university of groningen / faculty of science and engineering

Master Project Mathematics and Physics

Quantum Optimal Control: Two-Level Systems and Parameter Identification

Roy Rodenburg
S4130952

October 2025

Abstract Quantum optimal control provides a powerful framework for precisely controlling the dynamics of quantum systems by designing control fields which drive the system to a desired target state. In this thesis, we will first introduce the basic theory of quantum optimal control, including key theorems on controllability and on the absence of local (non-global) optima for control pulses. We will then introduce quantum optimal control algorithms and provide demonstrations of how they work in practice. Finally, we will explore how quantum optimal control can be used to identify system parameters based on experimental data. First, benchmarking this method using fabricated data, before applying it to real-world experimental data. Although the parameter identification algorithm works well for benchmark models, it performed less well for experimental data, with experimental noise likely being the main factor for reduced performance.

supervised by
Prof. Dr. S. Trenn
Prof. Dr. T.L.C Janssen

Acknowledgements

I am very grateful for the supervision I received for my thesis. I greatly valued the continued support and feedback throughout the project. I particularly appreciated the privilege of having two first supervisors, which stemmed from the combined double-degree nature of the project. I really enjoyed the interesting discussions which arose due to this interdisciplinary project.

I would also like to think Prof. Dr. Richard Hildner for having multiple insightful meetings with me, which led to the work presented in Chapters 7 and 8.

I would also like to thank the Berendsen Center of the University of Groningen for their support and for providing access to the Nieuwpoort high performance computing cluster.

List of Symbols

$\mathbb{N}$	Set of non-negative integers
$\mathbb{N}^+$	Set of positive integers
$\mathbb{Z}$	Set of integers
$\mathbb{R}$	Set of real numbers
$\mathbb{R}^{\geq 0}$	Set of non-negative real numbers
$\mathbb{C}$	Set of complex numbers
$\mathcal{H}$	Hilbert space
$\mathcal{H}^*$	Dual space of the Hilbert space
$B(\mathcal{H})$	Space of bounded operators on the Hilbert space
$\|\psi\rangle$	Element of Hilbert space
$\langle\phi\|$	Element of Dual space of the Hilbert space
$\hat{\rho}$	Density matrix
$\hat{H}$	Hamiltonian
$\hat{\mathcal{L}}$	Liouvillian
$(\cdot,\cdot) : \mathcal{H} \times \mathcal{H} \to \mathbb{C}$	Inner product associated with the Hilbert space
$\|\cdot\| : \mathcal{H} \to \mathbb{R}$	Norm associated with the Hilbert space
$\cdot^* : \mathbb{C} \to \mathbb{C}$	Complex conjugate operation
$\cdot^\dagger : \mathbb{C}^{n\times m} \to \mathbb{C}^{m\times n}$	Complex conjugate transpose operator
$\hat{A} : D(\hat{A}) \to \mathcal{H}$	Linear operator
$\hat{A}^\dagger : D(\hat{A}^\dagger) \to \mathcal{H}$	Adjoint operator
$\hat{I} : \mathcal{H} \to \mathcal{H}$	Identity operator
$\hat{U}(\cdot,\cdot) : \mathbb{R} \times \mathbb{R} : (\mathcal{H} \to \mathcal{H})$	Time-evolution operator
$\|\|\cdot\|\| : B(\mathcal{H}) \to \mathbb{R}^{\geq 0}$	Operator norm
$[\cdot,\cdot]$	Commutator of operators
$\hat{\sigma}_{x,y,z}$	Pauli matrices
$\hbar$	Reduced Planck constant
ω	Laser frequency
ω_R	Rabi frequency
Δ	Detuning
$\vec{\mu}$	Electric dipole moment
δ_{ij}	Kronecker delta
$\Delta\mathcal{F}$	Infidelity

Contents

Chapter 1

Introduction

Quantum optimal control provides a powerful framework for precisely controlling the dynamics of quantum systems by designing control fields which drive the system to a desired target state. In this presentation, we will first introduce the basic theory of quantum optimal control, including key theorems on controllability and on the absence of local (non-global) optima for control pulses. We will then introduce quantum optimal control algorithms and provide demonstrations of how they work in practice. Finally, we will explore how quantum optimal control can be used to identify system parameters based on experimental data. First, benchmarking this method using fabricated data, before applying it to real-world experimental data.

In this thesis, we will focus on quantum optimal control, which aims to understand how to precisely control the dynamics of quantum systems using control fields which drive the system to a desired quantum state with high accuracy. It has many applications, from control single atoms and molecules [1], steering chemical reactions [2] to creating stable quantum computers [3], to quantum metrology [4], the science of precisely measuring properties of quantum systems.

Quantum optimal control provides a powerful theoretical and computational framework to solve such problems. It allows us to translate such control problems to an optimisation task for which we have a large suite of tools to attack.

In Chapter 2, we will provide an introduction to quantum mechanics for the more mathematically-minded reader. We explain quantum mechanics using functional analysis and even provide the mathematical reasoning behind the bra-ket notation. We will discuss the axioms of quantum mechanics before discussing two-level quantum systems in such a way as to provide analytical solutions to the systems we will discuss in later chapters. We will then finish off by discussing open quantum systems, which will allow us to describe decay phenomena like excited state decay rigorously in quantum mechanics.

In Chapter 3, we will discuss rigged Hilbert spaces, an extension of the typical way that quantum mechanics is taught, which will allow for a fully rigorous description of infinite-dimensional systems, where both eigenstates of operators and physically realisable states have their own spaces that are well-defined and cooperate nicely.

In Chapter 4, we will introduce optimal control theory, which is used to find the optimal method of controlling a system, whilst also shaping the controls to some desired form. We will then apply it to quantum mechanics and show the optimal control solution to a simple two-level problem. We will then showcase some theorems about quantum optimal control, for example, the controllability of quantum systems and the non-existence of locally optimal control solutions, which are not globally optimal. These two results combine to explain why quantum optimal control problems have seemingly so few issues finding a globally optimal solution. These results will then be used to explain the GRAPE and GOAT algorithms in later chapters.

In Chapter 5, we will present the GRAPE algorithm, which allows us to numerically solve quantum optimal control problems. Its core simplification is to represent control functions as piecewise constant functions. This allows us to optimise for a very wide class of control functions. We can also impose an additional cost depending on the shape of the control functions. We will analytically show the relationship between those additional costs and the effectiveness of the pulse in driving the system to the target. We will then apply the GRAPE algorithm to some two-level systems.

In Chapter 6, we will describe the GOAT algorithm, which is a more advanced version of the GRAPE algorithm and its main innovation is that it allows us to describe the control functions and smooth, parametrised functions. This allows us far more control over the final optimal control functions and drastically lowers the number of degrees of freedom over which we optimise. However, because of this lower amount of independent variables, the optimal control problem is not necessarily controllable anymore and non-global locally optimal solutions start to exist. This leads to issues with convergence towards the global optimal solution. Which leads to the introduction of global optimisation algorithms such as the Basin-Hopping algorithm.

In Chapter 7, we will elucidate upon double-pulse excitation experiments. Where two Gaussian pulses successively interact with a single molecule, with a fixed delay time between them. By varying this delay time and recording the final excited state population, we can infer the lifetimes of certain decay processes and determine unknown parameters of the quantum system. In this chapter, we will focus on simulations to showcase the effects of certain decay processes upon the system.

In Chapter 8, we will showcase the parameter-identification algorithm, which will perform fits on experimental data with two- and three-level systems, enabling us to identify key unknown parameters of the system. The chapter starts with a short overview and mathematical introduction, before performing the algorithm on multiple benchmarks to assess the effect of experimental noise, for example. Afterwards, we will apply parameter identification on two different data sets.

For more specifics on the simulations and the IPYNB notebooks used to create the plots in this thesis, check out this Github repository:
`https://github.com/RoyRodenburg/MasterProjectQuantumOptimalControl`.

This thesis is submitted for the completion of a double degree in physics and mathematics. Chapters 2 and 7 consist of purely physics content, whilst Chapters 4 consist of mainly mathematics

content. The remaining Chapters are decidedly more mixed, with Chapter 3 providing a mathematically rigorous introduction to quantum mechanics with non-compact operators and the last remaining chapters being focused on quantum optimal control algorithms.

ChatGPT and Mistral were used for assistance in literature review. Grammarly and ChatGPT were used to aid in improving the readability of the text. GitHub Copilot was used for the programming aspect of the thesis.

Chapter 2

Introduction to Quantum Mechanics for Mathematicians

We will now introduce the basics of quantum mechanics for the mathematically-minded reader to understand the rest of the text. This is, of course, not a comprehensive introduction, but references will be provided for supplementary material. The goal of quantum mechanics is to accurately describe the physics at the atomic scale, where classical mechanics ceases to be accurate. Particles are no longer localised at one point in space and start to gain wave-like properties.

Most quantum states, which we will discuss, are quantised: they are only allowed to have a discrete set of values for energy, momentum and spin. A quantum particle is typically in a combination of such states, meaning that its true energy is uncertain until measurement, which causes the wave function to collapse into one of these quantised states. This is totally contrary to classical systems, where the measurement of a particle's momentum is treated as a simple observation, which does not affect the system. Many aspects of quantum mechanics are unintuitive and call for a philosophical investigation; however, this chapter will focus on treating quantum mechanics with mathematical rigour, without discussing many famous experiments, like the double-slit experiment.

In the following section, we will briefly discuss the prerequisites of a mathematically rigorous introduction to quantum mechanics, explain the postulates upon which quantum mechanics is built and introduce the Schrödinger equation. Afterwards, we will discuss the two-level system and give analytic solutions using the rotating-wave approximation. Lastly, we will introduce open quantum systems and the Lindblad master equation, which, contrary to the Schrödinger equation, is able to include decay processes into the system dynamics.

A good basic introduction would be [5]. For a more mathematically rigorous introduction, [6] can be used.

