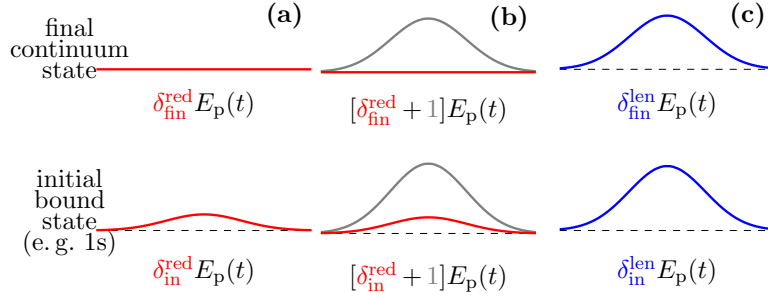


Figure 4.3. The Stark shift experienced by the initial bound and final continuum states as functions of the ponderomotive energy $E_p(t) = \frac{\mathcal{E}^2(t)}{4\omega^2}$. The Stark shift follows the envelope of the Gaussian laser pulse, hence the bell-shaped curves. a) the reduced gauge, b) the velocity gauge, c) the length gauge. Note that the effective Stark shift (the difference in energy between the initial bound and final continuum states) remains identical in all the gauges above. [120]

time domain by having the electron ionised directly to the continuum by an above-threshold Gaussian laser pulse. Each continuum state would then be dominantly populated at two points in time, i.e. when the energy difference between the ground state and the continuum state matches the laser frequency. Depending on the time delay between these two points, the two contributions interfere constructively or destructively, giving rise to a distinct interference pattern in the photo-electron spectrum. The authors further expanded their idea in refs. [10, 11].

Demekhin and Cederbaum described their findings by a minimal model consisting of two states, the ground state and a continuum state of energy ϵ ,

$$\begin{aligned} a_g(t) &= \exp(-\gamma_g^{\text{len}} J(t)/2) \\ a_\epsilon(t) &= -id_\epsilon \int_{t_0}^t dt' e^{-i\delta_g^{\text{len}} J(t')} E(t') e^{-i(E_g - \epsilon)t'} a_g(t), \end{aligned} \quad (4.35)$$

where $d_\epsilon = \langle \phi_{p=\sqrt{2\epsilon}} | \mathbf{r} \cdot \hat{\mathbf{e}} | \phi_g \rangle$ and $J(t) = \frac{1}{4\omega^2} \int_{t_0}^t \mathcal{E}^2(t') dt'$. Here the ground state decays exponentially as a result of its coupling to the continuum. The effect of all the states not included in the model is thought to have been injected through the Stark shift of the ground state δ_g^{len} . Another group seems to have followed the same idea [12].

In the following we will derive this model starting from first principles using

the method of essential states followed by the perturbation theory. Every state is labelled non-essential but the ground state and a p -continuum state of energy ϵ . The effect of the non-essential states on the essential states are expressed through the addition of an effective potential introduced in section 4.3. In this two-level system (the ground state and the single continuum state), the dynamic Stark effect potential has only three distinct elements, namely V_{gg}^{DSE} , $V_{g\epsilon}^{\text{DSE}}$, and $V_{\epsilon\epsilon}^{\text{DSE}}$. We shall evaluate these elements one by one.

The first element V_{gg}^{DSE} has been evaluated before in eq. (4.22) for the case of the ground state being the only essential state. But that scenario is different from the current model in that in eq. (4.22) every single continuum state is included in the sum over non-essential states, but in the current model there is one continuum state labelled essential which should not appear in that sum. However, the contribution of this single continuum state to the total sum is considered negligible, and therefore eq. (4.22) is used as an approximation to V_{gg}^{DSE} .

The second element $V_{g\epsilon}^{\text{DSE}}$ vanishes due to the selection rules governing a hydrogen-like atom.

The third element can be evaluated similarly to the first element. The Stark shift of a continuum state when all other states are labelled non-essential was derived in section 4.4. But here the ground state is also labelled essential. However, if we assume the contribution of the ground state to the Stark shift of the continuum state is negligible, we can use eq. (4.32) as an approximation to $V_{\epsilon\epsilon}^{\text{DSE}}$.

The real part of V^{DSE} will be used along with the field-free Hamiltonian to define the interaction picture. That is, the wave function is given by $|\psi(t)\rangle = a_g(t) |\varphi_g(t)\rangle + a_\epsilon(t) |\varphi_\epsilon(t)\rangle$, where the two basis functions are defined by

$$|\varphi_g(t)\rangle = e^{-i\delta_g^{\text{len}} J(t)} e^{-iE_g t} |\phi_g\rangle \quad (4.36)$$

$$|\varphi_\epsilon(t)\rangle = e^{-i\delta_\epsilon^{\text{len}} J(t)} e^{-i\epsilon t} |\phi_\epsilon\rangle. \quad (4.37)$$

$|\phi_g\rangle$ and $|\phi_\epsilon\rangle$ are the eigenfunctions of the field-free Hamiltonian. The evolution of the system in the interaction picture is then dictated by the dipole interaction between the ground state and continuum as well as the imaginary part of V^{DSE} ,

$$i\dot{a}_g(t) = -i \frac{\mathcal{E}^2(t)}{4\omega^2} \frac{\gamma_g^{\text{len}}}{2} a_g(t) + \langle \varphi_g(t) | \mathbf{r} \cdot \mathbf{E}(t) | \varphi_\epsilon(t) \rangle a_\epsilon(t) \quad (4.38)$$

$$i\dot{a}_\epsilon(t) = \langle \varphi_\epsilon(t) | \mathbf{r} \cdot \mathbf{E}(t) | \varphi_g(t) \rangle a_g(t). \quad (4.39)$$

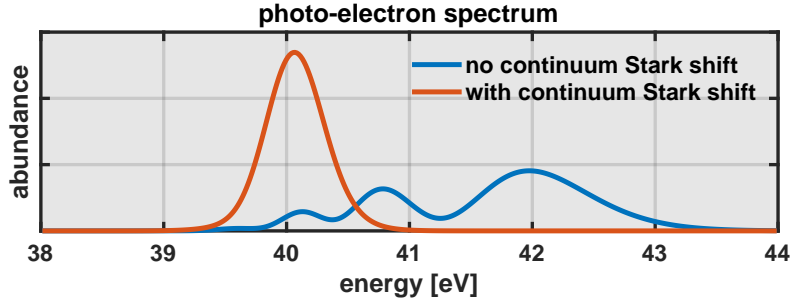


Figure 4.4. The comparison of the two analytic models, i.e. eq. (4.35) (blue) and eq. (4.42) (red), for $I = 7 \times 10^{16}$ W/cm², $\tau = 3$ fs, and $\omega = 53.6057$ eV. The wrongly observed dynamic interference vanishes as soon as the Stark shift of the continuum is taken into account.

These equations are not ready for the perturbation theory just yet. They have to be further simplified by multiplying the first equation from the left by $f(t) \equiv \exp(\gamma_g^{\text{len}} J(t)/2)$ and subsequently making the substitution $\tilde{a}_g(t) \equiv f(t)a_g(t)$. After a few lines of math we get

$$i\dot{\tilde{a}}_g(t) = \langle \varphi_g(t) | \mathbf{r} \cdot \mathbf{E}(t) | \varphi_\epsilon(t) \rangle f(t) a_\epsilon(t) \quad (4.40)$$

$$i\dot{\tilde{a}}_\epsilon(t) = \langle \varphi_\epsilon(t) | \mathbf{r} \cdot \mathbf{E}(t) | \varphi_g(t) \rangle f(t) \tilde{a}_g(t). \quad (4.41)$$

The perturbation theory can now be used to approximate the solution of these two coupled differential equations. The solution up to the first order, eq. (4.12), is (after substituting for $\tilde{a}_g(t)$)

$$\begin{aligned} a_g(t) &= \exp(-\gamma_g^{(l)} J(t)/2) \\ a_\epsilon(t) &= -id_\epsilon \int_{t_0}^t dt' e^{-i(\delta_g^{\text{len}} - \delta_\epsilon^{\text{len}})J(t')} E(t') e^{-i(E_g - \epsilon)t'} a_g(t). \end{aligned} \quad (4.42)$$

It is clearly seen that Demekhin and Cederbaum arrived at the wrong exponent inside the integral appearing in the expression for the population of the continuum state (cf. eqs. (4.35)). This mistake is equivalent to no Stark shift for the continuum state, which caused their model to overestimate the effective Stark shift (defined as the difference in energy between the Stark shifted ground state and the Stark shifted continuum). Had the authors used the reduced gauge introduced in the previous section ($\delta_g^{\text{len}} \rightarrow \delta_g^{\text{red}}$), their model would have been accurate (remember that continuum states don't undergo substantial Stark shift in the reduced gauge).

In fig. 4.4 two spectra are shown, one predicted by eq. (4.35) and one predicted by eq. (4.42). Dynamic interference is seen in the former due to the overestimation

of the effective Stark shift, whereas if the latter expression is used there is no dynamic interference. It will be explained in section 4.8 why this overestimation gave rise to erroneous dynamic interference.

The integral governing the population of the continuum states can be approximated by the stationary phase method, in which the integral is approximated by an expression containing the value of the integrand at a few points [119]. These points are where the derivative of the phase of the integrand vanishes. Given that this integral can be numerically evaluated almost instantly on a normal laptop, the stationary phase approximation may not seem to be of much practical use. However, it provides much insight into the mechanism behind the dynamic interference. It presents the phenomenon of dynamic interference as a temporal equivalent of the double-slit experiment, where the two temporal slits are the stationary points of the phase of the integrand. The phase of the integrand in the case of a linearly polarised laser pulse is given by $\phi(t) = (\delta_g^{\text{len}} - \delta_\epsilon^{\text{len}})J(t) + \omega t + (E_g - \epsilon)t$, see eqs. 4.42. The derivative of this phase can be easily calculated,

$$\begin{aligned} \frac{d\phi(t)}{dt} &= (\delta_g^{\text{len}} - \delta_\epsilon^{\text{len}}) \frac{1}{4\omega^2} \mathcal{E}^2(t) + \omega + (E_g - \epsilon) \\ &= (\delta_g^{\text{len}} - \delta_\epsilon^{\text{len}}) \frac{\mathcal{E}_0^2}{4\omega^2} \exp(-2t^2/\tau^2) + \omega + (E_g - \epsilon), \end{aligned} \quad (4.43)$$

where it is assumed the pulse has a Gaussian envelope $\mathcal{E}(t) = \mathcal{E}_0 \exp(-t^2/\tau^2)$. The laser pulse has a second term (resulting from the first term in $\cos(\omega t) = (e^{i\omega t} + e^{-i\omega t})/2$) which is ignored here because the phase of the integrand with that term never vanishes; that term corresponds to losing a photon in going from the ground state to the continuum. For any given continuum state, the derivative vanishes at most at two points given by

$$t_\pm = \pm \tau \sqrt{-\frac{1}{2} \ln \frac{\epsilon - E_g - \omega}{\delta_g E_p}}, \quad (4.44)$$

where $\delta_g = \delta_g^{\text{len}} - \delta_\epsilon^{\text{len}} \approx \delta_g^{\text{red}}$ is the effective Stark shift and $E_p = \frac{\mathcal{E}_0^2}{4\omega^2}$ is the ponderomotive energy. Depending on how far apart these two points are, their contributions to the photo-electron spectrum either add up or cancel out. This distance is a function of the energy of the continuum state, hence an interference pattern along the energy axis in the spectrum.

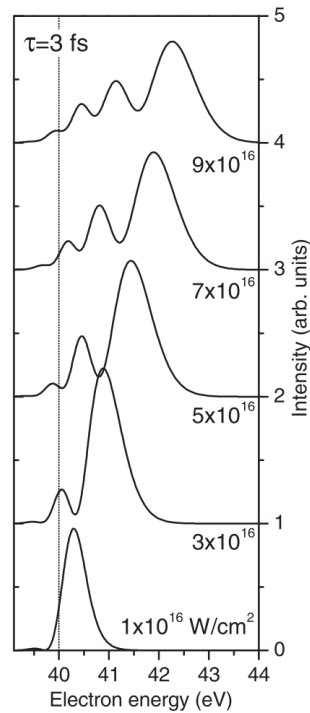


Figure 4.5. Photo-electron spectra of the hydrogen atom targeted by a Gaussian pulse of width $\tau = 3$ fs, and carrier frequency $\omega = 53.6057$ eV calculated in ref. [10]. It turns out these spectra are the results of unconverged numerical simulations. See fig. 4.7 for the converged results.

4.7. Notion of dynamic interference: numerics

The numerical simulations done in refs. [9, 10] seem to confirm Demekhin and Cederbaum's conjecture by showing an interference pattern in the photo-electron spectrum, see fig. 4.5. The simulations were carried out in the length gauge. The continuum part of the basis only consisted of the p -continuum, which might seem to be large enough for the current scenario.⁴ After all, the electron residing in the ground state is expected to end up in a p -state after absorbing a photon.

4. No numerical basis is ever complete, as a complete basis would usually contain an infinite number of basis functions. Therefore, the complete basis is usually reduced to the smallest number of basis functions required to describe the dynamics of the system correctly. This reduction is not always trivial and great care is needed not to misjudge the importance of the neglected basis functions.

This picture is only true for low intensities of the laser pulse. As the intensity increases, the two-photon process of dynamic Stark effect becomes so important that it can no longer be neglected. This means, the Stark shift of any state involved in the simulation has to be explicitly addressed. Even though it is possible in the analytics to split the states into two groups of essential and non-essential states, such a decomposition is not possible in the numerics. And there should be enough states included in the simulation to account for the Stark shift of the states labelled essential in the analytics. Demekhin and Cederbaum briefly discuss this issue by monitoring the effect of the included bound states on the final spectrum [10]. The discrete part of their basis is gradually expanded from only containing the $1s$ -state to containing ns -states with $n \leq 6$, mp -states with $m \leq 100$, and kd -states with $k \leq 6$. Even though the authors never go beyond the d -states arguing that their effect is very small, they manage to successfully account for the Stark shift of the ground state. Remember that in the length gauge both the bound and continuum states experience Stark shift. But unfortunately, the authors fail to provide a similar analysis for the Stark shift of the continuum states.

For this reason, we are showing the influence of an increasing set of continuum states on the final spectrum in fig 4.6, where it can be seen that the observed dynamic interference is a direct result of non-converged simulation, and depending on the gauge used in the numerics, the number of partial waves necessary for convergence varies (see appendix F). These partial waves are in fact needed for an accurate description of the Stark shift of the continuum during the laser pulse, and any failure to include these states in the simulation makes the implicit assumption of no/wrong Stark shift for the continuum. It seems that more partial waves are needed in the length gauge (there will be more on this soon).

Evidently, the effect of the mistake in the numerical simulation has exactly matched the effect of the mistake in the analytic model, i.e. they both fail to describe the Stark shift of the continuum properly, hence the agreement between them. It does not happen very often that a faulty numerical simulation is corroborated by a faulty analytic model. And in this case, this false agreement apparently assured the authors of the validity of their findings. The converged versions of the spectra shown in fig. 4.5 can be seen in fig. 4.7, where the simulations are done in the velocity gauge with $l_{\max} = 10$ (see appendix F). Obviously they look completely different from their non-converged counterparts. The prediction of the correct analytical model introduced in the previous section is depicted with the

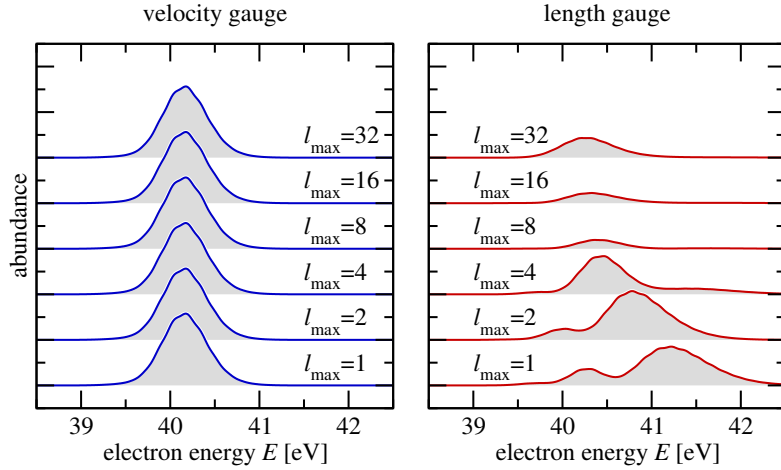


Figure 4.6. The photo-electron spectra in the velocity and length gauges for $\tau = 3$ fs, $I = 5 \times 10^{16}$ W/cm², and $\omega = 53.6$ eV. l_{\max} is the maximum angular momentum considered in the partial wave expansion, see appendix F. [120]

dash-dotted lines for two intensities.

In order to see why more partial waves are needed in the length gauge than in the velocity gauge, we have to go back to the gauge operator introduced in eq. (4.34). In the case of a linearly polarised electric field, i.e. $\mathbf{A} = A\hat{z}$, the gauge operator has a simple expansion,

$$e^{iAz} = \sum_{l=0}^{\infty} i^l \sqrt{4\pi(2l+1)} j_l(Ar) Y_l(\theta), \quad (4.45)$$

where $j_l(x)$ are the spherical Bessel function of the first kind, and $Y_l(\theta)$ is a shorthand for the spherical harmonic $Y_{l,m=0}(\theta, \phi)$; the dependence on the azimuthal angle ϕ drops automatically as a direct consequence of the polarisation of the electric field along the z -axis. Because of this symmetry, the wave function is usually expanded in numerical packages [91, 121] in terms of spherical harmonics,

$$\psi^{\text{vel}}(\mathbf{r}, t) = \frac{1}{r} \sum_{l=0}^{l_{\max}^{\text{vel}}} u_l^{\text{vel}}(r, t) Y_l(\theta), \quad (4.46)$$

where $u_l^{\text{vel}}(r, t)$ encompass the radial dependence of the wave function. l_{\max}^{vel} is the number of partial waves needed for convergence in the velocity gauge. Even though this upper bound should be in principle infinity, in reality its value is dictated by the available computational resources and desired accuracy.

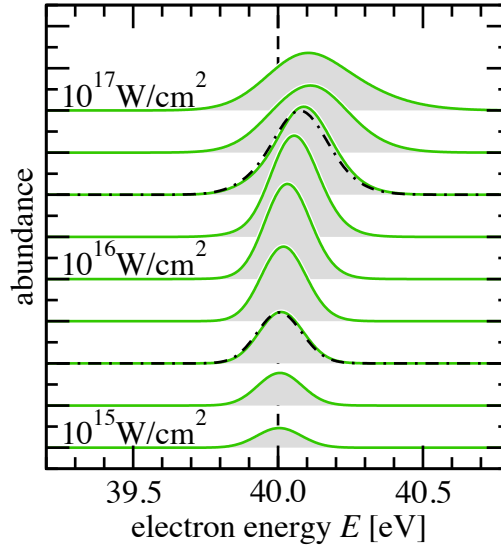


Figure 4.7. Photo-electron spectra from $1s$ hydrogen exposed to a pulse of $\tau = 10$ fs, $\omega = 53.6$ eV, and $I_k = 10^{k/4} \times 10^{15}$ W/cm 2 with $k = 0, 1, \dots, 8$. The dashed line marks the energy $E_\omega = E_{1s} + \omega = 40$ eV. The result of the minimal model (eq. (4.42)) is shown for two intensities by dot-dashed lines. [120]

The basis used for numerical simulation usually consists of the eigenfunctions of a field-free atom. The continuum states of such a system rapidly converge to the wave function of a free particle as the energy increases. For this reason, continuum states are approximated by free particle states here. The wave function of a free particle of linear momentum p , angular momentum l , and magnetic quantum number m is given in spherical coordinates by [34]

$$\phi_{plm}(r, \theta, \phi) = \sqrt{\frac{2}{\pi}} j_l(pr) Y_{lm}(\theta, \phi), \quad (4.47)$$

where $j_l(x)$ are again the spherical Bessel function of the first kind. When the electric field is polarised along the z -axis, only $m = 0$ states will be involved, and therefore there is no need to consider states with $m \neq 0$ in our analysis,

$$\phi_{pl}(r, \theta) = \sqrt{\frac{2}{\pi}} j_l(pr) Y_l(\theta), \quad (4.48)$$

where the dependence on the azimuthal angle ϕ has vanished automatically for $m = 0$.