2.1 Hilbert spaces

The fundamental object of quantum mechanics is the Hilbert space. Any quantum state of a system is an element of the Hilbert space associated with the system. Hilbert spaces will allow us to apply linear algebra and calculus to infinite-dimensional spaces. We will assume all of our Hilbert spaces to be complex vector spaces. Although real vector spaces can be Hilbert spaces, they have no application in quantum mechanics.

Definition 2.1.1 (Hilbert space)**.** *A complex vector space $\mathcal{H}$ equipped with an inner product $(\cdot,\cdot) : \mathcal{H}\times\mathcal{H}\to\mathbb{C}$ is a Hilbert space if for all $x, y, z \in \mathcal{H}$,*

1. *The inner product is conjugate symmetric in its arguments: $(x, y) = (y, x)^*$, where * denotes the complex conjugate.*
2. *It is linear in the 2nd argument: $(x, ay + bz) = a(x, y) + b(x, z)$ with $a, b \in \mathbb{C}$.*
3. *The inner product is positive definite: $(x, x) \geq 0$ with equality if and only if $x = 0$.*
4. *Equipped with the norm $|x| = \sqrt{(x, x)}$, the metric space $(\mathcal{H}, |\cdot|)$ should be complete, meaning that any Cauchy sequence converges[1] to a point in $\mathcal{H}$.*

A Hilbert space is separable if it has a dense countable[2] subset, which is equivalent to having a countable orthonormal basis.

Therefore, infinite-dimensional Hilbert spaces can be separable as long as they are countable infinite-dimensional [7], and all the Hilbert spaces which we will consider after this chapter will be separable.

In the mathematics literature, the inner product is typically assumed to be linear in its first argument. However, we will follow the physics convention and require the inner product to be linear in its second argument. Combining points 1 and 2, it follows that the inner product is conjugate linear in its first argument: $(ax + by, z) = a^*(x, z) + b^*(y, z)$ for all $a, b, x, y, z \in \mathbb{C}$.

Example 2.1.2. *For certain often-used Hilbert spaces, there is a standard choice of inner product, which we give here. These will come up often.*

- *If $\mathcal{H} = \mathbb{C}^n$, we can write the standard inner product as $(v, w) = v^\dagger w$ for $v, w \in \mathcal{H}$. Where the dagger operation $\cdot^\dagger : \mathbb{C}^{n\times m} \to \mathbb{C}^{m\times n}$ denotes the complex conjugate transpose. When we apply it to column vectors in $\mathbb{C}^n$, they get converted into row vectors in $\mathbb{C}^{1\times n}$. In the mathematics literature, the meaning of the $^*-$ and $^\dagger-$ operations is interchanged, but in this thesis, we will follow the physics convention.*

[1] A sequence $\{x_j\}_{j\in\mathbb{N}}$is Cauchy if for every $\epsilon > 0$, there exists a positive integer N such that $|x_n - x_m| < \epsilon$ for all integers $n, m > N$.

[2] A set is countable if it is finite or if it has a one-to-one correspondence to the set of positive integers $\mathbb{N}$, meaning that each element of the set can be assigned a unique integer.

- *If $\mathcal{H} = l^2(\mathbb{N})$, the space of square-summable infinite sequences $u = \{u_k\}_{k=0}^{\infty} \subset \mathbb{C}$ such that $\sum_{k=0}^{\infty} |u_k|^2 < \infty$, then the standard inner product is:*

$$(u, v) = \sum_{k=0}^{\infty} u_k^* v_k. \tag{2.1}$$

- *Let M be a manifold. If $\mathcal{H} = L^2(M)$, the space of square-integrable functions $f : M \to \mathbb{C}$, then the standard inner product is*

$$(\psi, \phi) = \int_M \psi^*(x)\phi(x)dx, \tag{2.2}$$

for $\psi, \phi \in L^2(M)$.

The first postulate of quantum mechanics relates to the state space associated to a quantum system.

Postulate 2.1.3. *The state of an isolated physical system at a fixed time t is represented by a state vector $|\psi\rangle$ belonging to a Hilbert space $\mathcal{H}$ called the state space.*

The state vector $|\psi\rangle \in \mathcal{H}$ is part of the Bra-ket notation in quantum mechanics, which we will now discuss. We will also generally assume that the elements of the Hilbert space are normalised: $(|\psi\rangle, |\psi\rangle) = 1$.

2.1.1 Bra-ket (Dirac) Notation

The Bra-ket or Dirac notation is one of the most commonly used ways to describe the state vectors used in quantum mechanics. The state vector $|\psi\rangle$ is an element of a Hilbert space $\mathcal{H}$. If the Hilbert space is separable, it can be represented by a column vector, which can be finite-dimensional or countably infinite-dimensional.

The bra $\langle\phi|$ is an element of the dual space $\mathcal{H}^*$. If the Hilbert space is separable, the bra can be viewed as the transpose of the ket and thus as a row vector. We can now define an isomorphism between the bra and ket states using the Riesz representation theorem.

Theorem 2.1.4 (Riesz representation theorem [8, Theorem 2.1.18])**.** *For all functionals $\langle\phi| \in \mathcal{H}^*$, there exists an unique vector $f_{\langle\phi|} \in \mathcal{H}$ such that*

$$\langle\phi|(|\psi\rangle) = (f_{\langle\phi|}, |\psi\rangle), \quad \forall |\psi\rangle \in \mathcal{H}. \tag{2.3}$$

Here we can identify $f_{\langle\phi|} := |\phi\rangle$. As a corollary, we can state:

Corollary 2.1.5. *The canonical map $\Phi : \mathcal{H} \to \mathcal{H}^*$ provides an isometric conjugate-linear[3] isomorphism between the Hilbert space and its dual space. It is defined by:*

[3]Meaning that $\Phi(a|\psi\rangle + b|\phi\rangle) = a^*\Phi(|\psi\rangle) + b^*\Phi(|\phi\rangle)$, for all $a, b \in \mathbb{C}$ and $|\psi\rangle, |\phi\rangle \in \mathcal{H}$.

$$\Phi : |\psi\rangle \mapsto (|\psi\rangle, \cdot) := \langle\psi| . \tag{2.4}$$

Its inverse $\Phi^{-1} : \mathcal{H}^* \to \mathcal{H}$ *is given using the above Riesz representation theorem to be:*

$$\Phi^{-1} : \langle\phi| \mapsto f_{\langle\phi|} := |\phi\rangle . \tag{2.5}$$

Combining the bra and the ket gives you a natural inner product. The main simplification in the bra-ket notation is the dropping of the parentheses in equation (2.3), i.e:

$$\langle\phi|\psi\rangle := \langle\phi| (|\psi\rangle) = (\Phi^{-1}(\langle\phi|), |\psi\rangle). \tag{2.6}$$

Elements of $\mathcal{H}$ which only differ by a global phase factor: $|\phi\rangle = e^{i\varphi} |\psi\rangle$ are considered equivalent since a global phase factor is not relevant to the dynamics of the system.

In parts of the physics literature, the bra is considered to be part of the Hilbert space. That rarely causes issues because of the isomorphism given in Corollary 2.1.5. However, for bra states, different transformation rules apply, which are more naturally explained if you view it as the dual of the ket.

We require that the state space of a quantum system is a Hilbert space, because we make heavy use of the inner product, and also because the dual of a separable Banach space may itself not be separable. This means there can't exist an isomorphism between the bras and kets, which would also cause many issues.

2.2 Linear Operators

Linear operators acting on Hilbert spaces are crucial in quantum mechanics because they represent measurements of physical observables, such as energy or position. We will now work towards the second postulate of quantum mechanics, which governs how measurements can be taken of a quantum system and how the system is affected by these measurements.

Definition 2.2.1 (Linear operator on Hilbert space)**.** *A linear operator on the Hilbert space* $\mathcal{H}$ *is a linear map* $\hat{A} : D(\hat{A}) \to \mathcal{H}$*, where* $D(\hat{A}) \subseteq \mathcal{H}$ *is the domain of* $\hat{A}$*. Such an operator can fulfil the following properties:*

- *An operator* $\hat{A}$ *is called bounded if there exists a constant* $M \in \mathbb{R}^+$ *such that* $|\hat{A} |\psi\rangle | \leq M | |\psi\rangle |$ *for all* $|\psi\rangle \in D(\hat{A})$*.*
- *A bounded linear operator* $\hat{A}$ *is called non-negative if* $(|\psi\rangle, \hat{A} |\psi\rangle) \geq 0$ *for all* $|\psi\rangle \in D(\hat{A})$*.*
- *If* $\hat{A} |\psi\rangle = \lambda |\psi\rangle$ *holds for some* $|\psi\rangle \in D(\hat{A})$ *and* $\lambda \in \mathbb{C}$*, then* λ *and* $|\psi\rangle$ *are referred to as an eigenvalue and eigenvector of the operator* $\hat{A}$*.*

All operators on the Hilbert space are equipped with a hat: $\hat{A} : D(\hat{A}) \to \mathcal{H}$, which is a convention in large parts of the physics literature.

Bounded linear operators are equivalent to continuous linear operators on Hilbert spaces. For infinite-dimensional operators, the notion of eigenvalues and eigenvectors can be extended to the notion of the spectrum of an operator, which is defined in Definition 2.2.7.

We will now also introduce the operator norm, as we will have to use it later.

Definition 2.2.2 (Operator norm)**.** *The operator norm of $\hat{A} : D(\hat{A}) \to \mathcal{H}$ is defined as*

$$||\hat{A}|| = \inf\{c \geq 0 : |\hat{A}\,|\psi\rangle\,| \leq c|\,|\psi\rangle\,|,\ \forall\,|\psi\rangle \in D(\hat{A})\}. \tag{2.7}$$

The space of bounded operators on $\mathcal{H}$ is denoted $B(\mathcal{H})$.

Bounded operators have the nice property that they can also be extended to the entire Hilbert space $\mathcal{H}$, using the Hahn-Banach theorem.