Now, assuming the wave function in the velocity gauge is given by eq. (4.46), it

is desirable to investigate what the largest angular momentum l_{\max}^{len} of a similar expansion of the wave function in the length gauge would have to be in order to have a similar accuracy to the velocity gauge. Consider the component along the largest angular momentum required in the velocity gauge, that is $u_{l_{\max}^{\text{vel}}}^{\text{vel}}(r, t)Y_{l_{\max}^{\text{vel}}}(\theta)/r$. This component is transformed to the length gauge by using the gauge operator $e^{iAz}u_{l_{\max}^{\text{vel}}}^{\text{vel}}(r, t)Y_{l_{\max}^{\text{vel}}}(\theta)/r$. The projection of this term on the continuum basis functions determines what basis functions are occupied in the length gauge,

$$\langle \phi_{pl^{\text{len}}} | \psi^{\text{len}} \rangle = \int \phi_{pl^{\text{len}}}(r, \theta) e^{iAz} u_{l_{\max}^{\text{vel}}}^{\text{vel}}(r, t) Y_{l_{\max}^{\text{vel}}}(\theta) r dr d\Omega, \quad (4.49)$$

which can be evaluated using the expressions for the wave function of a free particle and the gauge operator to yield

$$\langle \phi_{pl^{\text{len}}} | \psi^{\text{len}} \rangle = \sqrt{\frac{2}{\pi}} \sum_{l'}^{\infty} i^{l'} \sqrt{4\pi(2l'+1)} \times \int d\Omega Y_{l^{\text{len}}}(\theta) Y_{l'}(\theta) Y_{l_{\max}^{\text{vel}}}(\theta) \int_0^{\infty} dr r j_{l^{\text{len}}}(pr) j_{l'}(Ar) u_{l_{\max}^{\text{vel}}}^{\text{vel}}(r). \quad (4.50)$$

Let's assume that $u_{l_{\max}^{\text{vel}}}^{\text{vel}}(r) \approx 0$ for $r \gtrsim R$, which is a realistic assumption as the duration of the simulation is finite and the wave function can only travel so far within this time. The value of R should be of the order of the distance that the most energetic part of the wave function travels during the life of the simulation. This assumption allows us to change the upper limit of the corresponding integral from ∞ to R . To find the upper limit of l^{len} (denoted by l_{\max}^{len}) required for convergence, we will have to look at the two integrals separately. The angular integral can be expanded in terms of the Clebsch-Gordon coefficients and can only be non-zero if the triangle inequality is satisfied, i.e. $l^{\text{len}} \leq l' + l_{\max}^{\text{vel}}$. The second integral is slightly trickier. Bearing in mind that the spherical Bessel function $j_n(\rho)$ becomes negligible for $\rho \lesssim n$,⁵ it is inferred that two conditions have to be simultaneously satisfied for the integral to be non-zero,

$$l^{\text{len}} \lesssim pR \quad (4.51)$$

$$l' \lesssim AR. \quad (4.52)$$

5. Starting from $j_n(\rho) = \sqrt{\frac{\pi}{2\rho}} J_{n+\frac{1}{2}}(\rho)$, and $|J_\nu| \leq \frac{1}{\Gamma(\nu+1)} \left(\frac{\rho}{2}\right)^\nu$, we get $j_n(\rho) \leq \frac{1}{(n+1)!} \left(\frac{\rho}{2}\right)^n$. We can now estimate the value of ρ for which $j_n(\rho) < \epsilon$ where ϵ can be arbitrarily small, by solving $n \ln\left(\frac{\rho}{2}\right) < \ln \epsilon + (n+1) \ln(n+1)$. If we assume n is large enough, this relation can be simplified to $\rho \lesssim n$.

We can use the triangle inequality introduced earlier to rewrite the second line yielding

$$l^{\text{len}} \lesssim pR \quad (4.53)$$

$$l^{\text{len}} \lesssim AR + l_{\text{max}}^{\text{vel}}. \quad (4.54)$$

The first inequality reveals that partial waves of higher angular momentum have to be included in the length gauge if the photo-electron spectrum is required to be accurate for higher energies. This has been numerically observed by Cormier and Lambropoulos before [122]. The second inequality shows that around AR more angular momenta are needed in the length gauge, i.e. $l_{\text{max}}^{\text{len}} = l_{\text{max}}^{\text{vel}} + AR$, if $l_{\text{max}}^{\text{vel}}$ angular momenta are enough in the velocity gauge. This inequality has been derived before based on a classical mechanical analysis by Muller [123]. His analysis hinges upon the fact that the canonical momentum of a free electron in the velocity gauge is a constant of motion in the absence of any scalar potential or position-dependent vector potential. The difference between the two gauges has also been discussed in the context of high-order harmonic generation [124].

To sum, it seems in the regimes where the second order perturbation theory suffices, the velocity gauge is always a better option for numerical simulation.

4.8. Necessary conditions for dynamic interference

Following the discussion at the end of section 4.6, it is clear that in order for dynamic interference to happen, there should be two points during the course of the laser pulse at which the electron is ionised into the continuum. These two points are given by eq. (4.44). Obviously, the two points should be real-valued, and they are only real-valued if $\epsilon - E_g - \omega < \delta_g E_p$. Note that the left hand side of this inequality is of the order of the spectral width of the pulse, which can be approximated by $\sim 1/\tau$ (remember that the envelope is given by $\exp(-t^2/\tau^2)$ where τ determines the temporal width of the pulse). Therefore, the following condition has to be satisfied as a prerequisite for dynamic interference,

$$E_p \tau \gtrsim \frac{1}{\delta_g}. \quad (4.55)$$

This inequality can be interpreted in a different way. The effective Stark shift is of the order of $E_p \delta_g$ and should be larger than the bandwidth of the pulse $\sim 1/\tau$ in order to be energetically resolvable.

The second quantity affecting the spectrum is the depletion rate of the ground state. The depletion rate shouldn't be too high or the initial state will be completely depopulated in the first half of the laser pulse paralysing the second slit of the temporal double-slit setup. If the laser pulse is centred at $t_{\text{mid}} = 0$, the amplitude of the ground state at the end of the first half of the pulse is $a_g(0) = \exp(-\gamma_g J(0)/2)$ where $J(0) = \frac{1}{4\omega^2} \int_{-\infty}^{t=0} \mathcal{E}^2(t') dt'$ for a Gaussian pulse of envelope $\mathcal{E}(t) = \mathcal{E}_0 \exp(-t^2/\tau^2)$. Around half of the initial population of the initial state should reach the second half of the pulse, that is $|a_g(0)|^2 \gtrsim 1/2$, leading to the following inequality

$$E_p \tau \lesssim \frac{1}{\gamma_g}, \quad (4.56)$$

where $E_p = \frac{\mathcal{E}_0^2}{4\omega^2}$ is the ponderomotive energy.

These two inequalities have to be simultaneously satisfied before any dynamic interference can happen, requiring $\delta_g > \gamma_g$. Let's have a look at the values these two quantities take in the case of atomic hydrogen for a wide range of frequencies, see fig. 4.8. They are evaluated numerically in the reduced gauge using eqs. (4.24) and (4.25). The frequency used by Demekhin and Cederbaum is marked by an arrow in the figure. Not surprisingly, the condition is not satisfied, hence the lack of dynamic interference. However, because they failed to account for the Stark shift of the continuum, the value of the effective Stark shift that entered their model was the value marked here plus one, which clearly satisfies the requirement for dynamic interference, hence their observation of dynamic interference.

It is obvious in the plot that at much higher frequencies the two inequalities can be simultaneously satisfied, and it is expected that dynamic interference will occur for a proper choice of laser parameters.

The framework discussed so far is not the only way to accommodate interference in the photo-electron spectrum. There are other somewhat different scenarios where such interference can happen more easily. Some of them are discussed in the next section.

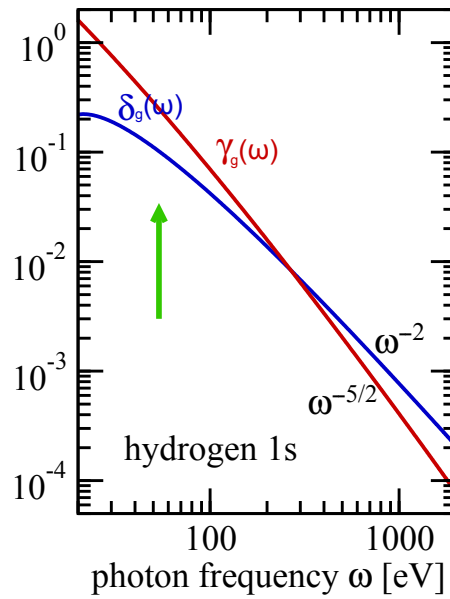


Figure 4.8. The effective Stark shift $\delta_g(\omega)$ and the depletion rate $\gamma_g(\omega)$ for the hydrogen 1s-state [120]. See eqs. (4.24) and (4.25) for the expressions for these quantities in the length gauge. Their asymptotic behaviours are given for $\omega \rightarrow \infty$ [125]. The green arrows shows the frequency $\omega = 53.6$ eV used in fig. 4.7 and in some previous publications [9, 10].

4.9. Other scenarios for interference in the photo-electron spectrum

There are several other schemes, beside the scheme discussed so far, which lead to an interference pattern in the energy domain. These schemes all share the same underlying mechanism, that is, the electron wave function is ionised into the continuum at two different times. The two ionised parts then interfere leaving behind a unique structure in the photo-electron spectrum.

4.9.1. Double pulse

The first scenario is the trivial case of a double pulse, where two consecutive pulses delayed by Δt interact with the atom. The frequency of the laser should be high enough to excite the electron directly into the continuum, which is $\omega > 13.6$ eV for atomic hydrogen. The wave packet ionised by the first pulse interferes with the

wave packet ionised by the second pulse giving rise to an interference pattern in the spectrum. If the temporal width of each of the two pulses is $\sim T$, the spectral width of the double pulse is $\sim 1/T$.

Let $\varphi_1(\epsilon)$ represent the phase of the ionised part of the wave function by the *first* pulse at the end of the *first* pulse where ϵ is the energy of the continuum state, and let $\varphi_2(\epsilon)$ be the phase of the ionised part of the wave function by the *second* pulse at the end of the *second* pulse. Assuming there is no significant Stark shift in the ground state or the continuum (in contrast to the main scheme of this chapter, the current setup does not need high intensities), we have $\varphi_2(\epsilon) = \varphi_1(\epsilon) - E_0 \Delta t$ where the second term is the phase picked up by the ground state population (of energy E_0) during the time delay between the two pulses. Furthermore, $\varphi_1(\epsilon)$ will evolve to $\varphi_1(\epsilon) - \epsilon \Delta t$ by the end of the second pulse.

Now the two contributions may interfere constructively or destructively depending on the relative phase of the two contributions. This relative phase is energy dependent,

$$\Delta\varphi(\epsilon) = (\varphi_1(\epsilon) - E_0 \Delta t) - (\varphi_1(\epsilon) - \epsilon \Delta t) = \epsilon \Delta t - E_0 \Delta t. \quad (4.57)$$

The difference in relative phase at the two ends of the spectrum dictates the number of fringes visible in the spectrum. Given the spectral width is $\sim 1/T$, this difference is $\sim \Delta t/T$. Therefore, the number of fringes is $\sim \Delta t/(2\pi T)$.

In the upper plot of fig. 4.9 a double pulse is shown where the envelope is given by $\mathcal{E}(t) = 2(t/\tau) \exp(-t^2/\tau^2)$. The intensity of the laser pulse is 10^{12} W/cm², at a frequency of 54.44 eV. The pulse width is $\tau = 0.6$ fs. For general details of the simulation please refer to appendix F. The two peaks of the envelope are located at $t_{\pm} = \pm\tau/\sqrt{2}$. The time delay between the two peaks is therefore $\Delta t = \sqrt{2}\tau$. The width of each pulse is $T = \tau/\sqrt{2}$. Thus, the number of fringes should be of the order of $1/\pi$, see the spectrum in the lower plot of fig. 4.9. The fact that this number is independent of the value of τ is a direct consequence of the particular choice of envelope. For other types of envelopes this independence may or may not follow. The careful reader might have noticed that because of the low intensity of the laser pulse used here, the corresponding photo-electron spectrum can be obtained simply by taking the Fourier transform of the laser pulse.

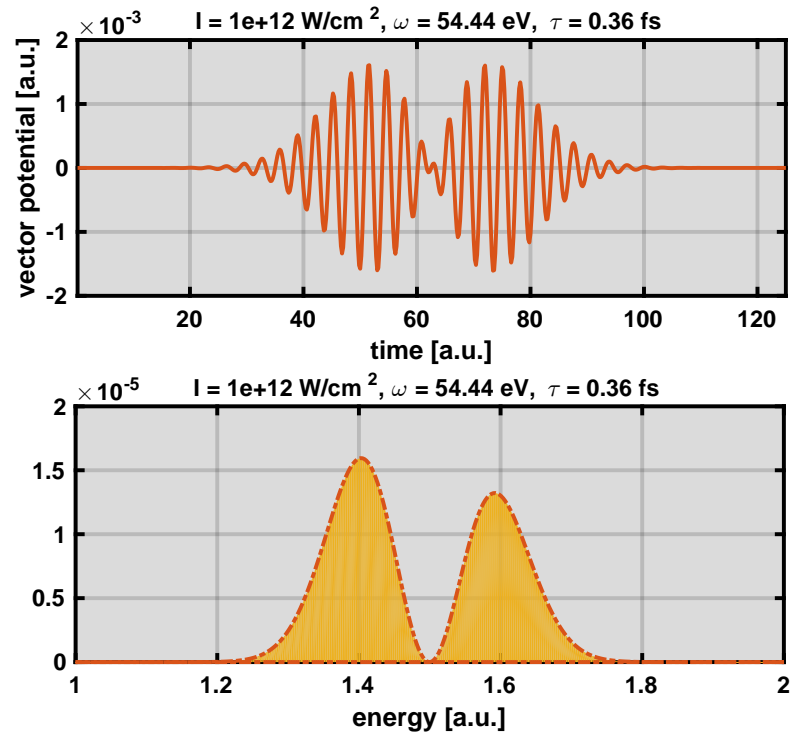


Figure 4.9. Upper plot: a double pulse used to induce a double-slit scenario in time. The further apart the two pulses are, the more fringes appear in the corresponding photo-electron spectrum. Lower plot: the corresponding photo-electron spectrum. The trough resulting from destructive interference is visible in the middle of the spectrum.

4.9.2. Stabilisation

In the previous example the ionisation mostly occurred around the two peaks of the double pulse, with no significant ionisation in between merely due to the absence of the electric field. If the aim is to reduce ionisation, suppressing the electric field is not the only option. It turns out if the intensity of the laser field exceeds a certain limit the ionisation rate plummets, a phenomenon known as the atomic stabilisation effect. It is not the purpose of this subsection to discuss this effect at length, and the interested reader is referred to ref. [93] for reviews of the topic. We will only see the atomic stabilisation as a tool to curb the ionisation rate.

If the peak intensity of the laser pulse ($\propto \mathcal{E}_0^2$) lies in the stabilisation regime, the pulse can be split into three parts. The rising part in which the instantaneous

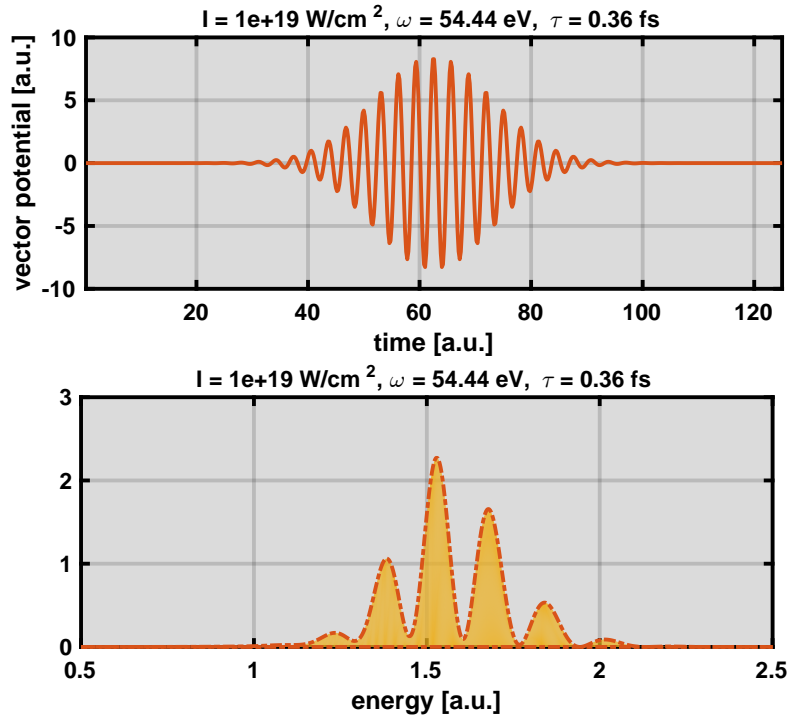


Figure 4.10. Upper plot: a Gaussian pulse. The intensity of this pulse is high enough to push the atom well into the stabilisation regime, as a result of which the ionisation rate plummets. Lower plot: the photoelectron spectrum after a hydrogen atom is targeted by such a pulse. The interference fringes are clearly visible.

intensity is lower than the threshold, the middle part in which the instantaneous intensity is higher than the threshold, and the falling part in which the instantaneous intensity is again lower than the stabilisation threshold. Not much ionisation occurs during the middle part, and this leaves the rising and falling parts effectively separated, each of which independently ionises the electron into the continuum. Similarly to the last subsection, the ionisation resulting from these two independent parts then interfere in the energy domain giving rise to a set of troughs and valleys in the spectrum.

In the upper plot of fig. 4.10 a pulse of Gaussian envelope $\mathcal{E}(t) = \exp(-t^2/\tau^2)$ is shown. The intensity of the laser pulse is $I = 10^{19} \text{ W/cm}^2$, the frequency is $\omega = 54.44 \text{ eV}$, and the width is $\tau = 0.6 \text{ fs}$. These values guarantee that the stabilisation against ionisation will indeed happen. Such a regime can be reached using free electron lasers [94, 95]. For general details of the simulation please refer

to appendix F. The intensity of the pulse is high enough for the instantaneous field strength around the peak of the pulse to exceed the stabilisation threshold, as a result of which the ionisation rate plunges. The electron is only ionised during the first part before the atomic stabilisation starts to take effect, and during the third part at the end of the pulse after the atomic stabilisation is over. Effectively, this is again equivalent to a double-slit setup in the time domain, hence interference in the photo-electron spectrum, see the lower plot of fig. 4.10. The number of fringes in the spectrum is determined by the duration of each part and the time delay between the two parts (see the discussion in the previous section).

4.9.3. Chirped pulse

In section 4.8 the criteria required for dynamic interference were discussed and it was seen that dynamic interference happens only if the effective Stark shift is stronger than the depopulation of the ground state. These two quantities are the properties of the system, and cannot be arbitrarily changed by modifying the laser parameters. However, there might just be another way to compensate for the smallness of the effective Stark shift. To see matters more clearly, let's go back to the final equation governing the population of the continuum states given in eqs. (4.42),

$$a_\epsilon(t) = -id_\epsilon \int_{t_0}^t dt' e^{-i\delta_g J(t')} E(t') e^{-i(E_g - \epsilon)t'} a_g(t), \quad (4.58)$$

where $\delta_g = \delta_g^{\text{len,vec}} - \delta_\epsilon^{\text{len,vec}}$ is the effective Stark shift. As appeared in the discussion towards the end of section 4.6, there need to be two stationary points in the phase of the integrand of the above expression if dynamic interference is to happen. One way to guarantee the existence of two real-valued stationary points is to have a large enough effective Stark shift as discussed in section 4.8. But in the event that the Stark shift is not strong enough, it may or may not be possible to alter the Stark shift merely by changing the carrier frequency. Taking a look at fig. 4.8 once again, it is apparent that if the frequency of the available laser should lie in the XUV regime, dynamic interference cannot happen.

There is, however, a second way to guarantee the existence of two real-valued stationary points. If the laser pulse is modulated with an exponential chirp, i.e. the instantaneous frequency is given by $\omega(t) = \omega_0 + \alpha \exp(-2t^2/\tau^2)$, the derivative

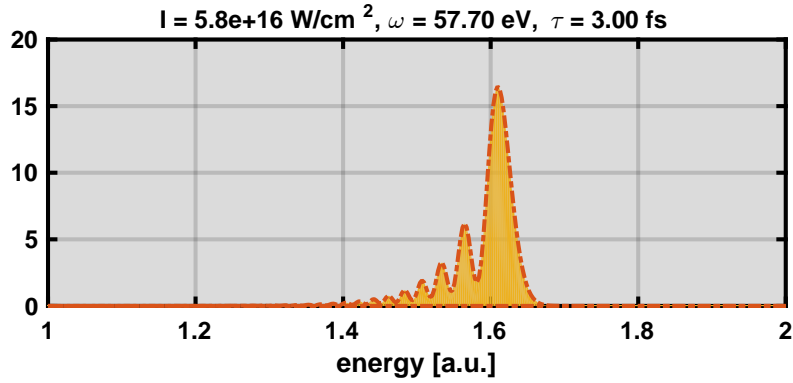


Figure 4.11. The photo-electron spectrum for hydrogen in the 1s-state after being targeted by a chirped laser pulse. The chirp strength is determined by $\alpha = 0.3$ a.u.

of the phase of the corresponding integrand is given by

$$\frac{d\phi(t)}{dt} = \delta_g E_p \exp(-2t^2/\tau^2) + \omega_0 + \alpha \exp(-2t^2/\tau^2) + (E_g - \epsilon), \quad (4.59)$$

where it can be seen that the role played before by δ_g is now played by $\delta_g + \alpha/E_p$, suggesting that by increasing the strength of the introduced chirp it is possible to satisfy eq. (4.55) (with δ_g replaced by $\delta_g + \alpha/E_p$) and as a result achieve dynamic interference. The exponential chirp can be approximated by a parabola,

$$\omega(t) = \omega_0 + \alpha \exp(-2t^2/\tau^2) \approx \omega_0 + \alpha - 2\alpha \frac{t^2}{\tau^2}. \quad (4.60)$$

The photo-electron spectrum from a parabolically chirped pulse with $\alpha = 0.3$ a.u. is shown in fig. 4.11. The intensity of the pulse is 5.8×10^{16} W/cm², the frequency is 53.6 eV, and the width is $\tau = 3$ fs. The chosen frequency and width match those used by Demekhin and Cederbaum [9, 10], and it is seen here that merely by introducing a chirp on the pulse, it is possible to recover dynamic interference. Notice that high intensities are by no means necessary in this case. At low intensities, however, the photo-electron spectrum is obtained simply by taking the Fourier transform of the pulse.

4.9.4. Ionisation from an excited state

So far we have always dealt with the hydrogen atom in its ground state. But a very interesting instance of dynamic interference happens if the atom is initially

excited. To be more clear, let's compare the Stark shift of the ground state of the hydrogen atom with that of the $2p$ -state. When the electron is ionised from the ground state, the frequency needed to excite the electron directly to the continuum is $\omega > 0.5$ a.u. At such frequencies the first fraction appearing in the expression for the Stark shift of the ground state, see eq. (4.24), can be safely neglected as all the non-essential states have higher energies than the ground state, as a consequence of which the Stark shift of the ground state is always positive.

This should be contrasted to the case of ionising the electron from an excited state, for example the $2p$ -state. The Stark shift of such a state is given by a similar expression to eq. (4.24) with index e (short for *excited*) replacing g ,

$$\delta_e^{(l)} = \omega^2 \text{p.v.} \sum_p \left(\frac{1}{\omega_{ep} - \omega} + \frac{1}{\omega_{ep} + \omega} \right) |\langle \phi_p | \mathbf{r} \cdot \hat{\mathbf{e}} | \phi_e \rangle|^2, \quad (4.61)$$

where index p runs over all the non-essential states. In this case the required frequency condition to ionise the electron directly into the continuum is $\omega > 0.125$ a.u. At such frequencies none of the two fractions appearing in the expression above can be generally neglected. The first fraction is dominant if 1. $|\phi_p\rangle = |\phi_g\rangle$, and 2. the frequency is close to the $1s$ - $2p$ resonance, otherwise the second fraction is dominant. Since the first fraction has a singularity at the $1s$ - $2p$ transition frequency, it is possible to arbitrarily increase the Stark shift by going near resonance, see fig. 4.12. Depending on what side of the resonance line the frequency lies, the Stark shift can be positive or negative. In the figure, a particular frequency is marked showing when the Stark shift of the $2p$ -state vanishes in the length gauge. Even though there is nothing fundamentally different about this particular frequency compared to the frequencies surrounding it, it makes the phenomenon of dynamic interference conceptually easier to grasp in the length gauge. If this frequency is used as the laser frequency, the initial state remains untouched in the length gauge throughout the pulse, whereas the continuum state experiences Stark shift, see fig. 4.13.

In fig. 4.14 the photo-electron spectrum resulting from such a frequency is shown for increasing laser intensities. The first thing to notice here is that since the electron is initially in a p -state, in the presence of a linearly polarised laser it can be ionised into both the s - and d - continua. The main peak is seen to become red shifted due to the fact that effective Stark shift is negative here. If the laser frequency was chosen below resonance, the main peak in the spectrum would move to higher energies for increasing laser intensities.

The numerical predictions of this subsection can be matched with those of a

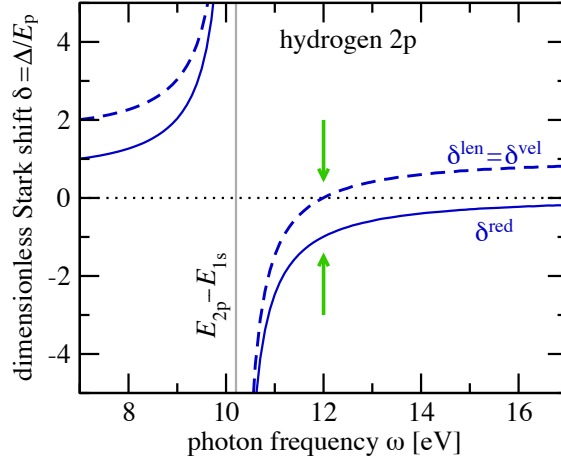


Figure 4.12. The dimensionless Stark shift for the $2p$ state of the hydrogen atom in the length, velocity, and reduced gauges. The Stark shift in the reduced gauge is $\delta^{\text{red}} = \delta^{\text{vel}} - 1$, and $\delta_{\text{vel}} = \delta_{\text{len}}$. The green arrows show the frequency at which $\delta^{\text{red}} = 0$; this frequency is used in fig. 4.14. [120]

minimal two-level model similar to what was introduced earlier in the chapter, except that the ground state should be replaced by the $2p$ -state, that is

$$\begin{aligned} a_e(t) &= \exp(-\gamma_e^{\text{len}} J(t)/2) \\ a_\epsilon(t) &= -i d_\epsilon \int_{t_0}^t dt' e^{-i(\delta_e^{\text{len}} - \delta_\epsilon^{\text{len}})J(t')} E(t') e^{-i(E_e - \epsilon)t'} a_e(t), \end{aligned} \quad (4.62)$$

where $d_\epsilon = \langle \phi_{p=\sqrt{2\epsilon}} | \mathbf{r} \cdot \hat{e} | \phi_e \rangle$. The prediction of this model is shown by the dash-dotted lines in fig. 4.7.

Similarly, dynamic interference can be observed if the electron is ionised from a Rydberg state directly into the continuum. For instance, if the electron is initially in the $25p$ state, the pulse shown in the upper plot of fig. 4.15 results in the spectrum shown in the lower plot. The peak intensity of the laser is $I = 3.5 \times 10^{14} \text{ W/cm}^2$, the frequency is $\omega = 2.72 \text{ eV}$, and the width is $\tau = 4.36 \text{ fs}$. An accurate description of the spectrum using a minimalist analytical model should in principle be possible, if the corresponding Stark shift and the depletion are known.

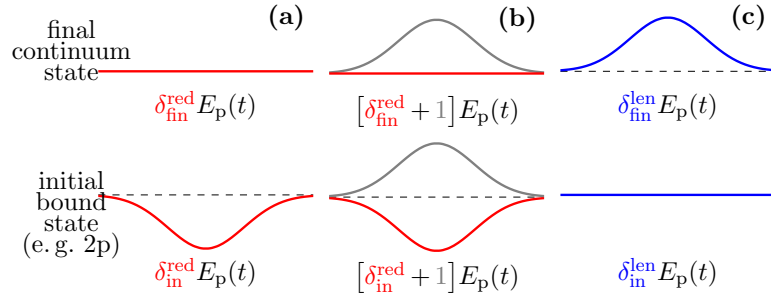


Figure 4.13. A similar sketch to fig. 4.3 except here the initial bound state is the $2p$ state, and the frequency is $\omega = 12$ eV. At this frequency the Stark shift of the initial state vanishes in the length gauge, see fig. 4.12. a) reduced gauge, b) velocity gauge, c) length gauge. [120]

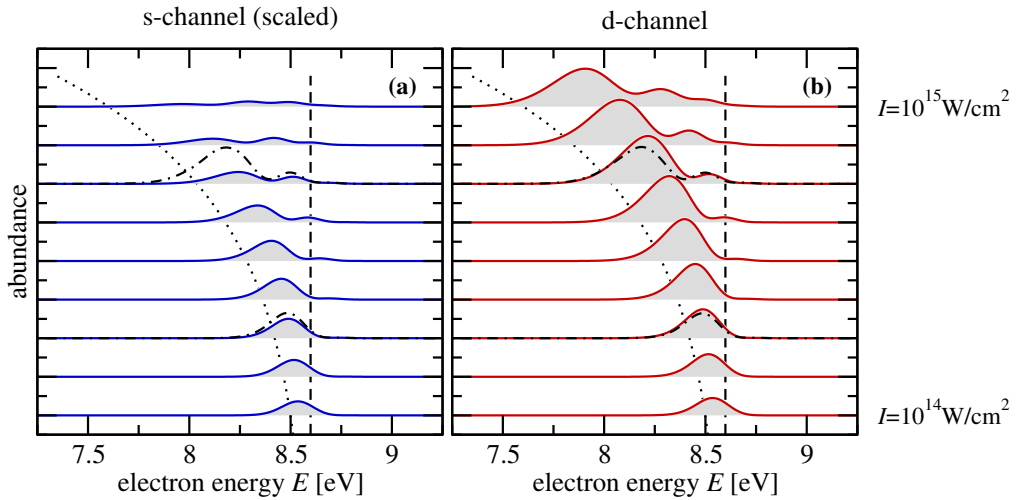


Figure 4.14. Photo-electron spectra for hydrogen in the $2p$ -state, after being irradiated with a 10 fs pulse of carrier frequency $\omega = 12$ eV. Ionisation into the a) s-channel, and b) d-channel is shown. The intensities from bottom to top are $I_k = 10^{k/8} \times 10^{14}$ W/cm² with $k = 0, 1, \dots, 8$. The dashed line shows the energy $E_\omega = E_{2p} + \omega = 8.59875$ eV. The dotted line is a fit to energies $E_{\omega,k} = E_{2p} + \omega - E_p(I_k)$. The s-channel is scaled by $f = d_{e,s}/d_{e,d} \approx 6.137$ so that the two channels have the same height in the perturbative regime (the lowest intensity shown here). The result of the analytic model (eq. 4.62) is shown for two intensities by the dot-dashed lines. [120]

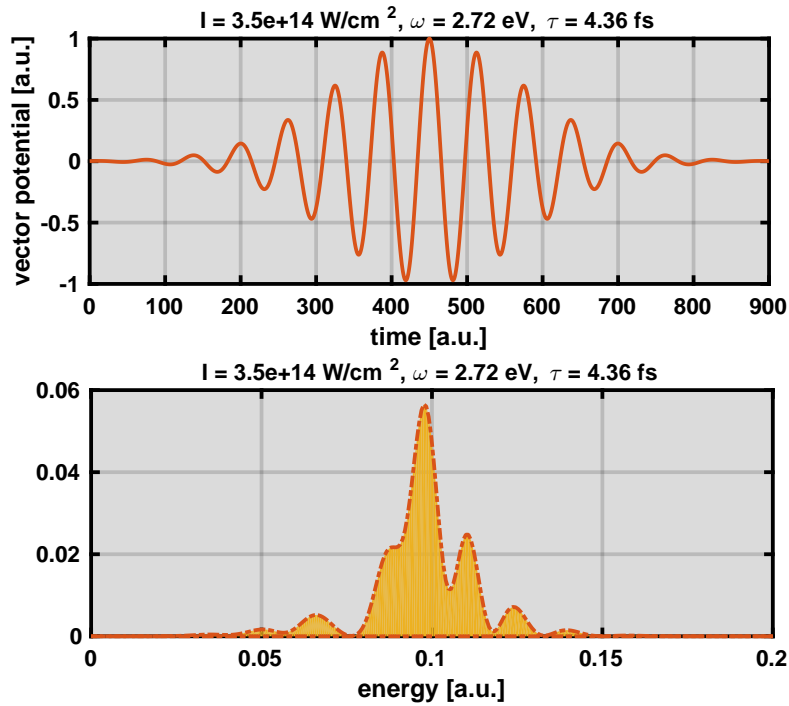


Figure 4.15. Upper plot: a Gaussian pulse of relatively low frequency used to excite the electron from a Rydberg state directly into the continuum. Rydberg states usually undergo stronger Stark shift compared to highly bound states, as a result of which a lower intensity suffices to observe dynamic interference. Lower plot: the corresponding photoelectron spectrum obtained by targeting a hydrogen atom in its $25p$ -state.

4.9.5. Resonance frequency

The last scenario that we will cover here is an attempt to make the last subsection experimentally more appealing (remember that hydrogen had to be initially in its $2p$ -state). There is no denying that it is experimentally easier to prepare the hydrogen atom in its ground state than in any excited state. Therefore, we modify a setup originally put forward by Demekhin and Cederbaum [126]. The electron is initially in the ground state, but the laser frequency is tuned near resonance with the $1s$ - $2p$ transition frequency, as a consequence of which the energy of a single photon is not enough to ionise the electron directly into the continuum.

In fig. 4.16 the d -channel of the photo-electron spectrum of a hydrogen atom

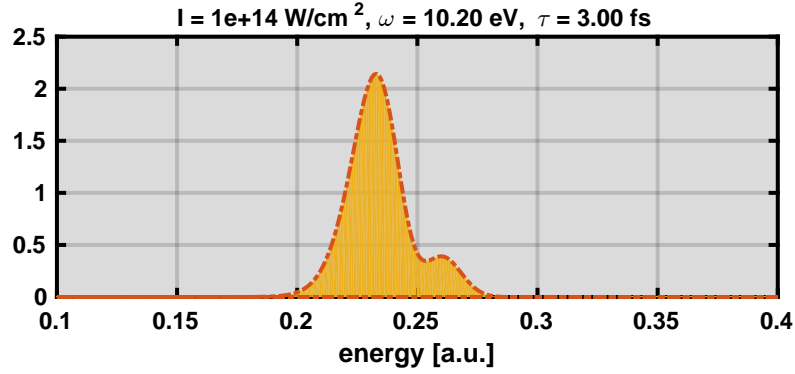


Figure 4.16. The d -channel of the photo-electron spectrum of a ground state hydrogen atom targeted by a Gaussian laser pulse tuned on resonance with the $1s$ - $2p$ transition frequency. The interference pattern is clearly visible.

targeted by a laser pulse of a Gaussian envelope $\mathcal{E}(t) = \mathcal{E}_0 \exp(-t^2/\tau^2)$ with $\tau = 3$ fs is shown. We will only consider the d -channel in what follows, but the analysis can be easily extended to the s -channel by inserting the proper parameter values. The intensity of the laser pulse is too low for the Stark shift to play any role. One point worth mentioning here is that contrary to the original paper [126], there is an obvious asymmetry in the spectrum about the energy $E_1 + 2\omega$ (this is the energy of the ground state plus twice the photon energy). The asymmetry will be soon described using a simple analytic model, which is an extension of the model used by the authors of the aforementioned publication (they modelled the hydrogen atom using a three-level system consisting of the $1s$, $2p$, and ϵd states). It will be seen that they observed the symmetric structure only because of their oversimplification of the problem.

In the analysis to follow, the electron is assumed to be initially in the ground state. Given that the laser is on resonance with the $1s$ - $2p$ transition frequency, and that the intensity is relatively low, we can safely assume the population stays mostly in the $1s$ - and $2p$ -states, which means a Rabi oscillation is expected between the two,

$$\begin{aligned} a_1(t) &= \cos\left(\frac{1}{2}d_{12}\zeta(t)\right) \\ a_2(t) &= -i \sin\left(\frac{1}{2}d_{12}\zeta(t)\right), \end{aligned} \tag{4.63}$$

where d_{12} is the dipole matrix element coupling the $1s$ - and $2p$ -states, and

$$\zeta(t) \equiv \int_{-\infty}^t \mathcal{E}(t') dt'. \quad (4.64)$$

The d -continuum states couple to the ground state through all the p -states (cf. eq. (4.21)). This is exactly what the authors of ref. [126] overlooked. If the ground state is only allowed to couple to the d -continuum through the $2p$ -state, a symmetric photo-electron spectrum follows. However, as soon as other p -states are included, this symmetry breaks down. For this very reason, we will include the p -continuum state at energy $E_1 + 2\omega + \delta$ in the model, where δ is a parameter determining the distance in energy from $E_1 + 2\omega$. Therefore, there are four levels, i.e. $1s$, $2p$, ϵp , and ϵd states.