Theorem 2.2.3 (Hahn-Banach theorem, [9, theorem 6.13])**.** *Let $\hat{A}_0 : D(\hat{A}_0) \to \mathcal{H}$ be a bounded linear operator, with $D(A_0) \subset \mathcal{H}$ a linear subspace of $\mathcal{H}$, then there exists an extension $\hat{A} : \mathcal{H} \to \mathcal{H}$ such that $||\hat{A}_0|| = ||\hat{A}||$ and $\hat{A}|_{D(\hat{A}_0)} = \hat{A}_0$.*

From now on, we will assume that all bounded linear operators are defined on the entirety of $\mathcal{H}$, as otherwise, the operator could be extended using the Hahn-Banach theorem.

However, not every operator defined on the entire Hilbert space is bounded. Consider, for example, the Hilbert space $\mathcal{H} = l^2(\mathbb{N})$, with the operator $\hat{A} : |j\rangle \mapsto j\,|j\rangle$, for $j = 0, \cdots, \infty$. Where $|j\rangle$ are the unit basis vectors of the Hilbert space. This operator is clearly defined on the entire Hilbert space, but it is also not bounded because

$$||\hat{A}|| \geq \lim_{j\to\infty} |\hat{A}\,|j\rangle\,| = \lim_{j\to\infty} |j\,|j\rangle\,| = \infty. \tag{2.8}$$

Unless mentioned otherwise, every operator we discuss is assumed to be bounded, because the vast majority of operators which we will come across will be bounded. We will now define the adjoint of an operator.

Definition 2.2.4 (Adjoint of an operator)**.** *For a linear operator $\hat{A} : D(\hat{A}) \to \mathcal{H}$, its adjoint operator $\hat{A}^\dagger : D(\hat{A}^\dagger) \to \mathcal{H}$ is defined as the unique operator such that:*

$$(\hat{A}^\dagger u, v) = (u, \hat{A}v), \quad u \in D(\hat{A}^\dagger),\ v \in D(\hat{A}), \tag{2.9}$$

where the domain of the adjoint is:

$$D(\hat{A}^\dagger) = \{u \in \mathcal{H} : \exists\ w \in \mathcal{H} : (u, \hat{A}v) = (w, v),\ \forall v \in D(\hat{A})\}. \tag{2.10}$$

A linear operator $\hat{A}$ is called self-adjoint if $\hat{A} = \hat{A}^\dagger$ and $D(\hat{A}) = D(\hat{A}^\dagger)$.

Using the Riesz representation theorem 2.1.4, we can show that the adjoint of a bounded operator is defined on the entire Hilbert space and is also bounded. Assume that $\hat{A}$ is bounded and define the map $\Pi : \mathcal{H} \to \mathcal{H}^*$, which is defined as $\Pi : |\phi\rangle \mapsto \Pi_{|\phi\rangle}$ with:

$$\Pi_{|\phi\rangle} : |\psi\rangle \mapsto (|\phi\rangle, \hat{A}\,|\psi\rangle). \tag{2.11}$$

We can now apply the isomorphism $\Phi^{-1} : \mathcal{H}^* \to \mathcal{H}$ to $\Pi_{|\phi\rangle}$, which is provided in Corollary 2.1.5. So, $\Phi^{-1}(\Pi_{|\phi\rangle}) = f_{\Pi_{|\phi\rangle}}$ with $f_{\Pi_{|\phi\rangle}} \in \mathcal{H}$ such that

$$\Pi_{|\phi\rangle}(|\psi\rangle) = (f_{\Pi_{|\phi\rangle}}, |\psi\rangle), \quad \forall\ |\psi\rangle \in \mathcal{H}. \tag{2.12}$$

This we can use to write the following:

$$(|\phi\rangle, \hat{A}\,|\psi\rangle) = \Pi_{|\phi\rangle}(|\psi\rangle) = (f_{\Pi_{|\phi\rangle}}, |\psi\rangle). \tag{2.13}$$

Here we can identify the adjoint operator as $\hat{A}^\dagger\,|\phi\rangle := f_{\Pi_{|\phi\rangle}}$, for all $|\phi\rangle \in \mathcal{H}$. This shows that $\hat{A}^\dagger$ is defined on the entire Hilbert space. We can then also define $\Pi_{|\phi\rangle} := \langle\phi|\,\hat{A}^\dagger$, where the adjoint $\hat{A}^\dagger$ acts on the operator to the left of itself. Since the isomorphism Φ is isometric, we can deduce that $||\hat{A}^\dagger|| = ||\hat{A}||$, showing that the adjoint is also bounded. Self-adjoint operators are often called Hermitian in quantum mechanics, and the adjoint operator $\hat{A}^\dagger$ operation is often called the Hermitian conjugate of $\hat{A}$.

In the physics literature, you will often come across terms like $\langle\phi|\hat{A}|\psi\rangle$, where it is unclear whether $\hat{A}$ acts upon the bra or the ket. If $\hat{A}$ is Hermitian, as almost all operators which we will be dealing with are, then this notation is no longer ambiguous. For all $|\psi\rangle, |\phi\rangle \in \mathcal{H}$, we have:

$$(\hat{A}\,|\phi\rangle, |\psi\rangle) = (|\phi\rangle, \hat{A}\,|\psi\rangle). \tag{2.14}$$

Rewriting both sides in bra-ket notation, we get

$$(\langle\phi|\hat{A})|\psi\rangle = \langle\phi|(\hat{A}|\psi\rangle). \tag{2.15}$$

This shows that we are actually allowed to drop the brackets and write $\langle\phi|\hat{A}|\psi\rangle$. If an operator is not Hermitian, this notation is not well-defined and should thus be avoided. We will now briefly state a few important properties of adjoint operators without proof.

Lemma 2.2.5. *For linear bounded operators $\hat{A}, \hat{B} : \mathcal{H} \to \mathcal{H}$, the following properties hold:*

- $\left(\hat{A}^\dagger\right)^\dagger = \hat{A}$,
- $\left(\hat{A}\hat{B}\right)^\dagger = \hat{B}^\dagger\hat{A}^\dagger$,
- $\left(\lambda\hat{A} + \hat{B}\right)^\dagger = \lambda^*\hat{A}^\dagger + \hat{B}^\dagger$ *with* $\lambda \in \mathbb{C}$,
- $\left|\left|\hat{A}\right|\right| = \left|\left|\hat{A}^\dagger\right|\right|$.
- *If $\hat{A}$ is Hermitian, then $\langle\phi|\hat{A}|\phi\rangle$ is real, for any $|\phi\rangle \in \mathcal{H}$.*

Using these basics, we can move towards trying to express states and operators using an orthonormal basis. We will find that compact operators have the required properties to make this work nicely.

Definition 2.2.6 (Compact operator [8, Definition 2.3.1])**.** *An operator $\hat{A} : \mathcal{H} \to \mathcal{H}$ is said to be compact if there exists a neighborhood U around the origin in $\mathcal{H}$ such that the closure $\overline{\hat{A}(U)}$ is a compact subset of $\mathcal{H}$.*

On finite-dimensional Hilbert spaces, all linear operators are bounded. Consider a neighbourhood around the origin U, then $\overline{\hat{A}(U)}$ is bounded and closed and thus compact. Therefore, all linear operators on finite-dimensional spaces are compact.

This does not generalise to infinite-dimensional spaces, where you have to be more careful. By the spectral theorem 2.2.9, which will follow, it can be shown that the eigenvectors of compact operators generate a basis for the Hilbert space. But generic infinite-dimensional operators might not even have any eigenvectors. For infinite-dimensional Hilbert spaces and operators, the idea of eigenvalues and eigenvectors can be extended through the notion of the spectrum of an operator.

Definition 2.2.7 (Spectrum of infinite-dimensional operators)**.** *The spectrum of a potentially unbounded Hermitian operator $\hat{A} : D(\hat{A}) \to \mathcal{H}$ is the set $\sigma(\hat{A})$, where $\lambda \in \mathbb{C}$ is part of $\sigma(\hat{A})$ if $\hat{A} - \lambda\hat{I}$ is not invertible, i.e. doesn't have a bounded inverse. Here $\hat{I} : \mathcal{H} \to \mathcal{H}$ denotes the identity operator.*

If $\lambda \in \mathbb{C}$ is an eigenvalue of $\hat{A}$, it is included in its spectrum, but the converse is not necessarily true. Meaning that there doesn't necessarily have to be an eigenvector $|\psi\rangle \in D(\hat{A})$ which satisfies

$$\hat{A}\,|\psi\rangle = \lambda\,|\psi\rangle\,. \tag{2.16}$$

We have the following lemma for Hermitian operators.

Lemma 2.2.8 ([6, theorem 1, 2–3])**.** *If the operator $\hat{A}$ is Hermitian, all its eigenvalues are real. The eigenvectors belonging to distinct eigenvalues are orthogonal.*

Eigenvalues can still be degenerate, meaning it has multiple associated linearly independent eigenvectors. We can now state the spectral theorem, which is one of the pillars of operator theory and guarantees that we have an orthonormal basis to work with.

Theorem 2.2.9 (Spectral theorem for compact Hermitian operators [10, Theorem 2.3.5])**.** *Suppose $\hat{A}$ is a compact Hermitian operator on a Hilbert space $\mathcal{H}$. Then there is a countable orthonormal basis of $\mathcal{H}$ consisting of eigenvectors of $\hat{A}$.*

These eigenvectors form a complete orthonormal basis in Hilbert space and, therefore, any vector can be expressed using this eigenvector basis. Given an operator which satisfies the conditions in Theorem 2.2.9, we can generate an orthonormal basis for the Hilbert space $\mathcal{H}$. Depending on the cardinality of the Hilbert space dimension, we will denote the eigenvector basis differently.