If the laser pulse is given by $E(t) = \mathcal{E}(t) \sin(\omega t)$, after incorporating the rotating wave approximation, the interaction term in the interaction picture reads (a detailed derivation of this term can be found in appendix G)

$$-\frac{i}{2}\mathcal{E}(t) \begin{pmatrix} 0 & d_{12} & 0 & 0 \\ -d_{12} & 0 & 0 & 0 \\ 0 & -d_{23}e^{i\delta t} & 0 & d_{34}(e^{i\omega t} - e^{-i\omega t}) \\ -d_{14}e^{i\delta t}e^{i\omega t} & 0 & d_{34}(e^{i\omega t} - e^{-i\omega t}) & 0 \end{pmatrix}, \quad (4.65)$$

where d_{23} is the coupling between the $2p$ - and ϵd -states, d_{14} is the coupling between the $1s$ - and ϵp -states, and d_{34} is the coupling between the ϵd - and ϵp -states. The upper right corner of the matrix is filled with zeros because it is assumed most of the population is shared between the $1s$ - and $2p$ -states, and therefore any contribution to these two states coming from the continuum is negligible. It turns out the corresponding Schrödinger equation in the interaction picture can be analytically solved. The derivation, even though straightforward, is tedious, and therefore we only give the final solution here (for a detailed derivation please refer to appendix G). The final population of the ϵd -state is

$$\begin{aligned} a_3(\infty) &= \frac{1}{2}d_{14} \int_{-\infty}^{\infty} a_1(t')\mathcal{E}(t')e^{(\delta+\omega)t'} \sin(d_{34}\phi(t')) dt' \\ &\quad + \frac{1}{2}d_{23} \int_{-\infty}^{\infty} a_2(t')\mathcal{E}(t')e^{i\delta t'} \cos(d_{34}\phi(t')) dt' \end{aligned} \quad (4.66)$$

where $\phi(t) = \int_{-\infty}^t E(t') dt'$. Since $\lim_{t \rightarrow \infty} \mathcal{E}(t) = 0$ for a Gaussian envelope, we can

use the following approximation ⁶

$$\phi(t) \approx \frac{\mathcal{E}(t)}{\omega} \sin(\omega t). \quad (4.69)$$

Now using Jacobi-Anger expansion⁷ we have

$$\sin(d_{34}\phi(t)) \approx 2J_1\left(d_{34}\frac{\mathcal{E}(t)}{\omega}\right) \cos(\omega t) \quad (4.73)$$

$$\cos(d_{34}\phi(t)) \approx J_0\left(d_{34}\frac{\mathcal{E}(t)}{\omega}\right). \quad (4.74)$$

Upon inserting these expressions in eq. (4.66) and neglecting the term with $e^{2i\omega t}$ we get

$$\begin{aligned} a_3(\infty) \approx & \frac{1}{2}d_{14} \int_{-\infty}^{\infty} a_1(t') J_1\left(d_{34}\frac{\mathcal{E}(t')}{\omega}\right) \mathcal{E}(t') e^{i\delta t'} dt' \\ & + \frac{1}{2}d_{23}^* \int_{-\infty}^{\infty} a_2(t') J_0\left(d_{34}\frac{\mathcal{E}(t')}{\omega}\right) \mathcal{E}(t') e^{i\delta t'} dt', \end{aligned} \quad (4.75)$$

where $J_n(x)$ is the Bessel function of the first kind. Substituting for $a_1(t')$ and $a_2(t')$ gives

$$\begin{aligned} a_3(\infty) \approx & \frac{1}{4}d_{14} \int_{-\infty}^{\infty} \left(e^{\frac{i}{2}d_{12}\zeta(t')+i\delta t'} + e^{-\frac{i}{2}d_{12}\zeta(t')+i\delta t'} \right) J_1\left(d_{34}\frac{\mathcal{E}(t')}{\omega}\right) \mathcal{E}(t') dt' \\ & - \frac{i}{4}d_{23}^* \int_{-\infty}^{\infty} \left(e^{\frac{i}{2}d_{12}\zeta(t')+i\delta t'} - e^{-\frac{i}{2}d_{12}\zeta(t')+i\delta t'} \right) J_0\left(d_{34}\frac{\mathcal{E}(t')}{\omega}\right) \mathcal{E}(t') dt'. \end{aligned} \quad (4.76)$$

6. Let $B(t) = g(t) \sin(\omega t)$ where $g(t)$ is a slowly-varying function. The temporal derivative of $B(t)$ is then given by

$$\frac{dB(t)}{dt} = \dot{g}(t) \sin(\omega t) + \omega g(t) \cos(\omega t), \quad (4.67)$$

where the term containing $\dot{g}(t)$ is negligible, therefore

$$\frac{dB(t)}{dt} \approx \omega g(t) \cos(\omega t). \quad (4.68)$$

7. The Jacobi-Anger expansion is given by [119]

$$\exp(i\alpha \sin(x)) = \sum_{n=-\infty}^{\infty} J_n(\alpha) \exp(inx), \quad (4.70)$$

from which we can derive the following expansions

$$\cos(\alpha \cos(\omega t)) = J_0(\alpha) + 2 \sum_{n=2,4,\dots} J_n(\alpha) \cos\left(n\left(\omega t + \frac{\pi}{2}\right)\right) \quad (4.71)$$

$$\sin(\alpha \cos(\omega t)) = 2 \sum_{n=1,3,\dots} J_n(\alpha) \sin\left(n\left(\omega t + \frac{\pi}{2}\right)\right) \quad (4.72)$$

and for $\alpha \ll 1$ we may only keep the first terms.

Based on the stationary phase approximation, the value of the integrals can be evaluated using the stationary points of the phase of the integrand. These points are the solutions to the following two equations

$$\begin{aligned}\frac{1}{2}d_{12}\zeta(t) + \delta t &= 0 \\ -\frac{1}{2}d_{12}\zeta(t) + \delta t &= 0.\end{aligned}\tag{4.77}$$

If we assume that the solutions of the first and second equations are t_1 and t_2 respectively for $\delta > 0$, then their solutions swap for $\delta < 0$, i.e. t_2 will be the solution to the second equation and t_1 will be the solution to the first equation. Consequently, the value of the first integral remains untouched by $\delta \rightarrow -\delta$, whereas the sign of the second integral changes (note that there is a minus sign between the two terms inside the brackets in the second integral). Therefore, the final result depends on the sign of δ . In other words, $a_3(\infty)$ is asymmetric about $\delta = 0$.

Interestingly, if the p -continuum state is excluded from the model (i.e. $d_{14} = 0$), the asymmetry vanishes, as observed in ref. [126].

4.10. Conclusion

The temporal equivalent of the double-slit experiment was discussed at length in this chapter. And it was seen how this simple phenomenon is encountered again and again in different situations. Dynamic interference can be adequately described by a simple model which can be derived from first principles using the method of essential states followed by the perturbation theory. The method of essential states replaces the full Hamiltonian with an effective Hamiltonian defined in a limited subspace of the original Hilbert space.

The effect of the neglected states is introduced through an extra term in the effective Hamiltonian. This extra term gives rise to what is referred to as the dynamic Stark effect in the literature. There are different methods to evaluate the Stark shift of different states in a system. We evaluated the Stark shift of the ground state of hydrogen using a few lines of straightforward math, whereas for the continuum we simply read the Stark shift off the Volkov solutions.

It was further shown how different gauges result in different degrees of Stark shift for different states, despite the fact that the end result should be identical.

This observation makes it very crucial to choose a gauge at the beginning of any task and not change it until the task is done. The difference among the gauges is even more noticeable in the numerics, where only a limited subspace of the entire Hilbert space can be used due to limited computational resources. It turns out the velocity gauge (or its reduced version) is computationally the cheapest and most reliable for the regime we were concerned with in this chapter.

The necessary conditions for dynamic interference were derived following a few lines of reasoning, as a result of which it was seen not only a substantial Stark shift, but also a relatively small depletion rate is required for dynamic interference. These two conditions are simultaneously satisfied only in very special circumstances.

The chapter ended with a discussion of a handful of setups which could facilitate interference in the photo-electron spectrum in regimes much more feasible than the main scenario discussed in the chapter.

A. Atomic units

Length, mass, time, and electric current¹ are four independent units of measure. In the MKSA system (a subsystem of SI) the appropriate units are **m**eter, **s**econd, **k**ilogram, and **a**mpere respectively. In atomic physics, the quantities are often expressed in terms of another system of units which helps to simplify the numerics as well as the analytic expressions. This system is called “atomic units” whose main four units are [119]

- **Length** is defined by setting the Bohr constant to one, $a_0 = 1$
- **Mass** is defined by setting the mass of the electron to one, $m_e = 1$
- **Electric charge** is defined by setting the charge of the electron to one, $e = 1$
- **Time** is defined by setting the Planck constant to one, $\hbar = 1$

Other quantities can be easily expressed in terms of these four, see table A.1.

One of the most confusing conversion relations is that of between the laser intensity in SI units and the electric field in atomic units,

$$(E[\text{a.u.}])^2 = \frac{I[\text{W}/\text{cm}^2]}{3.5095 \times 10^{16}}. \quad (\text{A.1})$$

To derive this relation, the electric field should be set to one in atomic units or 5.142×10^{11} V/m in eq. (E.21),

$$I = \frac{1}{8\pi\alpha} \frac{e^2}{\hbar} (5.142 \times 10^{11})^2 = 3.5095 \times 10^{16} \text{W}/\text{cm}^2. \quad (\text{A.2})$$

When the electric field is equal to one atomic unit, the corresponding intensity in SI units is $I_0 = 3.5095 \times 10^{16}$ W/cm². This, however, does not mean that I_0 is equal to one atomic unit of intensity. One atomic unit of intensity is easily obtained via dimensional analysis 6.436×10^{15} W/cm². In practice it is customary to only work

1. The unit of electric charge is defined in terms of the units of time and electric current.

Table A.1. Some physical quantities in atomic units

Physical Quantity	Expression	Value (a.u.)	Value (SI)
electron charge	e	1	1.602×10^{-19} C
electron mass	m_e	1	9.109×10^{-31} kg
Bohr constant	a_0	1	5.29×10^{-11} m
reduced Planck constant	\hbar	1	1.055×10^{-34} J·s
Coulomb's constant	$k_e = 1/(4\pi\epsilon_0)$	1	8.988×10^9 N·m ² /C ²
energy	$E = m_e e^4 / (4\pi\epsilon_0 \hbar)^2$	1	4.360×10^{-18} J
time	$t = \hbar/E$	1	2.419×10^{-17} s
velocity	$v = a_0/t$	1	2.188×10^6 m/s
electric field	$\varepsilon = E/(ea_0)$	1	5.142×10^{11} V/m

with intensity in SI units and electric field in atomic units, and convert them to one another using the relation given above.

B. Finite difference method

B.1. Basics

Finite difference methods are numerical methods for solving a partial differential equation (PDE) on a grid. The derivatives are approximated by expressions relating neighbouring grid points. When expressed on a grid, the original PDE transforms into a system of linear equations which can be easily solved using textbook methods.

The partial differential equation

$$\nabla^2 f(\mathbf{r}) + \lambda \nabla f(\mathbf{r}) - \sigma f(\mathbf{r}) = 0 \quad (\text{B.1})$$

is a second order differential equation which is defined on a domain shown in fig. B.1. Similarly to ordinary differential equations, the value of the function and its derivative are needed on the boundary for a unique solution throughout the domain [127]. This type of boundary condition is called the Cauchy boundary condition and is briefly discussed in the subsequent section below. A lack of knowledge of the function value or its derivative results in a class of solutions instead of a unique solution, which might seem troublesome at first sight. However, a unique solution is not always what is needed. Should a unique answer be required, it can always be expressed in terms of all the solutions satisfying the boundary conditions. There are different ways that such partial boundary conditions can be applied to a problem, and these different ways usually carry different meanings. For example, fixing the value of the solution function on the boundaries is equivalent to fixing the electric potential on the boundary, whereas fixing the normal derivative on the boundary is equivalent to fixing the electric field. These boundary conditions are discussed in the rest of this appendix.

For the sake of simplicity only a one-dimensional version of eq. (B.1) is considered

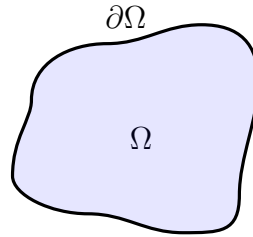
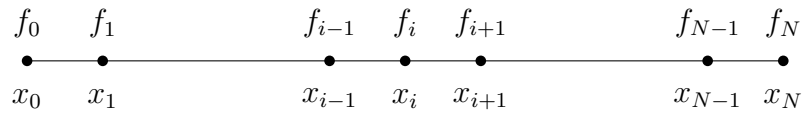
Figure B.1. domain Ω and its border $\partial\Omega$ 

Figure B.2. finite difference in one dimension

hereafter,

$$\frac{d^2 f(x)}{dx^2} + \lambda \frac{df(x)}{dx} - \sigma f(x) = 0, \quad (\text{B.2})$$

but all the results are easily generalisable to problems of higher dimensions. A one dimensional version of the domain is shown in fig. B.2.

To convert the partial differential equation into a system of linear equations, first we need to express the function on a grid,

$$f_i \equiv f(x_i), \quad (\text{B.3})$$

where $x_i = x_0 + i\Delta x$ are the grid points. The main idea behind any finite difference method is to approximate the derivatives appearing in the PDE with finite differences. The accuracy of these methods usually increases as the grid spacing shrinks. To approximate the first derivative, we write down the definition of the derivative,

$$f'(x) = \frac{f(x + \Delta x) - f(x - \Delta x)}{2\Delta x} + O((\Delta x)^2) \quad (\text{B.4})$$

and when expressed on a grid,

$$f'_i \equiv f'(x_i) \approx \frac{f_{i+1} - f_{i-1}}{2\Delta x}. \quad (\text{B.5})$$

The same has to be done with the second derivative. We write the definition,

$$f''(x) = \frac{f(x + \Delta x) - 2f(x) + f(x - \Delta x)}{(\Delta x)^2} + O((\Delta x)^2) \quad (\text{B.6})$$

and express it on the same grid,

$$f_i'' \equiv f''(x_i) \approx \frac{f_{i+1} - 2f_i + f_{i-1}}{(\Delta x)^2}. \quad (\text{B.7})$$

The careful reader may have noticed that we have used the second order central differences for the first and second derivatives. It is possible to use higher orders, or left and right differences [35]. But this choice depends on the problem at hand and the desired accuracy.

As will become apparent soon, it'll be convenient to write the function as a vector,

$$\mathbf{f} \equiv \begin{pmatrix} f_0 \\ f_1 \\ \vdots \\ f_N \end{pmatrix}, \quad (\text{B.8})$$

the size of which varies between $N - 1$ and $N + 1$ depending on the type of the boundary conditions applied. In a very similar fashion it is possible to define the first and second derivative vectors,

$$\mathbf{f}' \equiv \begin{pmatrix} f_0' \\ f_1' \\ \vdots \\ f_N' \end{pmatrix} = \frac{1}{2\Delta x} \begin{pmatrix} f_1 - f_{-1} \\ f_2 - f_0 \\ \vdots \\ f_{N+1} - f_{N-1} \end{pmatrix} \quad (\text{B.9})$$

and

$$\mathbf{f}'' \equiv \begin{pmatrix} f_0'' \\ f_1'' \\ \vdots \\ f_N'' \end{pmatrix} = \frac{1}{(\Delta x)^2} \begin{pmatrix} f_1 - 2f_0 + f_{-1} \\ f_2 - 2f_1 + f_0 \\ \ddots \\ f_{N+1} - 2f_N + f_{N-1} \end{pmatrix}. \quad (\text{B.10})$$

We are now in a position to rewrite the original PDE, eq. (B.2), as a system of linear equations,

$$\mathbf{f}'' + \lambda \mathbf{f}' - \sigma \mathbf{f} = 0, \quad (\text{B.11})$$

where we have replaced each derivative with its corresponding vector. After the boundary conditions are applied, this system of linear equations can be solved using textbook methods.

B.2. Boundary conditions

Cauchy boundary condition

Cauchy boundary condition specifies the values of the solution and its derivative on the boundary,

$$\begin{aligned} f(\mathbf{r}) &= f(\mathbf{r}) \quad \text{for } \mathbf{r} \in \partial\Omega \\ \left. \frac{\partial f}{\partial \mathbf{n}} \right|_{\mathbf{r}} &= g(\mathbf{r}) \quad \text{for } \mathbf{r} \in \partial\Omega, \end{aligned} \tag{B.12}$$

where \mathbf{n} is the normal to the boundary. This is the only boundary condition that guaranties a unique solution to the problem.

Dirichlet boundary condition

Dirichlet boundary condition (also known as the first-type boundary condition) is the first type of a category of boundary conditions which do not lead to a unique solution. This boundary condition specifies the value of the solution on the boundary,

$$f(\mathbf{r}) = g(\mathbf{r}) \quad \text{for } \mathbf{r} \in \partial\Omega, \tag{B.13}$$

where $g(\mathbf{r})$ is a predefined function. Such a boundary condition occurs when the electric potential between two different media is the quantity of interest.

In one dimension it takes a much simpler form,

$$\begin{aligned} f_0 &= \alpha \\ f_N &= \beta, \end{aligned} \tag{B.14}$$

in terms of which the function and its derivative vectors are expressed as

$$\mathbf{f} = \begin{pmatrix} f_1 \\ \vdots \\ f_{N-1} \end{pmatrix} \quad (\text{B.15})$$

$$\mathbf{f}' = \frac{1}{2\Delta x} \begin{pmatrix} 0 & 1 & & \\ -1 & 0 & 1 & \\ & \ddots & & \\ & & -1 & 0 & 1 \\ & & & -1 & 0 \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_{N-2} \\ f_{N-1} \end{pmatrix} + \frac{1}{2\Delta x} \begin{pmatrix} -\alpha \\ 0 \\ \vdots \\ 0 \\ \beta \end{pmatrix} \quad (\text{B.16})$$

$$\mathbf{f}'' = \frac{1}{(\Delta x)^2} \begin{pmatrix} -2 & 1 & & \\ 1 & -2 & 1 & \\ & \ddots & & \\ & & 1 & -2 & 1 \\ & & & 1 & -2 \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_{N-2} \\ f_{N-1} \end{pmatrix} + \frac{1}{(\Delta x)^2} \begin{pmatrix} \alpha \\ 0 \\ \vdots \\ 0 \\ \beta \end{pmatrix}. \quad (\text{B.17})$$

Note that these vectors are $N - 1$ dimensional. In general, fixing the value of the function on a boundary reduces the size of the vectors by one.

Neumann boundary condition

Neumann boundary condition (also known as the second-type boundary condition) specifies the derivative of the solution on the boundary,

$$\left. \frac{\partial f}{\partial \mathbf{n}} \right|_{\mathbf{r}} = g(\mathbf{r}) \quad \text{for } \mathbf{r} \in \partial\Omega, \quad (\text{B.18})$$

where \mathbf{n} is the normal vector to the boundary. This boundary condition is used when the electric field between two different media is the quantity of interest.

Regarding the derivatives on the boundaries of a one-dimensional domain, it is possible to use central differences at both ends, or use a left (right) difference at the right (left) end (left and right differences are called forward and backward differences if used in the time domain). It should be kept in mind that to maintain the level of accuracy throughout the domain, it is important to use one-sided (i.e. left or right) differences on the boundary which are of the same accuracy as the central difference used inside the domain. However, for more clarity we use the

simplest differences in each case regardless of the order. All the results are easily generalisable to an arbitrary order.

If the central differences are used at both ends,¹

$$\begin{aligned} -f'_0 &= \frac{f_{-1} - f_1}{2\Delta x} = \alpha \\ f'_N &= \frac{-f_{N-1} + f_{N+1}}{2\Delta x} = \beta. \end{aligned} \tag{B.19}$$

The derivative vectors will be given by

$$\begin{aligned} \mathbf{f}' &= \frac{1}{2\Delta x} \begin{pmatrix} 0 & 0 & & & & \\ -1 & 0 & 1 & & & \\ & & \ddots & & & \\ & & & -1 & 0 & 1 \\ & & & & 0 & 0 \end{pmatrix} \begin{pmatrix} f_0 \\ f_1 \\ \vdots \\ f_{N-1} \\ f_N \end{pmatrix} + \begin{pmatrix} -\alpha \\ 0 \\ \vdots \\ 0 \\ \beta \end{pmatrix} \\ \mathbf{f}'' &= \frac{1}{(\Delta x)^2} \begin{pmatrix} -2 & 2 & & & & \\ 1 & -2 & 1 & & & \\ & & \ddots & & & \\ & & & 1 & -2 & 1 \\ & & & & 2 & -2 \end{pmatrix} \begin{pmatrix} f_0 \\ f_1 \\ \vdots \\ f_{N-1} \\ f_N \end{pmatrix} + \frac{2}{\Delta x} \begin{pmatrix} \alpha \\ 0 \\ \vdots \\ 0 \\ \beta \end{pmatrix}. \end{aligned} \tag{B.20}$$

And if the backward difference is used for f'_0 , and the forward difference for f'_N , i.e.

$$\begin{aligned} -f'_0 &= \frac{f_{-1} - f_0}{\Delta x} = \alpha \\ f'_N &= \frac{-f_N + f_{N+1}}{\Delta x} = \beta, \end{aligned} \tag{B.21}$$

1. The reason why there is a minus sign on the left hand side of the first line as opposed to the second line, is that the normal vector always points outwards from the domain, i.e. $\mathbf{n}|_{\text{left}} = -\mathbf{n}|_{\text{right}}$.

the derivative vectors will be given by,

$$\mathbf{f}' = \frac{1}{2\Delta x} \begin{pmatrix} -1 & 1 & & & \\ -1 & 0 & 1 & & \\ & & \ddots & & \\ & & & -1 & 0 & 1 \\ & & & & -1 & 1 \end{pmatrix} \begin{pmatrix} f_0 \\ f_1 \\ \vdots \\ f_{N-1} \\ f_N \end{pmatrix} + \frac{1}{2} \begin{pmatrix} -\alpha \\ 0 \\ \vdots \\ 0 \\ \beta \end{pmatrix} \quad (\text{B.22})$$

$$\mathbf{f}'' = \frac{1}{(\Delta x)^2} \begin{pmatrix} -1 & 1 & & & \\ 1 & -2 & 1 & & \\ & & \ddots & & \\ & & & 1 & -2 & 1 \\ & & & & 1 & -1 \end{pmatrix} \begin{pmatrix} f_0 \\ f_1 \\ \vdots \\ f_{N-1} \\ f_N \end{pmatrix} + \frac{1}{\Delta x} \begin{pmatrix} \alpha \\ 0 \\ \vdots \\ 0 \\ \beta \end{pmatrix}.$$

Robin boundary condition

Robin boundary condition (or the third-type boundary condition) specifies a linear combination of the values of the solution and its derivative on the boundary of the domain,

$$A f(\mathbf{r}) + B \frac{\partial f}{\partial \mathbf{n}} \Big|_{\mathbf{r}} = g(\mathbf{r}) \quad \text{for } \mathbf{r} \in \partial\Omega, \quad (\text{B.23})$$

where \mathbf{n} is the normal vector to the boundary. In general the values of A and B can vary along the boundary. It is easily seen that in the limit of $A \rightarrow 0$ and $B \rightarrow 1$ this boundary condition becomes identical to the Neumann boundary condition, whereas in the limit of $A \rightarrow 1$ and $B \rightarrow 0$ it converges onto the Dirichlet boundary condition.