If the Hilbert space has dimension N, we will denote its basis set as $\{|n\rangle\}_{n=0,\cdots,N-1}$. The eigenkets are ordered by increasing eigenvalue, with the eigenket corresponding to the smallest eigenvalue denoted as $|0\rangle$. If the basis set is countable infinite, then it can be denoted as $\{|n\rangle\}_{n\in\mathbb{N}}$, if there is no lower and upper bound on the eigenvalues, $\mathbb{Z}$ can be used to index the set of eigenvalues. If the basis is uncountable infinite-dimensional, it will be indexed by its eigenvector itself, so

$$\{|\alpha\rangle : \hat{A}\,|\alpha\rangle = \lambda_\alpha\,|\alpha\rangle \text{ and } \langle\alpha|\alpha\rangle = 1\}. \tag{2.17}$$

Assuming that the basis dimension is finite or countable infinite, we use the basis to write any state $|\psi\rangle$ as a column vector. The component of $|\psi\rangle$ along the nth eigenvector is given by the inner product $\psi_n := \langle n|\psi\rangle$. Using this, we can write

$$|\psi\rangle = \sum_n \langle n|\psi\rangle\,|n\rangle = \sum_n \psi_n\,|n\rangle = \begin{pmatrix}\psi_1\\ \psi_2\\ \vdots\end{pmatrix}. \tag{2.18}$$

Because the basis is orthonormal and we assume that $\langle\psi|\psi\rangle = 1$, we can show by Parseval's identity [11], that $\sum_n |\langle n|\psi\rangle|^2 = 1$, which means that the infinite sum over the eigenvectors is well-defined.

This suggests we can view the state $|\psi\rangle$ as a vector in the basis generated by the eigenvalues of an operator satisfying the requirements of Theorem 2.2.9. By convention, $|\psi\rangle$ is considered to be a column vector.

This also invites us to define a basis for the identity operator: $\hat{I} := \sum_n |n\rangle\langle n|$. Applying it to the state $|\psi\rangle$, we get:

$$|\psi\rangle = \left(\sum_n |n\rangle\langle n|\right)|\psi\rangle = \sum_n \psi_n\,|n\rangle, \tag{2.19}$$

which is equal to $|\psi\rangle$, as discussed above. We can apply the same method to operators:

$$\hat{A} = \hat{I}\hat{A}\hat{I} \tag{2.20}$$

$$= \left(\sum_n |n\rangle\langle n|\right)\hat{A}\left(\sum_m |m\rangle\langle m|\right) \tag{2.21}$$

$$= \sum_{mn} A_{mn}\,|n\rangle\langle m|, \tag{2.22}$$

with $A_{mn} = \langle n|\hat{A}|m\rangle$ and we can thus also write $\hat{A}$ as a matrix with elements A_{mn}, where we use a matrix basis elements $|n\rangle\langle m|$. In matrix representation, Hermitian operators are complex-conjugate matrices since

$$A_{mn} = \langle n|\hat{A}|m\rangle \tag{2.23}$$
$$= \langle m|\hat{A}^\dagger|n\rangle^* \tag{2.24}$$
$$= \langle m|\hat{A}|n\rangle^* \tag{2.25}$$
$$= A^*_{nm}. \tag{2.26}$$

Finally, we can also write the bras as row vectors:

$$\langle\phi| = \sum_n \phi^*_n \langle n| = (\phi^*_1, \phi^*_2, \cdots). \tag{2.27}$$

This ensures that the inner product $\langle\phi|\psi\rangle$ can also be viewed as a product of row and column vectors which results in a scalar.

2.3 The Measurement Postulates of Quantum Mechanics

In quantum mechanics, a measurement is an active operation that impacts the system tremendously, in stark contrast to the classical world. Physical observables, such as energy, position, momentum, and angular momentum, are represented by Hermitian operators. For a system in state $|\psi\rangle$, the expectation value of the measurement of a physical observable is

$$\langle\psi|\hat{A}|\psi\rangle\,, \tag{2.28}$$

where $\hat{A}$ is the operator associated with the physical observable. The assumption that $\hat{A}$ is Hermitian ensures that this expectation value is real.

When the eigenvalue spectrum is discrete, the physical quantity is considered quantised. Often, the observables considered cannot be represented by compact operators, which leads to some technical complications, because the eigenkets are no longer included in the Hilbert space. These issues will be addressed using the rigged Hilbert space formalism in Chapter 3. However, in this chapter, we will only work with finite-dimensional operators, so all of our operators are compact, and we do not require the rigged Hilbert space formalism.

We will now present the measurement postulates without using the rigged Hilbert space formalism. At the end of Chapter 3, the measurement postulates will be reinstated in the rigged Hilbert space formalism.

Postulate 2.3.1. *Every measurable physical quantity is associated with a Hermitian operator $\hat{A} : D(\hat{A}) \to \mathcal{H}$. The result of measuring a physical quantity must be one of the eigenvalues of the corresponding observable $\hat{A}$.*

Therefore, when the system is in any arbitrary state, the value which a physical quantity is allowed to take is still restricted to the spectrum of its associated operator. The quantum state of the system, in turn, determines the probability that the physical quantity will take a particular value after measurement, which is determined through the following postulate.

Postulate 2.3.2. *When a physical quantity associated with the operator $\hat{A}$ is measured on a system in a normalized [4] state $|\psi\rangle$, the probability of obtaining the eigenvalue a_n with associated eigenket $|\psi_n\rangle$ is given by $\mathbb{P}[a_n] = |\langle\psi_n|\psi\rangle|^2$ in the case of a non-degenerate discrete spectrum. In the case of a non-degenerate dense spectrum, it is given by the probability measure $d\mathbb{P}[\alpha] = |\langle\alpha|\psi\rangle|^2 da$, where $|\alpha\rangle$ is the eigenket associated with the eigenvalue α.*

Now we discuss the effect of the measurement on the system. If, immediately after the first measurement, you perform a second measurement, it would only make sense that this second measurement will produce the same result as the first. This suggests that the quantum state has collapsed into the eigenspace of the measured eigenvalue, meaning that, contrary to the classical world, the act of measurement changed the quantum state.

Postulate 2.3.3. *If the measurement of the physical quantity associated with $\hat{A}$ on the system in the state $|\psi\rangle$ gives the result α, then the state of the system immediately after the measurement is the normalised projection of $|\psi\rangle$ onto the eigensubspace associated with α:*

$$|\psi\rangle \mapsto \frac{\hat{P}_\alpha |\psi\rangle}{\sqrt{\langle\psi|\hat{P}_\alpha|\psi\rangle}}, \tag{2.29}$$

Where $\hat{P}_\alpha$ is the projection operator to the eigensubspace belonging to α, if α is a nondegenerate eigenvalue, then the projection operator is given by:

$$\hat{P}_\alpha = |\alpha\rangle\langle\alpha|. \tag{2.30}$$

For compact operators, we know from the Spectral Theorem 2.2.9 that their eigenkets are part of the Hilbert space. However, this does not hold for non-compact operators, like the position and momentum operators, which we will discuss in Section 3.3. Therefore, the collapse of the wavefunction after a position measurement is ill-defined unless we move to the rigged Hilbert space formalism, which we discuss in Chapter 3.

2.4 Time Evolution Postulate

Although you can derive the Schrödinger equation for the time evolution of the system from a basic set of assumptions, we are going to assert it as a postulate, like many other texts on the topic. One pertinent way to derive the Schrödinger equation is actually through Bellman's equation from optimal control theory [12]. However, it is not one of the mainstream derivations, and is unfortunately outside of the scope of this text to discuss.

Postulate 2.4.1. *The time evolution of the state vector $|\psi(\cdot)\rangle : \mathbb{R} \to \mathcal{H}$ is governed by the Schrödinger equation:*

$$i\hbar\frac{\mathrm{d}}{\mathrm{d}t}|\psi(t)\rangle = \hat{H}(t)|\psi(t)\rangle, \tag{2.31}$$

[4] a normalized state has the unit norm: $\langle\psi|\psi\rangle = 1$.

where $\hat{H} : \mathbb{R} \times \mathcal{H} \to \mathcal{H}$ is the observable associated with the total energy of the system, which is also often called the Hamiltonian of the system.

The Hamiltonian is the operator associated with the energy of the system. In the case of a time-independent Hamiltonian $\hat{H} : \mathcal{H} \to \mathcal{H}$, the solution of the Schrödinger equation can be immediately written down using the matrix exponential:

$$|\psi(t)\rangle = e^{-i\hat{H}t/\hbar} |\psi(0)\rangle . \tag{2.32}$$

If we are using the energy eigenvector basis for both the quantum state and the Hamiltonian itself, then the Hamiltonian operator can be written as a diagonal matrix: $\hat{H} = \text{diag}(E_1, E_2, \cdots, E_n)$, with $E_{1,2,\cdots,n} \in \mathbb{R}$ its eigenvalues. The matrix exponential $e^{-i\hat{H}t/\hbar}$, then simply becomes:

$$e^{-i\hat{H}t/\hbar} = \text{diag}(e^{-iE_1t/\hbar}, e^{-iE_2t/\hbar}, \cdots, e^{-iE_nt/\hbar}). \tag{2.33}$$

Which means that in the energy eigenvector basis of a time-dependent Hamiltonian, the solution of the Schrödinger equation is rather straightforward. However, we will often deal with composite Hamiltonians like: $\hat{H}(t) = \hat{H}_0 + \hat{H}_I(t)$, with a time-independent part $\hat{H}_0$ and a time-dependent part $\hat{H}_I(t)$, in such cases, we will still write the quantum states in the eigenvector basis generated by $\hat{H}_0$, but the Schrödinger equation evolves using the composite Hamiltonian $\hat{H}(t)$, meaning that the Schrödinger equation is no longer easily solvable.

We can also easily show that if the state $|\psi(t)\rangle$ is normalised at the time $t = t_0$, it will remain normalised forever because $\frac{\mathrm{d}}{\mathrm{d}t}(\langle\psi|\psi\rangle) = 0$. To show this, we can take the Hermitian conjugate on both sides of the Schrödinger equation to get:

$$i\hbar \frac{\mathrm{d}}{\mathrm{d}t} \langle\psi(t)| = -\hat{H}(t) \langle\psi(t)| . \tag{2.34}$$

Using that, we can compute:

$$\frac{\mathrm{d}}{\mathrm{d}t}(\langle\psi(t)|\psi(t)\rangle) = \left(\frac{\mathrm{d}}{\mathrm{d}t} \langle\psi(t)|\right) |\psi(t)\rangle + \langle\psi(t)| \left(\frac{\mathrm{d}}{\mathrm{d}t} |\psi(t)\rangle\right) \tag{2.35}$$

$$= \frac{i}{\hbar} \left(\langle\psi(t)|\hat{H}(t)|\psi(t)\rangle - \langle\psi(t)|\hat{H}(t)|\psi(t)\rangle\right) . \tag{2.36}$$

$$= 0. \tag{2.37}$$

Which shows that quantum states do remain normalized under time evolution by the Schrödinger equation. In Section 2.6, we will give an example of a Hamiltonian operator on the Hilbert space $\mathcal{H} = \mathbb{C}^2$.

2.5 Heisenberg Representation of Quantum Mechanics

We will now use the Heisenberg representation of quantum mechanics, where the operators carry the time dependency of the system, and quantum states are time-independent. In contrast to

the Schrödinger representation introduced in postulate 2.4.1, where the quantum states are time-varying according to the Schrödinger equations and the operators are static. These two pictures end up being physically equivalent. The Schrödinger and Heisenberg representations can be written as follows:

$$\langle\psi(t)|\hat{A}_S|\phi(t)\rangle = \langle\psi(0)|\hat{A}_H(t)|\phi(0)\rangle\,, \tag{2.38}$$

where $\hat{A}_S$ denotes the operator in the Schrödinger picture and $\hat{A}_H(t)$ denotes it in the Heisenberg picture. We would like to now deduce the equation of motion of the operator in the Heisenberg picture. We can start by writing:

$$|\phi(t)\rangle = \hat{U}(t,0)\,|\phi(0)\rangle\,. \tag{2.39}$$

Where it can be argued through applying the Schrödinger equation that

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{U}(t,0) = \frac{i}{\hbar}\hat{H}_S(t). \tag{2.40}$$

We can then write

$$\langle\psi(t)|\hat{A}_S|\phi(t)\rangle = \langle\psi(0)|\hat{U}^\dagger(t)\hat{A}_S\hat{U}(t)|\phi(0)\rangle \tag{2.41}$$

Meaning that we can set

$$\hat{A}_H(t) = \hat{U}^\dagger(t)\hat{A}_S\hat{U}(t). \tag{2.42}$$

We can then take the time derivative on both sides:

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{A}_H(t) = \frac{\mathrm{d}}{\mathrm{d}t}\Big(\hat{U}^\dagger(t)\Big)\hat{A}_S\hat{U}(t) + \hat{U}^\dagger(t)\hat{A}_S\frac{\mathrm{d}}{\mathrm{d}t}\Big(\hat{U}(t)\Big) + \hat{U}^\dagger(t)\frac{\partial\hat{A}_S}{\partial t}\hat{U}(t), \tag{2.43}$$

$$= \frac{i}{\hbar}\hat{U}^\dagger(t)\hat{H}_S(t)\hat{A}\hat{U}(t) - \frac{i}{\hbar}\hat{U}^\dagger(t)\hat{A}\hat{H}_S(t)\hat{U}(t) + \hat{U}^\dagger(t)\frac{\partial\hat{A}_S}{\partial t}\hat{U}(t), \tag{2.44}$$

$$= \frac{i}{\hbar}\left[\hat{U}^\dagger(t)\hat{H}_S(t)\hat{U}^\dagger(t)\hat{U}(t)\hat{A}\hat{U}(t) - \hat{U}^\dagger(t)\hat{A}\hat{U}^\dagger(t)\hat{U}(t)\hat{H}_S(t)\hat{U}(t)\right] + \left(\frac{\partial\hat{A}_S}{\partial t}\right)_H, \tag{2.45}$$

$$= \frac{i}{\hbar}\left[\hat{H}_H(t)\hat{A}_H(t) - \hat{A}_H(t)\hat{H}_H(t)\right] + \left(\frac{\partial\hat{A}_S}{\partial t}\right)_H, \tag{2.46}$$

$$= \frac{i}{\hbar}\left[\hat{H}_H(t), \hat{A}_H(t)\right] + \left(\frac{\partial\hat{A}_S}{\partial t}\right)_H. \tag{2.47}$$

In the last line, we use the commutator operation $[\hat{A},\hat{B}] := \hat{A}\hat{B} - \hat{B}\hat{A}$. In the case that the Hamiltonian is time-independent, we have that it commutes with $\hat{U}(t)$ and therefore, $\frac{\mathrm{d}}{\mathrm{d}t}H_H(t) = 0$, meaning that $\hat{H}_H = \hat{H}_S$. If the Hamiltonian $\hat{H}$ and the operator $\hat{A}_S$ are both time-independent, the equation of motion for the operator $\hat{A}_H$ in the Heisenberg representation becomes

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{A}_H(t) = \frac{i}{\hbar}\left[\hat{H}, \hat{A}_H(t)\right]. \tag{2.48}$$

2.6 Two-level Quantum Systems

Whilst moving towards the quantum systems upon which we will apply quantum optimal control, we will start by discussing the simplest non-trivial quantum system, those with just two basis states. Their Hilbert space is $\mathcal{H} = \mathbb{C}^2$. All quantum states should, of course, be normalised, but if we ensure that our initial quantum states are normalised, the Schrödinger equation will preserve this normalisation. Because this Hilbert space is finite-dimensional, we won't have to resort to the rigged Hilbert space formalism.

We will now give some basic explanations about the real-world quantum systems we are trying to model with two-level quantum systems for readers with a mathematical background. However, it is not feasible to treat everything in full detail. If interested, any introductory quantum mechanics textbook like [13] would be a good starting point.

Many real-world systems can be modelled using a two-level system, as long as only two quantum states are relevant in the system. For example, the electron of a hydrogen atom interacting with an external laser can be treated as a two-level system if the laser can only transfer the electron between the ground state and the first excited state.

The Bohr model states that an atom consists of a positively charged nucleus consisting of protons and neutrons, surrounded by orbiting electrons, as depicted in Figure 2.1. It is one of the simplest quantum models, which can still predict the approximate energy levels of hydrogen.

The Bohr model of the hydrogen atom states that the electron can be in energy levels $E_n \propto -\frac{1}{n^2}$, for $n \in \mathbb{N}^+$. The ground state $|1\rangle$ is most tightly bound by the proton, and as you increase n, the electron is further separated from the proton. The electron can only transfer between different energy levels using a laser if the frequency of the laser corresponds to the difference in energy between those energy levels. In the real world, things are more complicated due to the conservation of spin and angular momentum. However, in this thesis, we will not discuss these complexities.

Atoms larger than hydrogen can also be modelled using the Bohr model. However, because of the complex interplay between the electrons when you have multiple, it is extremely difficult to retrieve theoretical predictions of the energy levels from the Bohr model without applying additional approximations. One common approximation is only computing the energy level of the outermost electron, while reducing the effective nuclear charge of the nucleus to account for screening by the inner electrons. The field of Quantum Chemistry concerns itself with accurately computing the energy levels of larger atoms and molecules by applying such approximations.

The energy related to a certain energy level can be theoretically predicted through the Bohr model (or more advanced methods). However, since we are taking a modelling approach, we will often simply impose each level to have a certain energy. Only the energy differences between states matter; you can arbitrarily shift the energy of the levels up or down, and it will only produce a global phase factor, which is not relevant.

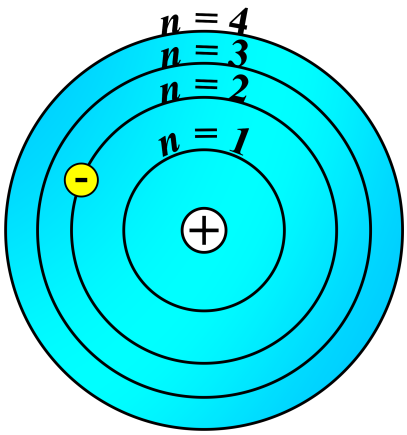

Figure 2.1: A schematic of the hydrogen atom in the Bohr model [14], the yellow point marked with a minus sign represents the electron moving around the nucleus shown as a white circle with a plus. Image taken from [15].

According to the Bohr model, an atom has many energy levels. However, oftentimes only two energy levels are relevant, for example, when the laser and the atom's internal dynamics don't push the atom into any third state. That means that there are no relevant decays which push the atom into a third state, and the laser energy is such that it is only resonant with the difference in energy between our two energy levels. In this thesis, we will often a priori impose the quantum system to have a certain number of states.

Internal Dynamics of Two-Level System

When the two-level system is not influenced by an outside force, the population of each quantum state will remain stable unless there is a decay process from a higher energy state to a lower energy state. Although it is possible to determine which decays are allowed or disallowed in physics, we will not describe this in detail. To describe quantum systems where states can decay with rigour in quantum physics, we will have to work with so-called open quantum systems, which we discuss in Section 2.9.

Quite often, the lifetime of a state with respect to a certain decay process is orders of magnitude larger than the time interval of the dynamics you are looking at; in those cases, we will not consider this decay process in our models.

The two-level quantum system has the Hilbert space $\mathcal{H} = \mathbb{C}^2$. By the Spectral Theorem 2.2.9, we are assured that it is possible to write the Hamiltonian of the two-level system using an orthonormal basis consisting of its own eigenvectors. The Hamiltonian representing the dynamics of the electron in the atom, without the presence of an external electric field, can be written as:

$$\hat{H}_0 = \begin{bmatrix} E_0 & 0 \\ 0 & E_1 \end{bmatrix}, \tag{2.49}$$

where the coordinate basis used is generated from the eigenvectors of the Hamiltonian $\hat{H}_0$, with

$|0\rangle = \begin{bmatrix}1\\0\end{bmatrix}$ and $|1\rangle = \begin{bmatrix}0\\1\end{bmatrix}$ being the associated eigenvectors. $E_{0,1}$ are the possible energy levels of the electron. We set $E_1 > E_0$. Any state $|\psi\rangle$ can be written as $\begin{bmatrix}c_1\\c_2\end{bmatrix}$, with $c_{1,2} \in \mathbb{C}$ and $|c_1|^2 + |c_2|^2 = 1$ to ensure normalisation.