In one dimension and using central differences at both ends²

$$\begin{aligned} Af_0 - Bf'_0 &= Af_0 + B \frac{f_{-1} - f_1}{\Delta x} = \alpha \\ Af_N + Bf'_N &= Af_N + B \frac{-f_{N-1} + f_{N+1}}{\Delta x} = \beta \end{aligned} \quad (\text{B.24})$$

2. See the footnote in the Neumann boundary condition section to understand why there is a minus sign in the middle of the left hand side of the first line while there is a plus sign in the second line.

gives rise to the following derivative vectors,

$$\begin{aligned} \mathbf{f}' &= \frac{1}{2\Delta x} \begin{pmatrix} \frac{2A\Delta x}{B} & 0 & & & & \\ -1 & 0 & 1 & & & \\ & & \ddots & & & \\ & & & -1 & 0 & 1 \\ & & & & 0 & -\frac{2A\Delta x}{B} \end{pmatrix} \begin{pmatrix} f_0 \\ f_1 \\ \vdots \\ f_{N-1} \\ f_N \end{pmatrix} + \frac{1}{B} \begin{pmatrix} -\alpha \\ 0 \\ \vdots \\ 0 \\ \beta \end{pmatrix} \\ \mathbf{f}'' &= \frac{1}{(\Delta x)^2} \begin{pmatrix} -\frac{2A\Delta x}{B} & -2 & 2 & & & \\ 1 & -2 & 1 & & & \\ & & \ddots & & & \\ & & & 1 & -2 & 1 \\ & & & & 2 & -\frac{2A\Delta x}{B} - 2 \end{pmatrix} \begin{pmatrix} f_0 \\ f_1 \\ \vdots \\ f_{N-1} \\ f_N \end{pmatrix} + \frac{2}{B\Delta x} \begin{pmatrix} \alpha \\ 0 \\ \vdots \\ 0 \\ \beta \end{pmatrix}. \end{aligned} \quad (\text{B.25})$$

If the backward difference is used for f'_0 and the forward difference for f'_N ,

$$\begin{aligned} Af_0 - Bf'_0 &= Af_0 + B\frac{f_{-1} - f_0}{\Delta x} = \alpha \\ Af_N + Bf'_N &= Af_N + B\frac{-f_N + f_{N+1}}{\Delta x} = \beta, \end{aligned} \quad (\text{B.26})$$

the derivative vectors will be given by,

$$\begin{aligned} \mathbf{f}' &= \frac{1}{2\Delta x} \begin{pmatrix} \frac{A\Delta x}{B} - 1 & 1 & & & & \\ -1 & 0 & 1 & & & \\ & & \ddots & & & \\ & & & -1 & 0 & 1 \\ & & & & -1 & -\frac{A\Delta x}{B} + 1 \end{pmatrix} \begin{pmatrix} f_0 \\ f_1 \\ \vdots \\ f_{N-1} \\ f_N \end{pmatrix} + \frac{1}{2} \begin{pmatrix} -\alpha \\ 0 \\ \vdots \\ 0 \\ \beta \end{pmatrix} \\ \mathbf{f}'' &= \frac{1}{(\Delta x)^2} \begin{pmatrix} -\frac{A\Delta x}{B} - 1 & 1 & & & & \\ 1 & -2 & 1 & & & \\ & & \ddots & & & \\ & & & 1 & -2 & 1 \\ & & & & 1 & -\frac{A\Delta x}{B} - 1 \end{pmatrix} \begin{pmatrix} f_0 \\ f_1 \\ \vdots \\ f_{N-1} \\ f_N \end{pmatrix} + \frac{1}{B\Delta x} \begin{pmatrix} \alpha \\ 0 \\ \vdots \\ 0 \\ \beta \end{pmatrix}. \end{aligned} \quad (\text{B.27})$$

Mixed boundary condition

Mixed boundary condition is a mixture of the Dirichlet boundary condition and Neumann boundary condition applied to different parts of the boundary. This is

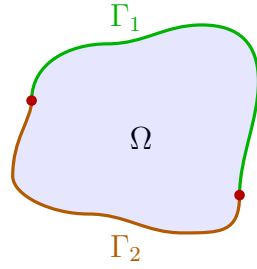


Figure B.3. Different types of boundary conditions applied to different parts of the boundary. This boundary condition is called the mixed boundary condition.

a special case of the Robin boundary condition with position dependent A and B coefficients.

If $\partial\Omega = \Gamma_1 \cup \Gamma_2$, see fig. B.3, the solution has to satisfy the following conditions

$$\begin{aligned} f(x_i, y_j) &= f(x_i, y_j) \quad \text{for } (x_i, y_j) \in \Gamma_1 \\ \frac{\partial f(x, y)}{\partial \mathbf{n}} \Big|_{(x_i, y_j)} &= g(x_i, y_j) \quad \text{for } (x_i, y_j) \in \Gamma_2, \end{aligned} \quad (\text{B.28})$$

where \mathbf{n} is the normal to the boundary.

Periodic boundary condition

This type of boundary condition is used when a large system is approximated by a small part of it, see fig. B.4. The periodic boundary condition for a rectangular boundary is

$$\begin{aligned} f(x, y) &= f(x \pm a, y) \\ f(x, y) &= f(x, y \pm b). \end{aligned} \quad (\text{B.29})$$

In a one dimensional space this boundary condition is equivalent to a ring with its ends tied together,

$$f_i = f_{i \pm N}. \quad (\text{B.30})$$

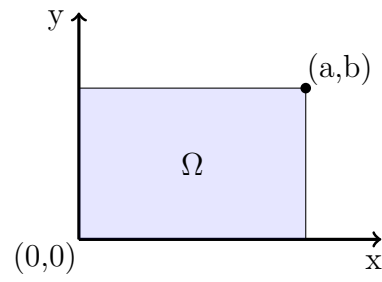


Figure B.4. A rectangular domain where the periodic boundary conditions can be applied.

The derivative vectors are given by

$$\mathbf{f}' = \frac{1}{2\Delta x} \begin{pmatrix} 0 & 1 & & -1 \\ -1 & 0 & 1 & \\ & \ddots & & \\ & & -1 & 0 & 1 \\ 1 & & & -1 & 0 \end{pmatrix} \begin{pmatrix} f_0 \\ f_1 \\ \vdots \\ f_{N-2} \\ f_{N-1} \end{pmatrix} \quad (\text{B.31})$$

$$\mathbf{f}'' = \frac{1}{(\Delta x)^2} \begin{pmatrix} -2 & 1 & & 1 \\ 1 & -2 & 1 & \\ & \ddots & & \\ & & 1 & -2 & 1 \\ 1 & & & 1 & -2 \end{pmatrix} \begin{pmatrix} f_0 \\ f_1 \\ \vdots \\ f_{N-2} \\ f_{N-1} \end{pmatrix}.$$

C. Further details regarding Gaussian processes

C.1. Derivatives of $m^*(\vec{x})$ and $k^*(\vec{x}, \vec{x})$

It is often necessary to find the minimum of a surface made of a linear combination of the posterior mean and covariance. In order to use textbook gradient based optimization algorithms, we have to calculate the derivatives of this surface with respect to \vec{x} (a $1 \times n$ matrix). This means the derivatives of the posterior mean and covariance are needed. The posterior mean and variance, respectively, are

$$\begin{aligned} m^*(\vec{x}) &= m(\vec{x}) + \vec{\Sigma}(\mathbf{X}, \vec{x})^\top \Sigma^{-1}(\vec{Y} - \vec{\mu}) \\ k^*(\vec{x}, \vec{x}) &= k(\vec{x}, \vec{x}) - \vec{\Sigma}(\mathbf{X}, \vec{x})^\top \Sigma^{-1} \vec{\Sigma}(\mathbf{X}, \vec{x}), \end{aligned} \tag{C.1}$$

where

$$\mathbf{X} = \begin{pmatrix} \vec{X}_1^\top \\ \vdots \\ \vec{X}_N^\top \end{pmatrix} \tag{C.2}$$

is a matrix whose rows are the different training points, and \vec{Y} is a vector of the corresponding target values. The covariance function is represented by $k(\vec{x}_1, \vec{x}_2)$ leading to the matrix Σ and vectors $\vec{\Sigma}$ defined as

$$\begin{aligned} \Sigma(\mathbf{X}, \mathbf{X}) &= k(\vec{X}_i, \vec{X}_j) \text{ matrix of size } N \times N \\ \vec{\Sigma}(\mathbf{X}, \vec{x}) &= k(\vec{X}_i, \vec{x}) \text{ vector of size } N \times 1, \end{aligned} \tag{C.3}$$

where \vec{x} is where the predictive probability is evaluated. Σ^{-1} is the inverse of $\Sigma(\mathbf{X}, \mathbf{X})$.

The derivative of the mean is

$$\begin{aligned}
\frac{dm^*(\vec{x})}{d\vec{x}} &= \frac{d}{d\vec{x}} \left\{ m(\vec{x}) + \vec{\Sigma}(\mathbf{X}, \vec{x})^\top \Sigma^{-1} (\vec{Y} - \vec{\mu}) \right\} \\
&= \frac{dm(\vec{x})}{d\vec{x}} + \left(\frac{d}{d\vec{x}} \vec{\Sigma}(\mathbf{X}, \vec{x})^\top \right) \Sigma^{-1} (\vec{Y} - \vec{\mu}) \\
&= \frac{dm(\vec{x})}{d\vec{x}} + \sum_i \left(\frac{d}{d\vec{x}} \Sigma(\vec{X}_i, \vec{x}) \right) \left(\Sigma^{-1} (\vec{Y} - \vec{\mu}) \right)_i,
\end{aligned} \tag{C.4}$$

while the derivative of the the covariance is

$$\begin{aligned}
\frac{dk^*(\vec{x}, \vec{x})}{d\vec{x}} &= \frac{dk(\vec{x}, \vec{x})}{d\vec{x}} - \frac{d}{d\vec{x}} \left\{ \vec{\Sigma}(\mathbf{X}, \vec{x})^\top \Sigma^{-1} \vec{\Sigma}(\mathbf{X}, \vec{x}) \right\} \\
&= \frac{dk(\vec{x}, \vec{x})}{d\vec{x}} - \frac{d}{d\vec{x}} \left(\vec{\Sigma}(\mathbf{X}, \vec{x})^\top \right) \Sigma^{-1} \vec{\Sigma}(\mathbf{X}, \vec{x}) - \vec{\Sigma}(\mathbf{X}, \vec{x})^\top \Sigma^{-1} \frac{d}{d\vec{x}} \left(\vec{\Sigma}(\mathbf{X}, \vec{x}) \right) \\
&= \frac{dk(\vec{x}, \vec{x})}{d\vec{x}} - \sum_{ij} \left(\frac{d}{d\vec{x}} \Sigma(\vec{X}_i, \vec{x}) \right) \Sigma_{ij}^{-1} \Sigma(\vec{X}_j, \vec{x}) - \sum_{ij} \Sigma(\vec{X}_i, \vec{x}) \Sigma_{ij}^{-1} \left(\frac{d}{d\vec{x}} \Sigma(\vec{X}_j, \vec{x}) \right).
\end{aligned} \tag{C.5}$$

C.2. Squared-exponential covariance function

The squared exponential covariance function is the most common covariance function used in the literature. It's functional simplicity has made it a very good choice.

$$k(\vec{x}_1, \vec{x}_2) = \sigma_f^2 \exp \left\{ -\frac{1}{2} (\vec{x}_1 - \vec{x}_2)^\top \Lambda^{-1} (\vec{x}_1 - \vec{x}_2) \right\}, \tag{C.6}$$

where Λ is a diagonal matrix with elements λ_i^2 representing correlation lengths in different dimensions.

For this choice of covariance function, it is conceivable to write down the derivatives analytically

$$\frac{d}{d\vec{x}} \Sigma(\vec{X}_i, \vec{x}) = \Sigma(\vec{X}_i, \vec{x}) (\vec{X}_i - \vec{x})^\top \Lambda^{-1}. \tag{C.7}$$

Substituting this relation in the formula for the derivatives of the mean from

previous subsection leads to

$$\begin{aligned}
\frac{dm^*(\vec{x})}{d\vec{x}} &= \frac{dm(\vec{x})}{d\vec{x}} + \sum_i \left(\Sigma^{-1}(\vec{Y} - \vec{\mu}) \right)_i \frac{d}{d\vec{x}} \Sigma(\vec{X}_i, \vec{x}) \\
&= \frac{dm(\vec{x})}{d\vec{x}} + \sum_i \left(\Sigma^{-1}(\vec{Y} - \vec{\mu}) \right)_i \Sigma(\vec{X}_i, \vec{x}) (\vec{X}_i - \vec{x})^\top \Lambda^{-1} \\
&= \frac{dm(\vec{x})}{d\vec{x}} + \left(\left(\Sigma^{-1}(\vec{Y} - \vec{\mu}) \right) \odot \vec{\Sigma}(\mathbf{X}, \vec{x}) \right)^\top \begin{pmatrix} \vec{X}_1^\top - \vec{x}^\top \\ \dots \\ \vec{X}_N^\top - \vec{x}^\top \end{pmatrix} \Lambda^{-1},
\end{aligned} \tag{C.8}$$

where \odot is the element wise product.

Similarly for the derivatives of the covariance

$$\begin{aligned}
\frac{dk^*(\vec{x}, \vec{x})}{d\vec{x}} &= - \sum_{ij} \left(\frac{d}{d\vec{x}} \Sigma(\vec{X}_i, \vec{x}) \right) \Sigma_{ij}^{-1} \Sigma(\vec{X}_j, \vec{x}) - \sum_{ij} \Sigma(\vec{X}_i, \vec{x}) \Sigma_{ij}^{-1} \left(\frac{d}{d\vec{x}} \Sigma(\vec{X}_j, \vec{x}) \right) \\
&= -2 \sum_{ij} \Sigma(\vec{X}_i, \vec{x}) \Sigma_{ij}^{-1} \Sigma(\vec{X}_j, \vec{x}) (\vec{X}_j - \vec{x})^\top \Lambda^{-1} \\
&= -2 \vec{\Sigma}(\mathbf{X}, \vec{x})^\top \Sigma^{-1} \begin{pmatrix} \Sigma(\vec{X}_1, \vec{x}) & & 0 \\ & \ddots & \\ 0 & & \Sigma(\vec{X}_N, \vec{x}) \end{pmatrix} \begin{pmatrix} \vec{X}_1^\top - \vec{x}^\top \\ \dots \\ \vec{X}_N^\top - \vec{x}^\top \end{pmatrix} \Lambda^{-1} \\
&= -2 \left(\left(\Sigma^{-1} \vec{\Sigma}(\mathbf{X}, \vec{x}) \right) \odot \vec{\Sigma}(\mathbf{X}, \vec{x}) \right)^\top \begin{pmatrix} \vec{X}_1^\top - \vec{x}^\top \\ \dots \\ \vec{X}_N^\top - \vec{x}^\top \end{pmatrix} \Lambda^{-1}.
\end{aligned} \tag{C.9}$$

C.3. Derivative of $\Lambda_1(\vec{x})$

We need the derivative of $\Lambda_1(\vec{x})$ in order to find the minimum using a gradient based method. Even though the derivation is not difficult, it is tedious, and here we only give the final expression

$$\begin{aligned}
\frac{d\Lambda_1(\vec{x})}{d\vec{x}} &= \frac{1}{2} \operatorname{erfc} \left(\frac{m^*(\vec{x}) - \eta}{\sqrt{2k^*(\vec{x}, \vec{x})}} \right) \frac{dm^*(\vec{x})}{d\vec{x}} \\
&\quad - \frac{1}{\sqrt{2\pi k^*(\vec{x}, \vec{x})}} \exp \left(-\frac{(m^*(\vec{x}) - \eta)^2}{2k^*(\vec{x}, \vec{x})} \right) \frac{1}{2} \frac{dk^*(\vec{x}, \vec{x})}{d\vec{x}},
\end{aligned} \tag{C.10}$$

where $\frac{dm^*(\vec{x})}{d\vec{x}}$ and $\frac{dk^*(\vec{x}, \vec{x})}{d\vec{x}}$ were derived earlier in the appendix.

D. Fourier transform

It is often desired to take the Fourier transform of a continuous function which has been sampled at discrete points. For example, the dipole moment of an atom at each time step given by a numerical simulation. Since all we have access to is a set of discrete values, sometimes it can be confusing to understand how the discrete Fourier transform of the discrete samples is connected to the continuous Fourier transform of the original quantity.

If $f(t)$ is a function of time, its Fourier transform is a function of frequency. One can be calculated given the other using the following expressions

$$\begin{aligned} F(\nu) &= \int_{-\infty}^{\infty} f(t)e^{-2\pi i\nu t} dt \\ f(t) &= \int_{-\infty}^{\infty} F(\nu)e^{2\pi i\nu t} d\nu, \end{aligned} \tag{D.1}$$

where ν is the frequency in hertz if t is time in seconds. In physics,¹ it is customary to use angular frequency² $\omega = 2\pi\nu$,

$$\begin{aligned} F(\omega) &= \int_{-\infty}^{\infty} f(t)e^{-i\omega t} dt \\ f(t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega)e^{i\omega t} d\omega. \end{aligned} \tag{D.2}$$

1. In quantum mechanics a slightly different version is used,

$$\begin{aligned} F(\omega) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t)e^{-i\omega t} dt \\ f(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} F(\omega)e^{i\omega t} d\omega, \end{aligned}$$

which is derived by making the substitution $F(\omega) \rightarrow \sqrt{2\pi}F(\omega)$ in eqs. (D.2). This version is unitary and expresses the symmetry between the Fourier transform and its inverse more clearly.

2. Even though ω is technically called the angular frequency, it is almost always referred to as frequency among physicists.

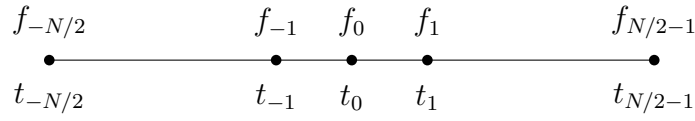


Figure D.1. A function sampled at different points in time. Even though the original function is continuous, we only have access to discrete measurements done at discrete times. It is crucial that we understand how the continuous Fourier transform of the original function can be evaluated from the discrete Fourier transform of the discrete samples.

It can be easily checked that upon plugging any of the various definitions of the Fourier transform in its corresponding inverse the original function is indeed reproduced.

D.1. Discrete Fourier transform

Assuming the sampling has been done at points Δt apart, see fig. D.1, the discrete Fourier transform (DFT) is defined as

$$\begin{aligned} F_m &= \sum_{n=-N/2}^{N/2-1} f_n e^{-2\pi i m n / N} \\ f_n &= \frac{1}{N} \sum_{m=-N/2}^{N/2-1} F_m e^{2\pi i m n / N}, \end{aligned} \quad (\text{D.3})$$

where $f_n = f(t_n)$ and $t_n = n\Delta t$. N represents the total number of samples.