Laser Interaction

In its simplest form[5], lasers are devices which output a focused beam of electromagnetic waves with a single frequency ω, some initial phase ϕ and time-varying amplitude $A(t)$. Here, $\hat{e}$ is a unit vector which sets the polarisation of the electromagnetic wave; it determines the direction in which the electric field points. The electric field experienced by the electron, which is fixed in space, can be written as:

$$E(t) = A(t)e^{i(\omega t+\phi)}\vec{e}. \tag{2.50}$$

Only the real part of the electric field is physically relevant; however, we will now describe it using a complex exponential and discard the complex part later, as that will ease the computations.

We will now examine how the laser light interacts with the quantum system. Our system, which is an electron bound by a nucleus, for simplicity, other electrons are ignored. Since the system is electrically neutral, its centre of mass will not move under the influence of an external electric field. However, the positively charged nucleus and the negatively charged electron move in opposite directions in the presence of an electric field.

This separation of charge can be described using the electric dipole moment, defined as:

$$\vec{\mu} = \sum_j q_j \vec{r}_j, \tag{2.51}$$

where q_j and $\vec{r}_j$ are the charge and position[6] of the jth point particle of the system.

When the external uniform electric field emitted by the laser is introduced, the potential energy of the classical system changes with:

$$\Delta U = -\vec{\mu} \cdot \vec{E}(t). \tag{2.52}$$

To simplify our computations, we want to get rid of that dot product in the equation above,

$$\Delta U = -\mu E_{||}(t), \tag{2.53}$$

[5]In the framework of quantum mechanics, the particle-wave duality implies that a laser can be viewed as a device emitting photons which all have the same energy. When this photon energy is equal to the difference between two energy levels of a system, the photon can be absorbed, bringing the electron from the lower to the higher state. However, this quantum-mechanical treatment of the laser is often unnecessary. Whenever appropriate, we will treat the laser output as an electromagnetic wave and describe it using classical electromagnetic theory.

[6]For neutral systems, the reference point from which we measure positions is not relevant.

where $E_{||} : \mathbb{R} \to \mathbb{R}$ denotes the electric field parallel to the dipole moment over time and $\mu = |\vec{\mu}|$.

After the two-level system is quantised, the electric dipole moment scalar μ becomes the operator $\hat{\mu} : \mathbb{C}^2 \to \mathbb{C}^2$. In classical mechanics, observables are scalar functions of the state of the system; however, in quantum mechanics, operators represent physical observables, and their expected value is given by $\langle\psi|\hat{\mu}|\psi\rangle$, which is a scalar itself. The electric dipole moment operator can be represented as the Hermitian matrix:

$$\hat{\mu} = \begin{bmatrix} 0 & \mu_{01} \\ \mu_{01}^* & 0 \end{bmatrix}. \tag{2.54}$$

Computing $\mu_{01} \in \mathbb{C}$ for a system of quantum particles is more complicated than for a system of classical particles because it requires integrating over each particle's wave function. However, we will not compute terms like μ_{01} from underlying principles, so we can skip over those difficulties. Generally, μ_{01} can be a complex number; however, we will absorb its complex phase into the complex phase of the electric field, such that we can simply work with a real dipole momentum μ.

The Hamiltonian of the two-level system represents the energy of the system. From a classical perspective, the electric field introduces an energy shift of $-E_{||}(t)\mu$. Therefore, we should add the term $-E_{||}(t)\hat{\mu}$ to the Hamiltonian:

$$\hat{H}(t) = \hat{H}_0 - E_{||}(t)\hat{\mu}. \tag{2.55}$$

This can be written in matrix form as:

$$\hat{H}(t) = \begin{bmatrix} E_0 & \mu^* E_{||}(t) \\ \mu(E_{||}(t))^* & E_1 \end{bmatrix} = \begin{bmatrix} E_0 & -\mu A(t) e^{i(\omega t+\phi)} \\ -\mu A(t) e^{-i(\omega t+\phi)} & E_1 \end{bmatrix} \tag{2.56}$$

Here, we have to take the complex conjugate of the electric field to ensure that the Hamiltonian remains Hermitian. For the laser electric field to resonantly drive the state transfer, it will require that $\omega_0 = \frac{E_1 - E_0}{\hbar}$. For simplicity, we will set $\hbar = 1$ as is common in parts of the physics literature[7], so we will also set $\omega_0 = E_1 - E_0$.

We can now write down the Schrödinger equation, using the basis coefficients $c_j(t) = \langle\Psi(t)|j\rangle$, for $j = 1, 2$ as follows:

$$i\frac{\partial}{\partial t}\begin{bmatrix} c_1(t) \\ c_2(t) \end{bmatrix} = \begin{bmatrix} E_0 & -\mu A(t) e^{i(\omega t+\phi)} \\ -\mu A(t) e^{-i(\omega t+\phi)} & E_1 \end{bmatrix} \cdot \begin{bmatrix} c_1(t) \\ c_2(t) \end{bmatrix}. \tag{2.57}$$

To ensure normalisation, we set $|c_1(0)|^2 + |c_2(0)|^2 = 1$. To simplify the dynamics, we assume that the electric field is resonant with the internal dynamics: $\omega = \omega_0$, and then we apply the time-dependent transformation $B(t) = \mathrm{diag}(e^{-iE_0 t}, e^{-iE_1 t})$. Such that $|\tilde{\Psi}(t)\rangle = B(t)^{-1}|\Psi(t)\rangle$. Which leads to:

[7] This is often done to avoid needing to write $\hbar$'s everywhere; for the final result, the $\hbar$'s can always be put back through dimensional analysis.

$$i\frac{\partial}{\partial t}\begin{bmatrix}\tilde{c}_1(t)\\ \tilde{c}_2(t)\end{bmatrix} = \begin{bmatrix}\mu A(t) & 0\\ 0 & \mu A(t)\end{bmatrix}\begin{bmatrix}\tilde{c}_1(t)\\ \tilde{c}_2(t)\end{bmatrix}. \tag{2.58}$$

We can define $u(t) := \mu A(t)$. We can now split $\tilde{c}_{1,2}(t)$ into their real and imaginary parts, $\tilde{c}_j(t) = x_j(t) + iy_j(t)$, for $j = 1, 2$. Doing this, and writing the system of coupled differential equations in terms of $x_j(t)$ and $y_j(t)$, we get:

$$\dot{x}_1(t) = u(t)y_2(t), \tag{2.59}$$
$$\dot{x}_2(t) = u(t)y_1(t), \tag{2.60}$$
$$\dot{y}_1(t) = -u(t)x_2(t), \tag{2.61}$$
$$\dot{y}_2(t) = -u(t)x_1(t). \tag{2.62}$$

Notice that both $x_1^2(t) + y_2^2(t) = \alpha_1$ and $x_2^2(t) + y_1^2(t) = \alpha_2$ are conserved quantities with $\alpha_{1,2} \in \mathbb{R}$ being determined by the initial conditions. We can then parametrise them as:

$$\begin{bmatrix}x_1(t)\\ x_2(t)\\ y_1(t)\\ y_2(t)\end{bmatrix} = \begin{bmatrix}\alpha_1\cos(\theta(t))\\ \alpha_2\cos(\phi(t))\\ \alpha_1\sin(\phi(t))\\ \alpha_2\sin(\theta(t))\end{bmatrix} \tag{2.63}$$

The differential equations then turn into just two equations:

$$\dot{\theta}(t) = -u(t), \tag{2.64}$$
$$\dot{\phi}(t) = -u(t). \tag{2.65}$$

Which can be tackled using standard control theory. Systems with more states can be reformulated to control problems on the surfaces of higher-dimensional spheres. For example, the three-state system can be formulated as a control problem on a three-dimensional sphere.

Supposing that a pulse starts in the ground state, it would be completely transferred to the excited state when $\theta = -\frac{\pi}{2}$, therefore, we get the condition that:

$$\frac{\pi}{2} = \int_0^T u(t)dt. \tag{2.66}$$

A control function $u(t)$ which satisfies the above equation is a π-pulse. If $u(t)$ is constant, it satisfies $u(t) = \frac{\pi}{2T}$. We define the Rabi frequency as the speed at which population inversion occurs. If the electric field amplitude is constant over time, the population of the ground state: $|x_1(t)|^2 + |x_2(t)|^2$ oscillates at a frequency of $\omega_R = 2\mu A$, this is thus becomes the Rabi frequency.

The electric field interaction is thus:

$$\mu E_{||}(t) = \frac{\pi}{2T}e^{i(\omega_0 t+\phi)}, \tag{2.67}$$

where the phase ϕ is arbitrary. However, only the real part of the electric field is physically relevant, so we should only use the real part:

$$\mu E_{||}(t) = \frac{\pi}{2T}\cos(\omega_0 t + \phi). \tag{2.68}$$

Note that if we would double the amplitude $u(t)$ of the pulse, then you get $\theta(T) = 2\pi$, meaning that you are back in the ground state again and that at some time $T' < T$, $\theta(T') = \pi$ and the system was fully in the excited state. If you would triple the amplitude $A(t)$, we would get that $\theta(T) = 3\pi$, meaning that the system is again fully in the excited state after already experiencing a full population inversion process.

Looking at the above equation, it becomes clear that if the system is initially in the excited state, then by adding laser light, you can force the system to transfer to the ground state. Since the laser light is adding energy to the system, this phenomenon does not make sense at first glance. The laser light is forcing the electron to emit radiation itself at the same frequency as the laser, through a process called stimulated emission. Unfortunately, we do not have space to discuss the origins of this process further, but any quantum physics textbook, like [13], will discuss it in detail.

2.7 Rotating Wave Approximation

We will now more closely examine the time dependence of the electric field $E_{||}(t)$, we will no longer require that its frequency matches the internal frequency ω_0 of the two-level system.

We rewrite the Hamiltonian $\hat{H}(t)$ as follows using the Pauli operators [8]:

$$\hat{H}_0 = -\frac{1}{2}\omega_0\hat{\sigma}_z, \tag{2.69}$$

$$\hat{H}_I(t) = -\mu E_{||}(t)\hat{\sigma}_x. \tag{2.70}$$

Here, $\hat{H}_0$ and $\hat{H}_I(t)$ denote the static and interaction part of the Hamiltonian, respectively. Using the Heisenberg equation of motion for the Pauli operators,

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{\sigma}_j(t) = \frac{i}{\hbar}[\hat{H}(t), \hat{\sigma}_j], \quad j = x, y, z, \tag{2.71}$$

and the commutation relationship for Pauli operators:

$$[\hat{\sigma}_j, \hat{\sigma}_k] = 2i\epsilon_{jkl}\hat{\sigma}_l, \tag{2.72}$$

where ϵ_{ijk} is the Levi-Civita symbol[9], we obtain the equations of motion for the Pauli operators:

[8] The Pauli operators are the set of matrices $(\hat{\sigma}_x, \hat{\sigma}_y, \hat{\sigma}_z) = \left(\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\right)$. The set $\{i\hat{\sigma}_x, i\hat{\sigma}_y, i\hat{\sigma}_z\}$ forms a basis of the Lie algebra of SU(2) and are fundamental to quantum mechanics.

[9] If (ijk) is an even permutation of (xyz), it is equal to one, if (ijk) is an odd permutation, it is equal to minus one and if there are repeated entries, it is equal to zero.

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix}\hat{\sigma}_x(t)\\ \hat{\sigma}_y(t)\\ \hat{\sigma}_z(t)\end{bmatrix} = \begin{bmatrix}0 & \omega_0 & 0\\ -\omega_0 & 0 & 2\mu E_{||}(t)\\ 0 & -2\mu E_{||}(t) & 0\end{bmatrix}\cdot\begin{bmatrix}\hat{\sigma}_x(t)\\ \hat{\sigma}_y(t)\\ \hat{\sigma}_z(t)\end{bmatrix}, \tag{2.73}$$

where $\omega_0 = \frac{E_1 - E_0}{\hbar}$. We can replace the Pauli operators $\hat{\sigma}_j(t)$ with their expectation values $s_j(t) := \langle\psi|\hat{\sigma}_j(t)|\psi\rangle$, for $j = x, y, z$; this is allowed because we are treating the laser field classically.

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix}s_x(t)\\ s_y(t)\\ s_z(t)\end{bmatrix} = \begin{bmatrix}0 & \omega_0 & 0\\ -\omega_0 & 0 & 2\mu E_{||}(t)\\ 0 & -2\mu E_{||}(t) & 0\end{bmatrix}\cdot\begin{bmatrix}s_x(t)\\ s_y(t)\\ s_z(t)\end{bmatrix}. \tag{2.74}$$

It can be shown that $s_1^2(t) + s_2^2(t) + s_3^2(t) = 1$, for all $t \in \mathbb{R}$ if the equality holds at $t = 0$, meaning that the probability is conserved at all times.

We can write equation (2.74) using a cross product as follows:

$$\frac{\mathrm{d}}{\mathrm{d}t}\vec{s}(t) = \vec{\Omega}^F(t) \times \vec{s}(t), \tag{2.75}$$

with the vector $\vec{\Omega}^F(t) = \begin{bmatrix}-2\mu E_{||}(t) & 0 & -\omega_0\end{bmatrix}^T$. We assume that $\omega \simeq \omega_0$, meaning that the frequency of the laser is close to the internal frequency of the atom, and the laser can drive a transition between the two levels. That means that we can write:

$$E_{||}(t) = A(t)\cos(\omega t + \phi). \tag{2.76}$$

We assume that $\omega_0 \gg \omega_R = |2\mu A(t)|$ because otherwise, the dynamics induced by the laser are happen at a similar frequency as the internal dynamics of the system, which would invalidate the rotating wave approximation. Additionally, we assume the amplitude $A(t)$ changes adiabatically, meaning that it changes slowly compared to the other dynamics involved; therefore, in the equations of motion, it can be treated as a constant. This is a broader concept in quantum mechanics, called the adiabatic theorem. For more information, consider [16].

We can decompose the vector $\Omega^F(t)$ into three parts: a stationary part, a clockwise-rotating part and a counter-clockwise-rotating part:

$$\vec{\Omega}^F(t) = \vec{\Omega}_0 + \vec{\Omega}_-(t) + \vec{\Omega}_+(t), \tag{2.77}$$

with

$$\vec{\Omega}_0 = \begin{bmatrix}0\\ 0\\ -\omega_0\end{bmatrix}, \quad \vec{\Omega}_\pm(t) = \begin{bmatrix}-\mu A(t)\cos(\omega t)\\ \mp\mu A(t)\sin(\omega t)\\ 0\end{bmatrix}. \tag{2.78}$$

With a external electric field amplitude $A(t) = 0$, the solution is

$$\vec{s}_{A=0}(t) = \begin{bmatrix}\cos(\omega_0 t)\\ -\sin(\omega_0 t)\\ 0\end{bmatrix}. \tag{2.79}$$

We will set $s_1(0) = 1, s_2(0) = s_3(0) = 0$ as initial conditions for simplicity. Because we assume $|2\mu A(t)| \ll \omega_0$, we can treat the electric field as a small perturbation and apply first-order perturbation theory to analyse its effects. We can expand $\vec{s}(t)$ as:

$$\vec{s}(t) = \vec{s}_{A(t)=0}(t) + \delta\vec{s}(t) + \mathcal{O}\left((\mu A(t))^2\right), \tag{2.80}$$

where $s_{A(t)}$ is the unperturbed solution and $\delta\vec{s}(t)$ is the first-order correction. Substituting the above equation into (2.74) and keeping all terms of order $\mathcal{O}(\mu A(t))$, we get

$$\frac{\mathrm{d}}{\mathrm{d}t}\delta\vec{s}(t) = (\vec{\Omega}_-(t) + \vec{\Omega}_+(t)) \times \vec{s}_{A(t)=0}(t). \tag{2.81}$$

The rotating wave approximation drops terms with a frequency of $\omega + \omega_0$ and keeps terms with a frequency of $\omega - \omega_0$ because the latter precesses so fast, that on average, they don't influence the dynamics on the timescales relevant to the system.

The term $\vec{\Omega}_-(t) \times \vec{s}_{A(t)=0}(t)$ produces such a high-frequency term.

$$\begin{aligned}\vec{\Omega}_-(t) \times \vec{s}(t)_{A(t)=0} &= -\mu A(t)\left[\cos(\omega t)\sin(\omega_0 t) + \sin(\omega t)\cos(\omega_0 t)\right] &(2.82)\\ &= -\mu A(t)\sin\left((\omega + \omega_0)t\right) &(2.83)\end{aligned}$$

However, $\vec{\Omega}_+(t) \times \vec{s}_{A(t)=0}(t)$ produces a term with a frequency of $\omega - \omega_0$ and thus $\vec{\Omega}_-(t)$ is discarded and $\vec{\Omega}_+(t)$ is kept. This is the core approximation of the rotating-wave approximation. That leaves us with

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\vec{s}(t) &= \left(\vec{\Omega}_0 + \vec{\Omega}_+(t)\right) \times \vec{s}(t), &(2.84)\\ &= \begin{bmatrix} -\omega_0 s_2(t) - \mu A(t) s_3(t)\sin(\omega t) \\ +\omega_0 s_1(t) + \mu A(t) s_3(t)\cos(\omega t) \\ -\mu A(t)[s_2(t)\cos(\omega t) - s_1(t)\sin(\omega t)]. \end{bmatrix} &(2.85)\end{aligned}$$

We can then make a coordinate transformation to a rotating frame with a frequency of ω:

$$\vec{\rho}(t) = \begin{bmatrix} u(t) \\ v(t) \\ w(t) \end{bmatrix} = \begin{bmatrix} \cos(\omega t) & \sin(\omega t) & 0 \\ -\sin(\omega t) & \cos(\omega t) & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \vec{s}(t). \tag{2.86}$$

In this new coordinate frame, the dynamics are:

$$\frac{\mathrm{d}}{\mathrm{d}t}\vec{\rho}(t) = \begin{bmatrix} 0 & -\Delta & 0 \\ +\Delta & 0 & \mu A(t) \\ 0 & -\mu A(t) & 0 \end{bmatrix} \vec{\rho}(t), \tag{2.87}$$

with $\Delta = \omega_0 - \omega$. We call this the optical Bloch equations, the nice thing is that we have no more terms of order $\mathcal{O}(\omega)$, meaning that $\vec{\rho}(t)$ no longer evolves at optical frequencies like $\vec{s}(t)$ did, which is a great improvement if you want to solve it analytically.

It can be checked that $u^2(t)+v^2(t)+w^2(t) = 1$ holds at all times. Take $|\psi(t)\rangle = c_0(t)\,|0\rangle + c_1(t)\,|1\rangle$, then

$$\begin{bmatrix} u(t) \\ v(t) \\ w(t) \end{bmatrix} = \begin{bmatrix} c_1(t)^* c_0(t) + c_1(t) c_0(t)^* \\ -i(c_1(t)^* c_0(t) - c_1(t) c_0(t)^*) \\ |c_0(t)|^2 - |c_1(t)|^2, \end{bmatrix} \tag{2.88}$$

where $w(t)$ is called the inversion and is related to the population difference of the two states, $v(t)$ is called the absorptive component, which enables energy transfers between the atom and the electric field. Lastly, $u(t)$ is the dispersive component, which governs the coherence between the two energy levels.