D.2. Continuous limit of DFT

Given that there are N samples at points Δt apart, the length of the total interval is

$$T = N\Delta t, \quad (\text{D.4})$$

in terms of which the frequencies obtained from the DFT can be written as

$$\omega_m = m\Delta\omega \quad (\text{with } \Delta\omega \equiv \frac{2\pi}{T}). \quad (\text{D.5})$$

For reasons that will soon become apparent we make the substitution

$$F_m \equiv \mathbf{F}_m \Delta t. \quad (\text{D.6})$$

Using these new variables, we can rewrite the exponents in eqs. (D.3),

$$\pm \frac{2\pi i m n}{N} = \pm i \frac{2\pi m}{N \Delta t} n \Delta t = \pm i \omega_m t_n, \quad (\text{D.7})$$

and the prefactor,

$$\frac{1}{N} = \frac{2\pi \Delta t}{2\pi N \Delta t} = \frac{\Delta \omega \Delta t}{2\pi}, \quad (\text{D.8})$$

and consequently eqs. (D.3) change to

$$\begin{aligned} F_m &= \sum_{n=-N/2}^{N/2-1} f_n e^{-i\omega_m t_n} \Delta t \\ f_n &= \frac{1}{2\pi} \sum_{m=-N/2}^{N/2-1} F_m e^{i\omega_m t_n} \Delta \omega, \end{aligned} \quad (\text{D.9})$$

which are clearly reminiscent of eqs. (D.2). We are now in a position to study the two continuous limits of the DFT, i.e. $\Delta t \rightarrow 0$ and $\Delta \omega \rightarrow 0$.

D.2.1. Fourier series

In the limit of $\Delta t \rightarrow 0$ (while $N \rightarrow \infty$ such that $N \Delta t = T \ll \infty$) the summation in the inverse DFT turns into an integral,

$$\begin{aligned} F_m &= \sum_{n=-N/2}^{N/2-1} f_n e^{-i\omega_m t_n} \Delta t \quad \rightarrow \quad F_m = \int_{-T/2}^{T/2} f(t) e^{-i\omega_m t} dt \\ f_n &= \frac{1}{2\pi} \sum_{m=-N/2}^{N/2-1} F_m e^{i\omega_m t_n} \Delta \omega \quad \rightarrow \quad f(t) = \frac{1}{2\pi} \sum_{m=-N/2}^{N/2-1} F_m e^{i\omega_m t} \Delta \omega. \end{aligned} \quad (\text{D.10})$$

Upon defining $\mathcal{F}_m = \Delta \omega F_m / 2\pi = F_m / T$ we get

$$\begin{aligned} \mathcal{F}_m &= \frac{1}{T} \int_{-T/2}^{T/2} f(t) e^{-i\omega_m t} dt \\ f(t) &= \sum_{m=-N/2}^{N/2-1} \mathcal{F}_m e^{i\omega_m t}, \end{aligned} \quad (\text{D.11})$$

which is the well-known Fourier series.

D.2.2. Fourier transform

In the limit of $\Delta\omega \rightarrow 0$ (or equivalently $T \rightarrow \infty$) the only summation in eqs. (D.10) turns into an integral,

$$\begin{aligned} F_m &= \int_{-T/2}^{T/2} f(t)e^{-i\omega_m t} dt \quad \rightarrow F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t} dt \\ f(t) &= \frac{1}{2\pi} \sum_{m=-N/2}^{N/2-1} F_m e^{i\omega_m t} \Delta\omega \rightarrow f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{i\omega t} d\omega, \end{aligned} \quad (\text{D.12})$$

which is the Fourier transform as defined in eq. (D.2).

D.3. Parseval's theorem

Parseval's theorem states that the integral of the square of function $f(t)$ is equal to the integral of the square of its Fourier transform $F(\omega)$, i.e.

$$\int_{-\infty}^{\infty} |f(t)|^2 dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |F(\omega)|^2 d\omega, \quad (\text{D.13})$$

where the vertical bars indicate the absolute value. A similar relation can be derived for the discrete Fourier transform.

D.4. Convolution theorem

The convolution of function $f(t)$ with function $g(t)$ is denoted by $(f * g)(t)$ and is defined as

$$(f * g)(t) \equiv \int_{-\infty}^{\infty} f(\tau)g(t - \tau) d\tau. \quad (\text{D.14})$$

The convolution is in fact the reaction of a linear system with impulse response function $g(t)$ to input signal $f(t)$.

The convolution theorem states that the Fourier transform of the product of two functions is equal to the convolution of their Fourier transforms,

$$\mathcal{FT}\{f(t) \cdot g(t)\} = (\mathcal{FT}\{f\} * \mathcal{FT}\{g\})(t), \quad (\text{D.15})$$

where \mathcal{FT} is the Fourier transform, and \cdot denotes the elementary multiplication operation. It also works the other way around .

$$\mathcal{FT}\{f(t) * g(t)\} = \mathcal{FT}\{f(t)\} \cdot \mathcal{FT}\{g(t)\} \quad (\text{D.16})$$

Similar relations can be written for the discrete Fourier transform,

$$\begin{aligned} \mathcal{DFT}\{f_n \cdot g_n\} &= \mathcal{DFT}\{f_n\} * \mathcal{DFT}\{g_n\} \\ \mathcal{DFT}\{f_n * g_n\} &= \mathcal{DFT}\{f_n\} \cdot \mathcal{DFT}\{g_n\} \end{aligned} \quad (\text{D.17})$$

where \mathcal{DFT} denotes the discrete Fourier transform.

D.5. Windowed Fourier transform

Since computer simulations cannot run from $-\infty$ to ∞ , we always have to set a starting point $-\tau$ and an end point τ . This has the effect that the numerical output is in fact the product of the output we would get if the simulation could run from $-\infty$ to ∞ , and a rectangular window function with value 1 in the interval $[-\tau, \tau]$ and zero everywhere else,

$$w(t) = \begin{cases} 1 & \text{if } |t| < \tau \\ 0 & \text{else.} \end{cases} \quad (\text{D.18})$$

According to the convolution theorem the Fourier transform of the product of two functions is equal to the convolution of their Fourier transforms, eq. (D.15). The Fourier transform of the rectangular window function is the sinc function³

$$\mathcal{FT}\{w(t)\} = \frac{2 \sin(\omega\tau)}{\omega} = 2\tau \text{sinc}(\omega\tau) , \quad (\text{D.19})$$

which is a very jagged function. The convolution of the Fourier transform of almost any function with the sinc function results in a jagged function. This is the reason why the dipole matrix element between two continuum states normalised in a box shows an oscillatory behaviour when plotted as a function of one of the continuum state energies, whereas if the Dirac delta normalisation is used the behaviour is smooth.

3. The prefactor may be different if a different normalisation is used in the definition of the Fourier transform. This prefactor is correct for the normalisation defined in eq. (D.2)

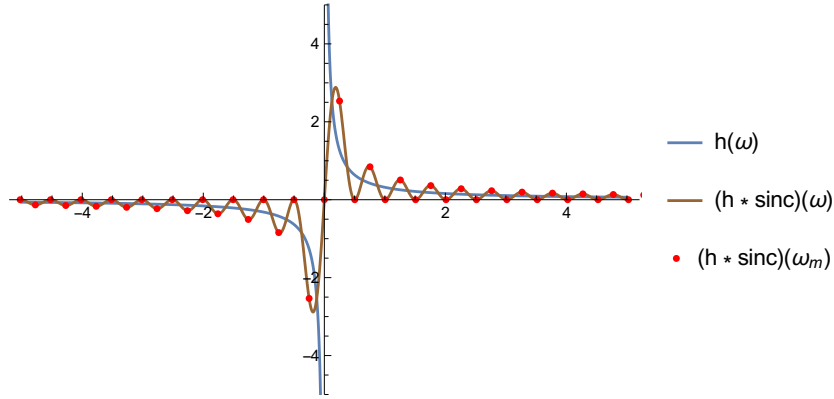


Figure D.2. $h(\omega)$ is the Fourier transform of $i \operatorname{sgn}(t)$; $(h * \operatorname{sinc})(\omega)$ is the Fourier transform of the same function limited to the interval $[-\tau, \tau]$; $(h * \operatorname{sinc})(\omega_m)$ is the discrete Fourier transform of $i \operatorname{sgn}(t_n)$.

As an example, consider the Fourier transform below

$$h(\omega) \equiv \mathcal{FT}\{i \operatorname{sgn}(t)\} = \int_{-\infty}^{\infty} i \operatorname{sgn}(t) e^{-i\omega t} dt = \frac{2}{\omega}, \quad (\text{D.20})$$

where $\operatorname{sgn}(\dots)$ is the sign function. If the time is limited to $[-\tau, \tau]$, then

$$\int_{-\tau}^{\tau} i \operatorname{sgn}(\tau) e^{-i\omega t} dt = \frac{2(1 - \cos(\omega\tau))}{\omega}, \quad (\text{D.21})$$

which is equal to $(h * \operatorname{sinc})(\omega)$. And if the function has been sampled at points Δt apart,

$$\begin{aligned} \mathcal{DFT}\{i \operatorname{sgn}(t_n)\} &= (h * \operatorname{sinc})(\omega_m) \\ &= \frac{2(1 - \cos(\omega_m \tau))}{\omega_m} = \begin{cases} 0 & \text{for } m \text{ even} \\ \frac{4}{\omega_m} & \text{for } m \text{ odd,} \end{cases} \end{aligned} \quad (\text{D.22})$$

where $\omega_m = m \Delta\omega$ with $\Delta\omega = 2\pi/2\tau$. The three different eqs. (D.20), (D.21), and (D.22) are compared in fig. D.2.

Now it should be apparent from the discussion above how crucial it is to use a well-justified window function in order to avoid troublesome convolutions. The window method commonly used among physicists is the Gabor transform which is the subject of the next section.

D.6. Gabor transform

One of the shortcomings of the Fourier transform is the lack of time resolution, meaning there is no knowledge of when a certain frequency is emitted and for how long. The Gabor transform tries to alleviate this problem by introducing a Gaussian window around the point of interest in time.⁴ The Gabor transform of function $f(t)$ is defined as

$$S(\omega, t) \equiv \int_{-\infty}^{\infty} w(t' - t, \sigma) f(t') e^{-i\omega t'} dt', \quad (\text{D.23})$$

where $S(\omega, t)$ is called the spectrogram, and $w(t, \sigma) = \exp(-t^2/2\sigma^2)$ is the window function with width σ . The smaller the width of the window, the finer the time resolution of the spectrogram. However, as the time resolution improves, the resolution along the frequency domain worsens and vice versa; this is the known Heisenberg uncertainty from quantum mechanics.

By taking the integral of the spectrogram along the time axis, it is possible to acquire information about a wider region than the original width of the window. In this case, the effective window function is derived by writing out the integral,

$$\int_{t-\frac{T}{2}}^{t+\frac{T}{2}} S(\omega, t') dt' = \int_{-\infty}^{\infty} \left(\int_{t-\frac{T}{2}}^{t+\frac{T}{2}} w(t'' - t', \sigma) dt' \right) f(t'') e^{-i\omega t''} dt''. \quad (\text{D.24})$$

The integral in parentheses can be analytically evaluated

$$\int_{t-\frac{T}{2}}^{t+\frac{T}{2}} w(t' - t'', \sigma) dt'' = \sqrt{\frac{\pi}{2}} \sigma \left(\operatorname{erf} \left(\frac{t' + \frac{T}{2} - t}{\sqrt{2}\sigma} \right) - \operatorname{erf} \left(\frac{t' - \frac{T}{2} - t}{\sqrt{2}\sigma} \right) \right). \quad (\text{D.25})$$

This effective window needs normalising as it takes on values greater than 1 for $t + T/2 < t' < t - T/2$. The easiest workaround is to include a prefactor equal to the maximum value of the effective window function (which is at $t' = t$) when integrating the spectrogram over time, i.e.

$$\mathcal{S}(\omega, t) \equiv \sqrt{2\pi} \sigma \operatorname{erf} \left(\frac{T}{2\sqrt{2}\sigma} \right) \int_{t-\frac{T}{2}}^{t+\frac{T}{2}} S(\omega, t') dt'. \quad (\text{D.26})$$

The original Gabor window along with the normalised effective window are shown in fig. D.3.

4. The Gabor transform is in fact a special case of the short-time Fourier transform.

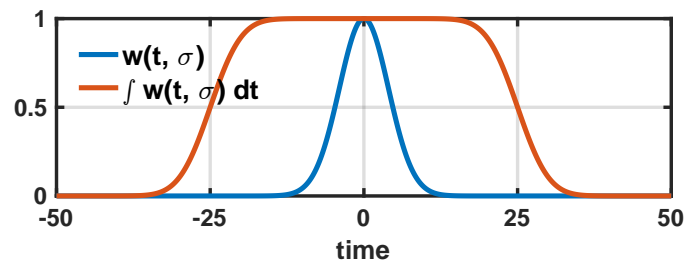


Figure D.3. The window function used in the Gabor transform is shown for $\sigma = 4$. Also drawn is the normalised effective window function resulting from integrating the Gabor window along the time axis from -25 to 25 .

E. Laser pulses: essentials

It is customary in the literature to split the vector potential of a laser pulse into two parts: the envelope, and the carrier wave

$$A(t) = \mathcal{A}(t) \cos(\omega t), \quad (\text{E.1})$$

where $\mathcal{A}(t)$ is the envelope, and $\cos(\omega t)$ is the carrier wave. The introduction of such an envelope is particularly helpful when the rate at which the envelope changes is small compared to the carrier frequency

$$\frac{d}{dt} \log \mathcal{A}(t) = \frac{1}{\mathcal{A}(t)} \frac{d}{dt} \mathcal{A}(t) \ll \omega. \quad (\text{E.2})$$

This regime is known as slowly-varying envelope regime.

The electric field is the time-derivative of the vector potential

$$E(t) = \frac{dA(t)}{dt} = \dot{\mathcal{A}}(t) \cos(\omega t) - \omega \mathcal{A}(t) \sin(\omega t), \quad (\text{E.3})$$

where, based on eq. (E.2), the first term can be neglected in the slowly-varying regime, yielding

$$E(t) \approx -\omega \mathcal{A}(t) \sin(\omega t), \quad (\text{E.4})$$

and if we define a similar envelope for the electric field as $E(t) = \mathcal{E}(t) \sin(\omega t)$, the envelope will be

$$\mathcal{E}(t) \approx -\omega \mathcal{A}(t). \quad (\text{E.5})$$

E.1. Realistic laser pulses

The laser pulse is often introduced by its vector potential rather than its electric field for a technical reason. To be more clear, imagine a charged particle in free space, and a laser pulse of width τ set to interact with the particle. It is known

from classical as well as quantum mechanics that a single electron cannot absorb or emit a photon in the absence of a second particle.¹ This means the velocity of the electron has to remain unchanged after the pulse has vanished

$$v(t \rightarrow +\infty) = v(t \rightarrow -\infty). \quad (\text{E.6})$$

The solution of Newton's equation of motion $\dot{v}(t) = E(t)$ for the electron ($e = 1$ since atomic units are used) is given by

$$v(t) = v(t = -\infty) + \int_{-\infty}^t \frac{dA(t)}{dt} dt. \quad (\text{E.7})$$

Applying the condition mentioned above (eq. (E.6)) to this solution gives rise to a new constraint on the vector potential

$$\int_{-\infty}^{+\infty} \frac{dA(t)}{dt} dt = 0. \quad (\text{E.8})$$

This condition is satisfied for any choice of vector potential as long as it vanishes at $t = +\infty$ and $t = -\infty$

$$\int_{-\infty}^{+\infty} \frac{dA(t)}{dt} dt = A(t = +\infty) - A(t = -\infty) = 0 - 0 = 0. \quad (\text{E.9})$$

The reader is urged to note that if instead of the vector potential the electric field had been used, the constraint on the electric field would not be as straightforward, and would be only approximately satisfied by any finite pulse in the slowly-varying envelope regime.

Using the vector potential has one more advantage. It can be directly used in the Schrödinger equation written in the velocity gauge. The velocity gauge is usually preferred over the length gauge. For more details please refer to the discussion at the end of section 4.7.

E.2. Energy density and intensity of laser pulses

It is known from classical electrodynamics that the density of the energy contained in an electromagnetic field in the vacuum is given by [128]

$$u(t) = \frac{1}{2} \left(\epsilon_0 E^2(t) + \frac{1}{\mu_0} B^2(t) \right), \quad (\text{E.10})$$

1. This is in fact why Bremsstrahlung and inverse Bremsstrahlung only happen in the presence of a nucleus.

where ϵ_0 is the electric permittivity of vacuum and μ_0 is the magnetic permeability of vacuum.

By taking advantage of the relation $E(t)/B(t) = c$ for an electromagnetic wave, and the expression for the speed of light in vacuum $c = 1/\sqrt{\epsilon_0\mu_0}$, it is easy to show that the two constituent parts of the energy density are equal in magnitude. This makes it possible to rewrite the energy density of an electromagnetic field solely in terms of the electric field,

$$\begin{aligned} u(t) &= \frac{1}{2} (\epsilon_0 E^2(t) + \mu_0^{-1} B^2(t)) \\ &= \frac{1}{2} \left(\epsilon_0 E^2(t) + \frac{1}{\mu_0 c} E^2(t) \right) \\ &= \epsilon_0 E^2(t). \end{aligned} \tag{E.11}$$

It is worth noting that if the dipole approximation² is used prior to this step, the formalism fails to reproduce the magnetic component of the wave (because $\nabla \times \hat{z} = 0$), wrongly implying that the magnetic component does not contribute to the energy of the wave. This discrepancy is a direct consequence of blindly generalising a local approximation to a global scale. The dipole approximation is an unrealistic assumption if applied to the entire space as it does not satisfy the wave equation.

The intensity of the electromagnetic wave is the amount of energy passing through unit area per unit time. In a plane perpendicular to the propagation direction it is the product of the energy density and the speed of light,

$$I(t) = cu(t) = c\epsilon_0 E^2(t). \tag{E.12}$$

E.3. Energy density and intensity of laser pulses in the slowly-varying envelope regime

In order to calculate the energy density (eq. (E.11)) and intensity (eq. (E.12)) of a laser pulse, we need to know the square of its electric field. By substituting the

2. which assumes the electric field has no spatial dependence, $E(x, t) = E(t)$

electric field as given in eq. (E.3), we get

$$u(t) = \epsilon_0 \left(\frac{dA(t)}{dt} \right)^2 \quad (\text{E.13})$$

$$I(t) = c\epsilon_0 \left(\frac{dA(t)}{dt} \right)^2. \quad (\text{E.14})$$

Since we are only interested in time scales much larger than the frequency of the carrier wave, it is easier to work with mean quantities which are the average of the above mentioned quantities over a few cycles. The mean energy density and intensity are

$$\langle u(t) \rangle = \epsilon_0 \left\langle \left(\frac{dA(t)}{dt} \right)^2 \right\rangle \quad (\text{E.15})$$

$$\langle I(t) \rangle = c\epsilon_0 \left\langle \left(\frac{dA(t)}{dt} \right)^2 \right\rangle. \quad (\text{E.16})$$

In the slowly-varying envelope regime the time derivative can be replaced with eq. (E.4), yielding

$$u(t) \approx \epsilon_0 \omega^2 \mathcal{A}^2(t) \sin^2(\omega t) \approx \epsilon_0 \mathcal{E}^2(t) \sin^2(\omega t) \quad (\text{E.17})$$

$$I(t) \approx c\epsilon_0 \omega^2 \mathcal{A}^2(t) \sin^2(\omega t) \approx c\epsilon_0 \mathcal{E}^2(t) \sin^2(\omega t), \quad (\text{E.18})$$

and when averaged over a few cycles ($\langle \sin^2(\omega t) \rangle = \frac{1}{2}$)

$$\langle u(t) \rangle \approx \frac{\epsilon_0}{2} \omega^2 \mathcal{A}^2(t) \approx \frac{\epsilon_0}{2} \mathcal{E}^2(t) \quad (\text{E.19})$$

$$\langle I(t) \rangle \approx \frac{c\epsilon_0}{2} \omega^2 \mathcal{A}^2(t) \approx \frac{c\epsilon_0}{2} \mathcal{E}^2(t). \quad (\text{E.20})$$

The intensity of the wave is sometimes expressed in terms of the fine-structure constant $\alpha = \frac{1}{4\pi\epsilon_0} \frac{e^2}{\hbar c}$

$$\langle I(t) \rangle \approx \frac{1}{8\pi\alpha} \frac{e^2}{\hbar} \mathcal{E}^2(t). \quad (\text{E.21})$$

The energy of a laser pulse is evaluated by integrating the intensity over time

$$U = \int_{-\infty}^{+\infty} I(t) dt. \quad (\text{E.22})$$

E.4. Spectral and temporal widths of pulses

In the discussion of laser pulses, two of the most important parameters are the temporal and spectral widths of the pulse. As to the temporal width, there are many definitions but the one we will use here is the full width at half maximum of the electric field of the laser pulse. And the spectral width that we will use refers to the full width at half maximum of the Fourier transform of the electric field. For example, a Gaussian pulse in the Fourier domain is given by

$$E(\omega) = \exp\left(-4 \ln 2 \frac{(\omega - \omega_0)^2}{\Omega^2}\right), \quad (\text{E.23})$$

where ω_0 is the carrier wave frequency, and Ω is the spectral width of the pulse (which is the FWHM of the Fourier transform of the pulse). The temporal profile of this pulse is given by its inverse Fourier transform

$$E(t) \propto \exp\left(-4 \ln 2 \frac{t^2}{T^2}\right), \quad (\text{E.24})$$

where T is the temporal width of the pulse (which is the FWHM of the electric field). The temporal and spectral widths are related by $T\Omega = 8 \ln 2$.

E.5. High harmonic generation: inhomogeneous wave equation

High harmonic generation (HHG) is a non-linear process in which a sample is illuminated by a laser pulse of frequency ω . The sample then radiates at frequencies equal to odd multiples of the laser frequency, that is $\omega, 3\omega, 5\omega, \dots$. This phenomenon has been explained semi-classically and quantum mechanically. For reviews of HHG please refer to refs. [96, 97].

In order to study and possibly control HHG, a precise calculation of the spectrum of the radiation field is needed. Now we will try and see how this can be done, and how to keep to a minimum the numerical artefacts that come about when using simulations of finite durations.

It is known that when an atom is in the presence of an oscillating electric field, a dipole moment is induced inside the atom giving rise to a non-zero dipole operator

expectation value $\langle \mathbf{r} \rangle = \langle \psi | \hat{\mathbf{r}} | \psi \rangle$, where \mathbf{r} is the position of the negative charge within the atom with respect to the nucleus. In order to calculate the spectrum of the radiation field, first it has to be determined if the spectrum is related to the Fourier transform of $\langle \mathbf{r} \rangle$, $\langle \dot{\mathbf{r}} \rangle$ or $\langle \ddot{\mathbf{r}} \rangle$. The derivation in this subsection closely follows that of ref. [129].

We start off with the inhomogeneous electromagnetic wave equation derived directly from Maxwell's equations [128]

$$\frac{1}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2} - \nabla^2 \mathbf{E} = - \left(\frac{1}{\epsilon_0} \nabla \rho + \mu_0 \frac{\partial \mathbf{J}}{\partial t} \right), \quad (\text{E.25})$$

where ρ is the source charge density (charge per unit volume), and \mathbf{J} is the source current density (current per unit area). If the only source present in the problem is a pointlike dipole, then the overall charge density is zero, i.e. $\rho = 0$, which is a direct consequence of the charge neutrality of a dipole. In the lack of any free-charge current density or magnetization current density, the source current density can be written in terms of the polarisation density

$$\mathbf{J} = \frac{\partial \mathbf{P}}{\partial t}, \quad (\text{E.26})$$

where \mathbf{P} is the polarisation density.

The polarisation density is given by $\mathbf{P} = \rho_b \mathbf{d}$, where \mathbf{d} is the displacement vector between the two bound charges measured from the negative charge to the positive charge. ρ_b is the bound charge density, and in the case of a pointlike source it is simply $q\delta(\mathbf{r} - \mathbf{r}_0)$ with q being the bound charge and \mathbf{r}_0 the position of the dipole. The source current density in the case of a pointlike dipole is therefore

$$\mathbf{J} = \frac{\partial \mathbf{P}}{\partial t} = \frac{\partial}{\partial t} q\delta(\mathbf{r} - \mathbf{r}_0)\mathbf{d} = q\delta(\mathbf{r} - \mathbf{r}_0) \frac{\partial \mathbf{d}}{\partial t}, \quad (\text{E.27})$$

where in the last step it is assumed the position of the dipole does not change over time. Plugging this in the original wave equation, we obtain

$$\frac{1}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2} - \nabla^2 \mathbf{E} = -\mu_0 q\delta(\mathbf{r} - \mathbf{r}_0) \frac{\partial^2 \mathbf{d}}{\partial t^2}, \quad (\text{E.28})$$

where it is clearly seen that the source of the radiation is $q \partial^2 \mathbf{d} / \partial t^2$. Rewriting in terms of the vectors introduced at the beginning, for the hydrogen atom (in atomic units) we get

$$q \frac{\partial^2 \mathbf{d}}{\partial t^2} = - \langle \ddot{\mathbf{r}} \rangle, \quad (\text{E.29})$$

where the minus sign originates from the fact that the displacement vector is measured from the negative charge to the positive charge while \mathbf{r} is measured from the positive charge to the negative charge.

The radiation spectrum is calculated by using $f(t) = -\langle \ddot{\mathbf{r}} \rangle$ in the Gabor transform introduced in appendix D.6 with $\sigma = 70$ a.u. The spectrogram is then integrated over the time window $[-T/2, T/2]$ where $T = 3.6\tau$ with τ being the FWHM of the pulse. The center of the pulse is always assumed to be at $t = 0$. This choice of time window makes sure the value of the integrated window function is negligible at the two ends of the interval (remember the simulations always run from $t = -2.5\tau$ to $t = 2.5\tau$ in this thesis) which will cause minimal ringing artefacts (for more on ringing artefacts please refer to ref. [130]). The spectrum is then given by

$$S(\omega) = \mathcal{S}(\omega, t = 0), \quad (\text{E.30})$$

where $\mathcal{S}(\omega, t)$ is defined in eq. (D.26).

F. Numerical propagation

The SCID-TDSE (short for Spherical-Coordinate Implicit Derivatives - Time Dependent Schrödinger Equation) package is repeatedly used in this thesis. For a complete description of the details of the package the reader is referred to the original paper accompanying the package [91]. Here we will only discuss the matter briefly, and report the parameter values used.

The wave function is expanded in terms of spherical harmonics,

$$\psi(\mathbf{r}) = \frac{1}{r} \sum_{l=0}^{l_{\max}} \sum_{m=-l}^l u_{lm}(r) Y_{lm}(\hat{r}), \quad (\text{F.1})$$

where \mathbf{r} is the position of the electron, r is its magnitude, and \hat{r} is a unit vector pointing in the direction of \mathbf{r} . l_{\max} represents the maximum angular momentum included in the simulation and has to be increased until convergence is reached. When the laser pulse is linearly polarised, only states of identical m couple. That is, if the electron is initially in the ground state (where $m = 0$), the wave function takes a much simpler form,

$$\psi(\mathbf{r}) = \frac{1}{r} \sum_{l=0}^{l_{\max}} u_l(r) Y_l(\hat{r}), \quad (\text{F.2})$$

where all the dependence on m has vanished. The radial components $u_{lm}(r)$ are expressed on a grid defined by $r = 0, dr, 2dr, \dots, R_{\max}$, where dr is the grid spacing, and R_{\max} is the maximum radius considered, which should be increased until no part of the wave function reaches the end of the simulation box over the course of the simulation. This wave function is then propagated in time using the Crank-Nicolson scheme with time step dt .

The field-free Hamiltonian $H = p^2/2 + V(r)$ expressed on this grid is diagonalised to find the eigenvalues (energies) and the eigenfunctions of the system. At the end of each simulation, the photo-electron spectrum can be calculated by projecting the wave function on the eigenfunctions.

The following parameters are used throughout this thesis unless specifically stated otherwise, $l_{\max} = 30$, $dr = 0.4$, $dt = 0.01$.

G. Four-level system

The four-level system presented in section 4.9.5 consists of one bound s -state, one bound p -state, one continuum p -state, and one continuum d -state. This system is set to interact with laser pulse $E = \mathcal{E}(t) \sin(\omega t)$ where $\mathcal{E}(t)$ is the envelope, and ω is the frequency of the pulse. The energies of the four levels are given by

$$\begin{aligned} E_1 &= 0 \\ E_2 &= \omega \\ E_3 &= \delta + 2\omega \\ E_4 &= \delta + 2\omega, \end{aligned} \tag{G.1}$$

where δ is the shift from energy 2ω . It is obvious that the first two states are coupled resonantly.

The Hamiltonian of such a system in its eigenbasis and in the absence of the laser pulse reads

$$H_0 = \begin{pmatrix} E_1 & 0 & 0 & 0 \\ 0 & E_2 & 0 & 0 \\ 0 & 0 & E_3 & 0 \\ 0 & 0 & 0 & E_4 \end{pmatrix}. \tag{G.2}$$

The interaction term resulting from the laser pulse in the same basis reads

$$V(t) = E(t) \begin{pmatrix} 0 & d_{12} & 0 & d_{14} \\ d_{12} & 0 & d_{23} & 0 \\ 0 & d_{23} & 0 & d_{34} \\ d_{14} & 0 & d_{34} & 0 \end{pmatrix}. \tag{G.3}$$

where d_{ij} is the dipole matrix element coupling state i and state j . The same

interaction term can be represented in the interaction picture

$$V_{\text{int}}(t) = -\frac{i}{2}\mathcal{E}(t) \times \begin{pmatrix} 0 & d_{12}(1 - e^{-2i\omega t}) & 0 & d_{14}e^{-i(\delta+\omega)t}(1 - e^{-2i\omega t}) \\ -d_{12}(1 - e^{2i\omega t}) & 0 & d_{23}e^{-i\delta t}(1 - e^{-2i\omega t}) & 0 \\ 0 & -d_{23}e^{i\delta t}(1 - e^{2i\omega t}) & 0 & d_{34}(e^{i\omega t} - e^{-i\omega t}) \\ -d_{14}e^{i(\delta+\omega)t}(1 - e^{2i\omega t}) & 0 & d_{34}(e^{i\omega t} - e^{-i\omega t}) & 0 \end{pmatrix}. \quad (\text{G.4})$$

Now we assume that the wave function remains mainly in the two bound states due to the resonant coupling between them. This assumption has the direct consequence that the population in the two continuum states is negligible, and therefore any transition from the continuum to the discrete states can be neglected. In other words, the upper right corner of the above matrix can be set to zero yielding

$$V_{\text{int}}(t) \approx -\frac{i}{2}\mathcal{E}(t) \times \begin{pmatrix} 0 & d_{12}(1 - e^{-2i\omega t}) & 0 & 0 \\ -d_{12}(1 - e^{2i\omega t}) & 0 & 0 & 0 \\ 0 & -d_{23}e^{i\delta t}(1 - e^{2i\omega t}) & 0 & d_{34}(e^{i\omega t} - e^{-i\omega t}) \\ -d_{14}e^{i(\delta+\omega)t}(1 - e^{2i\omega t}) & 0 & d_{34}(e^{i\omega t} - e^{-i\omega t}) & 0 \end{pmatrix}. \quad (\text{G.5})$$

Thanks to the rotating wave approximation, this term can be further simplified

$$V_{\text{int}}(t) \approx -\frac{i}{2}\mathcal{E}(t) \begin{pmatrix} 0 & d_{12} & 0 & 0 \\ -d_{12} & 0 & 0 & 0 \\ 0 & -d_{23}e^{i\delta t} & 0 & d_{34}(e^{i\omega t} - e^{-i\omega t}) \\ -d_{14}e^{i(\delta+\omega)t} & 0 & d_{34}(e^{i\omega t} - e^{-i\omega t}) & 0 \end{pmatrix}, \quad (\text{G.6})$$

which is identical to eq. (4.65).

The populations of the first two states, i.e. $a_1(t)$ and $a_2(t)$, can be easily found by solving the corresponding two-level system. The exact derivation can be looked up in almost any textbook on quantum mechanics [119]. The final expressions are

$$\begin{aligned} a_1(t) &= \cos\left(\frac{1}{2}d_{12}\zeta(t)\right) \\ a_2(t) &= -i \sin\left(\frac{1}{2}d_{12}\zeta(t)\right), \end{aligned} \quad (\text{G.7})$$

where $\zeta(t) \equiv \int_{-\infty}^t \mathcal{E}(t') dt'$.

In order to calculate the populations of the other two levels, let's first write the corresponding Schrödinger equation

$$i\partial_t \begin{pmatrix} a_3(t) \\ a_4(t) \end{pmatrix} = d_{34}E(t) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_3(t) \\ a_4(t) \end{pmatrix} + \frac{i}{2}\mathcal{E}(t)e^{i\delta t} \begin{pmatrix} d_{23}a_2(t) \\ e^{i\omega t}d_{14}a_1(t) \end{pmatrix}. \quad (\text{G.8})$$

These two coupled differential equations can be decoupled using transformation $T = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}$ yielding

$$\begin{aligned} i\partial_t \tilde{a}_3(t) &= -d_{34}E(t)\tilde{a}_3(t) + \frac{i}{2\sqrt{2}}e^{i\delta t}\mathcal{E}(t) (e^{i\omega t}d_{14}a_1(t) - d_{23}a_2(t)) \\ i\partial_t \tilde{a}_4(t) &= +d_{34}E(t)\tilde{a}_4(t) + \frac{i}{2\sqrt{2}}e^{i\delta t}\mathcal{E}(t) (e^{i\omega t}d_{14}a_1(t) + d_{23}a_2(t)), \end{aligned} \quad (\text{G.9})$$

where $\begin{pmatrix} \tilde{a}_3(t) \\ \tilde{a}_4(t) \end{pmatrix} \equiv T \begin{pmatrix} a_3(t) \\ a_4(t) \end{pmatrix}$. These two equations can be solved analytically

$$\begin{aligned} \tilde{a}_3(t) &= \frac{\sqrt{2}}{4} \int_{-\infty}^t \mathcal{E}(t')e^{i\delta t'} (e^{i\omega t'}d_{14}a_1(t') - d_{23}a_2(t')) e^{id_{34}(\phi(t)-\phi(t'))} dt' \\ \tilde{a}_4(t) &= \frac{\sqrt{2}}{4} \int_{-\infty}^t \mathcal{E}(t')e^{i\delta t'} (e^{i\omega t'}d_{14}a_1(t') + d_{23}a_2(t')) e^{-id_{34}(\phi(t)-\phi(t'))} dt', \end{aligned} \quad (\text{G.10})$$

where $\phi(t) = \int_{-\infty}^t E(t') dt'$ and it is assumed that the initial populations are zero, i.e. $\tilde{a}_3(0) = 0$ and $\tilde{a}_4(0) = 0$. The population of the third state can be found by transforming back to $a_3(t)$ yielding

$$\begin{aligned} a_3(t) &= \frac{1}{2}d_{14} \int_{-\infty}^t a_1(t')\mathcal{E}(t')e^{(\delta+\omega)t'} \sin(d_{34}(\phi(t') - \phi(t))) dt' \\ &\quad + \frac{1}{2}d_{23} \int_{-\infty}^t a_2(t')\mathcal{E}(t')e^{i\delta t'} \cos(d_{34}(\phi(t') - \phi(t))) dt'. \end{aligned} \quad (\text{G.11})$$

If the envelope of the pulse is Gaussian, then we have $\lim_{t \rightarrow \infty} \mathcal{E}(t) = 0$, and consequently $\lim_{t \rightarrow \infty} \phi(t) = 0$. Therefore, the final population of the third state reads

$$\begin{aligned} a_3(\infty) &= \frac{1}{2}d_{14} \int_{-\infty}^{\infty} a_1(t')\mathcal{E}(t')e^{(\delta+\omega)t'} \sin(d_{34}\phi(t')) dt' \\ &\quad + \frac{1}{2}d_{23} \int_{-\infty}^{\infty} a_2(t')\mathcal{E}(t')e^{i\delta t'} \cos(d_{34}\phi(t')) dt', \end{aligned} \quad (\text{G.12})$$

which is identical to eq. (4.66).

References

- [1] “Complex Systems”, Wikipedia, The Free Encyclopedia.
- [2] P. Di Cintio, U. Saalman, and J.-M. Rost, “Proton Ejection from Molecular Hydride Clusters Exposed to Strong x-Ray Pulses”, *Physical Review Letters* **111**, 123401 (2013).
- [3] P. P. Mattson, “Why Haven’t CPU Clock Speeds Increased in the Last Few Years?”, COMSOL Blog (2014).
- [4] G. H. Booth, A. J. W. Thom, and A. Alavi, “Fermion Monte Carlo without fixed nodes: a game of life, death, and annihilation in Slater determinant space”, *The Journal of Chemical Physics* **131**, 054106 (2009).
- [5] J. R. McClean and A. Aspuru-Guzik, “Clock quantum Monte Carlo technique : An imaginary-time method for real-time quantum dynamics”, *Physical Review A* **91**, 012311 (2015).
- [6] J. Werschnik and E. K. U. Gross, “Quantum optimal control theory”, *Journal of Physics B* **40**, R175–R211 (2007).
- [7] J. Werschnik and E. K. U. Gross, “Tailoring laser pulses with spectral and fluence constraints using optimal control theory”, *Journal of Optics B* **7**, S300–S312 (2005).
- [8] R. E. Goetz, A. Karamatskou, R. Santra, and C. P. Koch, “Quantum optimal control of photoelectron spectra and angular distributions”, *Physical Review A* **93**, 013413 (2016).
- [9] P. V. Demekhin and L. S. Cederbaum, “Dynamic Interference of Photoelectrons Produced by High-Frequency Laser Pulses”, *Physical Review Letters* **108**, 253001 (2012).
- [10] P. V. Demekhin and L. S. Cederbaum, “ac Stark effect in the electronic continuum and its impact on the photoionization of atoms by coherent intense short high-frequency laser pulses”, *Physical Review A* **88**, 043414 (2013).

-
- [11] P. V. Demekhin, D. Hochstuhl, and L. S. Cederbaum, “Photoionization of hydrogen atoms by coherent intense high-frequency short laser pulses: Direct propagation of electron wave packets on large spatial grids”, *Physical Review A* **88**, 023422 (2013).
- [12] C. Yu, N. Fu, G. Zhang, and J. Yao, “Dynamic Stark effect on XUV-laser-generated photoelectron spectra: Numerical experiment on atomic hydrogen”, *Physical Review A* **87**, 043405 (2013).
- [13] J. B. Anderson, “Quantum Monte Carlo: Atoms, Molecules, Clusters, Liquids, and Solids”, in *Reviews in Computational Chemistry*, Vol. 13, edited by K. B. Lipkowitz and D. B. Boyd (John Wiley & Sons, Inc., 1999), pp. 133–182.
- [14] W. M. C. Foulkes, L. Mitas, R. Needs, and G. Rajagopal, “Quantum Monte Carlo simulations of solids”, *Reviews of Modern Physics* **73**, 33–83 (2001).
- [15] M. Head-Gordon, “Quantum Chemistry and Molecular Processes”, *Journal of Physical Chemistry* **100**, 13213 (1996).
- [16] K. Raghavachari and J. B. Anderson, “Electron correlation effects in molecules”, *Journal of Physical Chemistry* **100**, 12960–12973 (1996).
- [17] G. H. Booth and A. Alavi, “Approaching chemical accuracy using full configuration-interaction quantum Monte Carlo: a study of ionization potentials”, *The Journal of Chemical Physics* **132**, 174104 (2010).
- [18] D. Cleland, G. H. Booth, and A. Alavi, “A study of electron affinities using the initiator approach to full configuration interaction quantum Monte Carlo”, *The Journal of Chemical Physics* **134**, 024112 (2011).
- [19] G. H. Booth, D. Cleland, A. J. W. Thom, and A. Alavi, “Breaking the carbon dimer: the challenges of multiple bond dissociation with full configuration interaction quantum Monte Carlo methods”, *The Journal of Chemical Physics* **135**, 084104 (2011).
- [20] G. H. Booth, A. Grüneis, G. Kresse, and A. Alavi, “Towards an exact description of electronic wavefunctions in real solids”, *Nature* **493**, 365–70 (2013).
- [21] J. S. Spencer, N. S. Blunt, and W. M. C. Foulkes, “The sign problem and population dynamics in the full configuration interaction quantum Monte Carlo method”, *The Journal of Chemical Physics* **136**, 054110 (2012).

-
- [22] M. H. Kolodrubetz, J. S. Spencer, B. K. Clark, and W. M. C. Foulkes, “The effect of quantization on the full configuration interaction quantum Monte Carlo sign problem”, *The Journal of Chemical Physics* **138**, 024110 (2013).
- [23] A. Goldberg and J. L. Schwartz, “Integration of the Schrödinger equation in imaginary time. I”, *Journal of Computational Physics* **1**, 433–447 (1967).
- [24] A. Goldberg and J. L. Schwartz, “Integration of the Schrödinger equation in imaginary time. II”, *Journal of Computational Physics* **1**, 448–453 (1967).
- [25] L. Mitas, “Diffusion Monte Carlo”, in *Quantum Monte Carlo methods in physics and chemistry*, edited by M. Nightingale and C. J. Umrigar (Springer, 1999), pp. 247–261.
- [26] T. Pang, “Diffusion Monte Carlo: A powerful tool for studying quantum many-body systems”, *American Journal of Physics* **82**, 980–988 (2014).
- [27] G. H. Booth, S. D. Smart, and A. Alavi, “Linear-scaling and parallelisable algorithms for stochastic quantum chemistry”, *Molecular Physics* **112**, 1855–1869 (2014).
- [28] N. S. Blunt, T. W. Rogers, J. S. Spencer, and W. M. C. Foulkes, “Density-matrix quantum Monte Carlo method”, *Physical Review B* **89**, 245124 (2014).
- [29] S. Ten-no, “Stochastic determination of effective Hamiltonian for the full configuration interaction solution of quasi-degenerate electronic states”, *The Journal of Chemical Physics* **138**, 164126 (2013).
- [30] A. D. McLachlan, “A variational solution of the time-dependent Schrodinger equation”, *Molecular Physics* **8**, 39–44 (1964).
- [31] A. Eckardt, “Colloquium: Atomic quantum gases in periodically driven optical lattices”, *Reviews of Modern Physics* **89**, 011004 (2017).
- [32] W. Pauli, *General Principles of Quantum Mechanics* (Springer, 1980).
- [33] J. Hilgevoord, “Time in quantum mechanics”, *American Journal of Physics* **70**, 301–306 (2002).
- [34] J. J. Sakurai, *Modern Quantum Mechanics* (Addison-Wesley, 1994).
- [35] R. J. LeVeque, *Finite Difference Methods for Ordinary and Partial Differential Equations* (SIAM, 2007).

- [36] J. R. McClean, J. A. Parkhill, and A. Aspuru-Guzik, “Feynman’s clock, a new variational principle, and parallel-in-time quantum dynamics”, in Proceedings of the National Academy of Sciences of the United States of America, Vol. 110, edited by F. A. Wilczek (2013), E3901–E3909.
- [37] “Big O notation”, Wikipedia, The Free Encyclopedia.
- [38] M. Campbell, A. Hoane, and F.-h. Hsu, “Deep Blue”, *Artificial Intelligence* **134**, 57–83 (2002).
- [39] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis, “Mastering the game of Go with deep neural networks and tree search”, *Nature* **529**, 484–489 (2016).
- [40] Z. Li, J. R. Kermode, and A. De Vita, “Molecular Dynamics with On-the-Fly Machine Learning of Quantum-Mechanical Forces”, *Physical Review Letters* **114**, 096405 (2015).
- [41] P. B. Wigley, P. J. Everitt, A. van den Hengel, J. W. Bastian, M. A. Sooriyabandara, G. D. McDonald, K. S. Hardman, C. D. Quinlivan, P. Manju, C. C. N. Kuhn, I. R. Petersen, A. N. Luiten, J. J. Hope, N. P. Robins, and M. R. Hush, “Fast machine-learning online optimization of ultra-cold-atom experiments”, *Scientific Reports* **6**, 25890 (2016).
- [42] M. Rupp, R. Ramakrishnan, and O. A. von Lilienfeld, “Machine Learning for Quantum Mechanical Properties of Atoms in Molecules”, *The Journal of Physical Chemistry Letters* **6**, 3309–3313 (2015).
- [43] R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. von Lilienfeld, “Big Data Meets Quantum Chemistry Approximations: The Δ -Machine Learning Approach”, *Journal of Chemical Theory and Computation* **11**, 2087–2096 (2015).
- [44] J. Cui and R. V. Krems, “Gaussian Process Model for Collision Dynamics of Complex Molecules”, *Physical Review Letters* **115**, 073202 (2015).
- [45] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, “Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons”, *Physical Review Letters* **104**, 136403 (2010).

-
- [46] A. P. Bartók and G. Csányi, “Gaussian approximation potentials: A brief tutorial introduction”, *International Journal of Quantum Chemistry* **115**, 1051–1057 (2015).
- [47] M. Rupp, “Machine learning for quantum mechanics in a nutshell”, *International Journal of Quantum Chemistry* **115**, 1058–1073 (2015).
- [48] A. Hentschel and B. C. Sanders, “Machine Learning for Precise Quantum Measurement”, *Physical Review Letters* **104**, 063603 (2010).
- [49] J. Bang, J. Ryu, S. Yoo, M. Pawłowski, and J. Lee, “A strategy for quantum algorithm design assisted by machine learning”, *New Journal of Physics* **16**, 073017 (2014).
- [50] E. Magesan, J. M. Gambetta, A. D. Córcoles, and J. M. Chow, “Machine Learning for Discriminating Quantum Measurement Trajectories and Improving Readout”, *Physical Review Letters* **114**, 200501 (2015).
- [51] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis* (CRC Press, 2014).
- [52] U. von Toussaint, “Bayesian inference in physics”, *Reviews of Modern Physics* **83**, 943–999 (2011).
- [53] Z. Ghahramani, “Bayesian non-parametrics and the probabilistic approach to modelling”, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **371**, 20110553 (2012).
- [54] Z. Ghahramani, “Probabilistic machine learning and artificial intelligence”, *Nature* **521**, 452–459 (2015).
- [55] Z. Ghahramani, “Unsupervised Learning”, in *Advanced Lectures on Machine Learning*, Vol. 3176, edited by O. Bousquet, U. von Luxburg, and G. Rätsch (Springer, 2004), pp. 72–112.
- [56] “Introduction to Deep Neural Networks”, *Deep Learning 4 Java (DL4J)*.
- [57] K. P. Murphy, *Machine Learning: A Probabilistic Perspective* (MIT Press, 2012).
- [58] N. Houlsby, F. Huszár, Z. Ghahramani, and M. Lengyel, “Bayesian Active Learning for Classification and Preference Learning”, arXiv:1112.5745 (2011).
- [59] “Newton’s method in optimization”, *Wikipedia, The Free Encyclopedia*.
- [60] “Nelder-Mead method”, *Wikipedia, The Free Encyclopedia*.

-
- [61] A. O’Hagan, “Monte Carlo is Fundamentally Unsound”, *The Statistician* **36**, 247 (1987).
- [62] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning* (MIT Press, 2006).
- [63] C. E. Rasmussen, “Gaussian Processes in Machine Learning”, in *Advanced Lectures on Machine Learning*, edited by O. Bousquet, U. von Luxburg, and G. Rätsch (Springer, 2004), pp. 63–71.
- [64] M. A. Osborne, S. J. Roberts, A. Rogers, S. D. Ramchurn, and N. R. Jennings, “Towards Real-Time Information Processing of Sensor Network Data Using Computationally Efficient Multi-output Gaussian Processes”, in *Proceedings of the 2008 International Conference on Information Processing in Sensor Networks (IPSN ’08)* (2008), pp. 109–120.
- [65] R. Garnett, M. A. Osborne, and S. J. Roberts, “Bayesian optimization for sensor set selection”, in *Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN ’10)* (2010), p. 209.
- [66] S. Roberts, M. Osborne, M. Ebdon, S. Reece, N. Gibson, and S. Aigrain, “Gaussian processes for time-series modelling”, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **371**, 20110550 (2012).
- [67] E. Keogh and A. Mueen, “Curse of Dimensionality”, in *Encyclopedia of Machine Learning*, edited by C. Sammut and G. I. Webb (Springer US, 2010), pp. 257–258.
- [68] A. B. Downey, *Think Bayes - Bayesian Statistics Made Simple* (Green Tea Press, 2012).
- [69] M. Zhu and A. Y. Lu, “The Counter-intuitive Non-informative Prior for the Bernoulli Family”, *Journal of Statistics Education* **12**, 1–10 (2004).
- [70] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms* (Cambridge University Press, 2003).
- [71] D. J. C. MacKay, “Comparison of Approximate Methods for Handling Hyperparameters”, *Neural Computation* **11**, 1035–1068 (1999).

-
- [72] M. Osborne, R. Garnett, Z. Ghahramani, D. K. Duvenaud, S. J. Roberts, and C. E. Rasmussen, “Active Learning of Model Evidence Using Bayesian Quadrature”, in *Advances in Neural Information Processing Systems 25* (NIPS ’12), edited by F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (2012), pp. 46–54.
- [73] R. Garnett, M. A. Osborne, and P. Hennig, “Active Learning of Linear Embeddings for Gaussian Processes”, arXiv:1310.6740 (2013).
- [74] T. Gunter, M. A. Osborne, R. Garnett, P. Hennig, and S. J. Roberts, “Sampling for Inference in Probabilistic Models with Fast Bayesian Quadrature”, arXiv:1411.0439 (2014).
- [75] G. C. Cawley and N. L. C. Talbot, “Preventing Over-Fitting during Model Selection via Bayesian Regularisation of the Hyper-Parameters”, *Journal of Machine Learning Research* **8**, 841–861 (2007).
- [76] G. C. Cawley and N. L. C. Talbot, “On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation”, *Journal of Machine Learning Research* **11**, 2079–2107 (2010).
- [77] S. Pressé, K. Ghosh, J. Lee, and K. A. Dill, “Principles of maximum entropy and maximum caliber in statistical physics”, *Reviews of Modern Physics* **85**, 1115–1141 (2013).
- [78] “Surrogate model”, Wikipedia, The Free Encyclopedia.
- [79] J. González, M. Osborne, and N. D. Lawrence, “GLASSES: Relieving The Myopia Of Bayesian Optimisation”, arXiv:1510.06299 (2015).
- [80] J. González, Z. Dai, P. Hennig, and N. D. Lawrence, “Batch Bayesian Optimization via Local Penalization”, in *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics* (PMLR 51), edited by A. Gretton and C. C. Robert (2016), pp. 648–657.
- [81] R. Bardenet and B. Kégl, “Surrogating the surrogate: accelerating Gaussian-process-based global optimization with a mixture cross-entropy algorithm”, in *27th International Conference on Machine Learning* (ICML), edited by T. J. J. Fürnkranz (2010), pp. 55–62.
- [82] A. Jalali, J. Azimi, X. Fern, and R. Zhang, “A Lipschitz Exploration-Exploitation Scheme for Bayesian Optimization”, in *Machine Learning and Knowledge Discovery in Databases*, edited by H. Blockeel, K. Kersting, S. Nijssen, and F. Železný (Springer, 2013), pp. 210–224.

-
- [83] J. Azimi, A. Jalali, and X. Fern, “Dynamic Batch Bayesian Optimization”, arXiv:1110.3347 (2011).
- [84] M. Frean and P. Boyle, “Using Gaussian Processes to Optimize Expensive Functions”, in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 5360, edited by W. Wobcke and M. Zhang (Springer, 2008), pp. 258–267.
- [85] E. Contal, V. Perchet, and N. Vayatis, “Gaussian Process Optimization with Mutual Information”, in Proceedings of the 31st International Conference on Machine Learning (PMLR 32), edited by E. P. Xing and T. Jebara (2013), pp. 253–261.
- [86] M. A. Osborne, R. Garnett, and S. J. Roberts, “Gaussian processes for global optimization”, in Proceedings of the 3rd International Conference on Learning and Intelligent Optimization (LION 3) (2009), pp. 1–15.
- [87] N. Srinivas, A. Krause, S. M. Kakade, and M. Seeger, “Gaussian Process Optimization in the Bandit Setting: No Regret and Experimental Design”, *IEEE Transactions on Information Theory* **58**, 3250–3265 (2009).
- [88] J. Snoek, H. Larochelle, and R. P. Adams, “Practical Bayesian Optimization of Machine Learning Algorithms”, in *Advances in Neural Information Processing Systems 25 (NIPS ’12)*, edited by F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (2012), pp. 2951–2959.
- [89] D. R. Jones, M. Schonlau, and W. J. Welch, “Efficient Global Optimization of Expensive Black-Box Functions”, *Journal of Global Optimization* **13**, 455–492 (1998).
- [90] T. Bäck, *Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms* (Oxford University Press, 1996).
- [91] S. Patchkovskii and H. Muller, “Simple, accurate, and efficient implementation of 1-electron atomic time-dependent Schrödinger equation in spherical coordinates”, *Computer Physics Communications* **199**, 153–169 (2016).
- [92] “Quantum harmonic oscillator”, Wikipedia, The Free Encyclopedia.
- [93] M. Gavrilá, “Atomic stabilization in superintense laser fields”, *Journal of Physics B* **35**, R147–R193 (2002).

-
- [94] B. W. J. McNeil and N. R. Thompson, “X-ray free-electron lasers”, *Nature Photonics* **4**, 814–821 (2010).
- [95] J. Feldhaus, J. Arthur, and J. B. Hastings, “X-ray free-electron lasers”, *Journal of Physics B: Atomic, Molecular and Optical Physics* **38**, S799 (2005).
- [96] K. L. Ishikawa, “High-Harmonic Generation”, in *Advances in Solid State Lasers Development and Applications*, edited by M. Grishin (InTech, 2010), pp. 439–464.
- [97] M. Kohler, T. Pfeifer, K. Hatsagortsyan, and C. Keitel, “Frontiers of Atomic High-Harmonic Generation”, in *Advances in Atomic, Molecular and Optical Physics*, Vol. 61, edited by P. Berman, E. Arimondo, and C. Lin (Academic Press, 2012), pp. 159–208.
- [98] A. R. Alberto Castro and E. K. Gross, “Enhancing and controlling single-atom high-harmonic generation spectra: a time-dependent density-functional scheme”, *The European Physical Journal B* **88**, 191 (2015).
- [99] I. Schaefer and R. Kosloff, “Optimal-control theory of harmonic generation”, *Physical Review A* **86**, 063417 (2012).
- [100] J. Solanpää, J. A. Budagosky, N. I. Shvetsov-Shilovski, A. Castro, A. Rubio, and E. Räsänen, “Optimal control of high-harmonic generation by intense few-cycle pulses”, *Physical Review A* **90**, 053402 (2014).
- [101] A. Fleischer and N. Moiseyev, “Amplification of high-order harmonics using weak perturbative high-frequency radiation”, *Physical Review A* **77**, 010102 (2008).
- [102] F. Krausz and M. Ivanov, “Attosecond physics”, *Reviews of Modern Physics* **81**, 163–234 (2009).
- [103] A. Fleischer, “Generation of higher-order harmonics upon the addition of high-frequency XUV radiation to IR radiation: Generalization of the three-step model”, *Physical Review A* **78**, 053413 (2008).
- [104] L. B. Madsen, “Strong-field approximation in laser-assisted dynamics”, *American Journal of Physics* **73**, 57–62 (2005).
- [105] C. Winterfeldt, C. Spielmann, and G. Gerber, “Colloquium: Optimal control of high-harmonic generation”, *Reviews of Modern Physics* **80**, 117–140 (2008).

-
- [106] D. R. Jones, C. D. Perttunen, and B. E. Stuckman, “Lipschitzian optimization without the Lipschitz constant”, *Journal of Optimization Theory and Applications* **79**, 157–181 (1993).
- [107] E. Brochu, V. M. Cora, and N. de Freitas, “A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning”, arXiv:1012.2599 (2010).
- [108] D. R. Jones, “A Taxonomy of Global Optimization Methods Based on Response Surfaces”, *Journal of Global Optimization* **21**, 345–383 (2001).
- [109] J. Mockus, “Application of Bayesian approach to numerical methods of global and stochastic optimization”, *Journal of Global Optimization* **4**, 347–365 (1994).
- [110] S. J. Glaser, U. Boscain, T. Calarco, C. P. Koch, W. Köckenberger, R. Kosloff, I. Kuprov, B. Luy, S. Schirmer, T. Schulte-Herbrüggen, D. Sugny, and F. K. Wilhelm, “Training Schrödinger’s cat: quantum optimal control”, *The European Physical Journal D* **69**, 279 (2015).
- [111] C. P. Koch, “Controlling open quantum systems: tools, achievements, and limitations”, *Journal of Physics: Condensed Matter* **28**, 213001 (2016).
- [112] D. Duvenaud, J. R. Lloyd, R. Grosse, J. B. Tenenbaum, and Z. Ghahramani, “Structure Discovery in Nonparametric Regression through Compositional Kernel Search”, in *Proceedings of the International Conference on Machine Learning (PMLR 28)*, Vol. 30, edited by S. Dasgupta and D. McAllester (2013), pp. 1166–1174.
- [113] D. B. Milošević, G. G. Paulus, D. Bauer, and W. Becker, “Above-threshold ionization by few-cycle pulses”, *Journal of Physics B* **39**, R203–R262 (2006).
- [114] F. H. M. Faisal, *Theory of Multiphoton Processes* (Springer, 1986).
- [115] B. W. Shore, *The Theory of Coherent Atomic Excitation*, Vol. 1,2 (John Wiley & Sons, Inc., 1990).
- [116] B. J. Sussman, “Five ways to the nonresonant dynamic Stark effect”, *American Journal of Physics* **79**, 477–484 (2011).
- [117] E. Brion, L. H. Pedersen, and K. Mølmer, “Adiabatic elimination in a lambda system”, *Journal of Physics A* **40**, 1033–1043 (2007).
- [118] L. B. Madsen, “Gauge invariance in the interaction between atoms and few-cycle laser pulses”, *Physical Review A* **65**, 053417 (2002).

-
- [119] F. Grossmann, *Theoretical Femtosecond Physics* (Springer, 2013).
- [120] M. Baghery, U. Saalmann, and J. M. Rost, “Essential Conditions for Dynamic Interference”, *Physical Review Letter* **118**, 143202 (2017).
- [121] D. Bauer and P. Koval, “Qprop: A Schrödinger-solver for intense laser–atom interaction”, *Computer Physics Communications* **174**, 396–421 (2006).
- [122] E. Cormier and P. Lambropoulos, “Optimal gauge and gauge invariance in non-perturbative time-dependent calculation of above-threshold ionization”, *Journal of Physics B* **29**, 1667–1680 (1996).
- [123] H. G. Muller, “An efficient propagation scheme for the time-dependent Schrödinger equation in the velocity gauge”, *Laser Physics* **9**, 138 (1999).
- [124] Y.-C. Han and L. B. Madsen, “Comparison between length and velocity gauges in quantum simulations of high-order harmonic generation”, *Physical Review A* **81**, 063430 (2010).
- [125] M. Gavrilă, “Atomic structure and decay in high-frequency fields”, in *Atoms in intense laser fields*, edited by M. Gavrilă (Academic Press, Boston, 1992), pp. 435–510.
- [126] P. V. Demekhin and L. S. Cederbaum, “Coherent intense resonant laser pulses lead to interference in the time domain seen in the spectrum of the emitted particles”, *Physical Review A* **86**, 063412 (2012).
- [127] W. E. Boyce and R. C. DiPrima, *Elementary Differential Equations and Boundary Value Problems, 10th edition* (John Wiley & Sons, Inc., 2012).
- [128] J. D. Jackson, *Classical Electrodynamics, 3rd edition* (John Wiley & Sons, Inc., 1998).
- [129] J. A. Pérez-Hernández and L. Plaja, “Comment on ‘On the dipole, velocity and acceleration forms in high-order harmonic generation from a single atom or molecule’”, *Journal of Physics B* **45**, 028001 (2012).
- [130] “Ringling artifacts”, Wikipedia, The Free Encyclopedia.