In the case of a constant amplitude pulse, when the laser frequency ω is slightly off-resonant compared to the internal frequency ω_0, the Rabi frequency $\omega_R = 2\mu A$ is affected by the detuning of the pulse and will shift upwards:

$$\omega_{R,\Delta} = \sqrt{\Delta^2 + (2\mu A)^2}. \tag{2.89}$$

Additionally, it will no longer be possible to fully transfer the atom from the ground state to the excited state, instead the maximal excited population which is achieved becomes:

$$P_{exc}(\Delta) = \frac{(2\mu A)^2}{\Delta^2 + (2\mu A)^2} = \frac{\omega_{R,0}^2}{\omega_{R,\Delta}^2}. \tag{2.90}$$

To recap, we have transformed the original Hamiltonian (2.69), to the following Hamiltonian:

$$\hat{H}(t) = -\frac{1}{2}\Delta\hat{\sigma}_z - \mu A(t)\hat{\sigma}_x, \tag{2.91}$$

with $A(t)$ being obtained from $E_{||}(t)$ using the relation:

$$E_{||}(t) = A(t)\cos(\omega t + \phi). \tag{2.92}$$

Analytic solutions are easier to obtain from this Hamiltonian because $\Delta \ll \omega_0$, thus if the laser is near-resonance, the Hamiltonian will not produce any rapid oscillations. For the rotating wave approximation, we need the following relations to hold: $\omega_R \ll \omega_0 \simeq \omega$. We will employ the rotating wave approximation in Section 4.4, and the optical Bloch equations (2.87) will be used for simulations in Chapter 7.

2.8 Density Matrix Formalism

Before we can go and discuss open quantum systems, we first have to introduce the concept of the density matrix, which will prove to be a very useful new way to write down quantum states. We will follow the introduction given in [17] in the remaining sections of this chapter.

Definition 2.8.1 (Density matrix)**.** *A density matrix is defined as a linear combination*

$$\hat{\rho} := \sum_j p_j \left|\psi_j\right\rangle \left\langle\psi_j\right|, \quad p_j \in \mathbb{R}^+, \left|\psi_j\right\rangle \in \mathcal{H} \tag{2.93}$$

where the $|\psi_j\rangle$ are (not necessarily orthogonal) state vectors in the Hilbert space. The populations p_j give the probability of the system being in the jth state. The density matrix $\hat{\rho}$ is a positive semi-definite Hermitian operator, which satisfies $\mathrm{Tr}[\hat{\rho}] = 1$.

If we want to work in a pre-specified Hilbert space basis $\{|i\rangle \text{ for } i = 1, \cdots, N\}$, then any density matrix can be written as:

$$\hat{\rho} = \sum_{i,j} \rho_{ij} \left|i\right\rangle \left\langle j\right|, \tag{2.94}$$

where $\mathrm{Tr}[\hat{\rho}] = 1$ and the diagonal elements ρ_{ii} represent the probability of being in the ith state. The off-diagonal elements ρ_{ij} are called the coherences and are conjugate-symmetric: $\rho_{ij} = \rho^*_{ji}$.

You can have pure states and mixed states in the density matrix formalism. Density matrices of pure states can be written down using just one state. For example, consider:

$$\hat{\rho} = \frac{1}{2}\left[\left|0\right\rangle \left\langle 0\right| + \left|0\right\rangle \left\langle 1\right| + \left|1\right\rangle \left\langle 0\right| + \left|1\right\rangle \left\langle 1\right|\right], \tag{2.95}$$

which can be written as $\hat{\rho} = |\psi\rangle \langle\psi|$, with $|\psi\rangle = \frac{1}{\sqrt{2}}\left[|0\rangle + |1\rangle\right]$.

Density matrices for mixed states can't be written down using just one state. Equivalently, we have that pure states have $\mathrm{Tr}[\hat{\rho}^2] = 1$ and mixed states fulfil $\mathrm{Tr}[\hat{\rho}^2] < 1$. One example of such a mixed state is:

$$\hat{\rho} = \frac{1}{2}[\left|0\right\rangle \left\langle 0\right| + \left|1\right\rangle + \left\langle 1\right|]. \tag{2.96}$$

A brief sketch of the proof is that only for pure states, there exists a basis transformation such that $\tilde{\rho}_{00} = 1$, which clearly means that $\mathrm{Tr}\left[\tilde{\hat{\rho}}^2\right] = 1$, because basis transformations preserve trace, also $\mathrm{Tr}[\hat{\rho}^2] = 1$. However, for mixed states, such a basis transformation is impossible by definition and thus $\mathrm{Tr}[\hat{\rho}^2] < 1$.

Given a state $|\psi\rangle = \sum_j c_j |j\rangle$, we can write the associated density matrix as

$$\hat{\rho} = \sum_{j,k} c_j c_k^* \left|j\right\rangle \left\langle k\right|. \tag{2.97}$$

Therefore, in the two-level system the density matrix is given by:

$$\hat{\rho} = \begin{bmatrix} |c_0|^2 & c_0 c_1^* \\ c_0^* c_1 & |c_1|^2. \end{bmatrix} \tag{2.98}$$

Which means that it indeed makes sense to talk about populations and coherences. The expected value of a state operator $\hat{O}$ is given as

$$\langle \hat{O} \rangle := \langle \psi | \hat{O} | \psi \rangle = \mathrm{Tr}\Big[\hat{O}\hat{\rho}\Big], \tag{2.99}$$

this can be checked by writing everything out in the same basis.

The time evolution of the density matrix is governed by the von-Neumann equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{\rho} = -\frac{i}{\hbar}[\hat{H}, \hat{\rho}] := \mathcal{L}\hat{\rho}, \tag{2.100}$$

where $\mathcal{L}$ is called the Liouvillian superoperator. Through routine computations, it can be checked that this is consistent with the Schrödinger equation. It can also be shown that the purity of the system is conserved over time: $\frac{\mathrm{d}}{\mathrm{d}t}\,\mathrm{Tr}[\hat{\rho}^2] = 0$.

2.9 Open Quantum Systems

We will now discuss composite systems, which consist of multiple connected subsystems, for which we need to introduce a new postulate for quantum mechanics to describe them. We will primarily work with open quantum systems, where our system of interest interacts with its environment.

The environment itself is too large to be affected by our system of interest and is modelled as a thermodynamic heat bath. Moreover, we only care about our system of interest, so we must find a way to only describe it together with its interactions with the environment, but not the environment itself.

Postulate 2.9.1 (Composite systems)**.** *The Hilbert space of a system contains N subsystems is the tensor product of the Hilbert spaces belonging to the subsystem: $\mathcal{H} = \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_N$. A state of the composite system is a unit vector in $\mathcal{H}$, and it can be written as the tensor product of the state in each subsystem: $|\psi\rangle = |\psi_1\rangle \otimes \cdots \otimes |\psi_N\rangle$.*

We will clarify the definition of the tensor product of Hilbert spaces in the case of the tensor product of two Hilbert spaces. However, it can readily be extended to the tensor product of an arbitrary number of Hilbert spaces. The underlying vector space of $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$ is spanned by the set:

$$\{|\alpha\rangle_1 \otimes |\beta\rangle_2 \,|\, \alpha \in A,\ \beta \in B)\}, \tag{2.101}$$

where A and B are the basis sets of $\mathcal{H}_1$ and $\mathcal{H}_2$ respectively. The inner product associated to $\mathcal{H}$ is defined as:

$$(|\phi_1\rangle \otimes |\phi_2\rangle\,, |\psi_1\rangle \otimes |\psi_2\rangle) = (|\phi_1\rangle\,, |\psi_1\rangle)_1 \cdot (|\phi_2\rangle\,, |\psi_2\rangle), \quad \forall\, |\phi_j\rangle\,, |\psi_j\rangle \in \mathcal{H}_j, j = 1, 2. \tag{2.102}$$

Which is extended by linearity to satisfy:

$$(|\phi_1\rangle \otimes |\phi_2\rangle\,, |\psi_1\rangle \otimes |\psi_2\rangle + |\pi_1\rangle \otimes |\pi_2\rangle) = (|\phi_1\rangle\,, |\psi_1\rangle)_1 \cdot (|\phi_2\rangle\,, |\psi_2\rangle) + (|\phi_1\rangle\,, |\psi_1\rangle)_1 \cdot (|\pi_2\rangle\,, |\pi_2\rangle), \tag{2.103}$$

for all $|\phi_j\rangle, |\psi_j\rangle, |\pi_j\rangle \in \mathcal{H}_j, j = 1, 2$. Through this extension, $\mathcal{H}$ now satisfies all the requirements to be considered a Hilbert space.

A basis for the composite Hilbert space can be created by taking the tensor product of all possible combinations of basis elements from each subsystem. For example, for a system containing two two-level subsystems, a possible basis is the set:

$$\{|0\rangle \otimes |0\rangle, |0\rangle \otimes |1\rangle, |1\rangle \otimes |0\rangle, |1\rangle \otimes |1\rangle\}. \tag{2.104}$$

We might be interested in only studying the dynamics of one of the subsystems, although it might still interact with the other subsystems. We can do this using the *reduced density matrix*, which is defined as follows in a system containing just two subsystems a and b:

$$\hat{\rho}_a := \mathrm{Tr}_b\, \hat{\rho}, \tag{2.105}$$

with the partial trace defined as:

$$\mathrm{Tr}_b \left[\sum_{i,j,k,l} |a_i\rangle \langle a_j| \otimes |b_k\rangle \langle b_l| \right] := \sum_{i,j} |a_i\rangle \langle a_j| \,\mathrm{Tr} \left[\sum_{k,l} |b_k\rangle \langle b_l| \right]. \tag{2.106}$$

When we study open quantum systems, we want to study the dynamics of a system which is affected by its environment. We do not care about the environment itself, so we can trace out the environment such as to only study the reduced density matrix of the system of interest.

2.10 Derivation of the Lindblad master equation

We will now study the Lindblad master equation, which combines the time evolution from the Schrödinger equation with the evolution due to the collapse of the wave function caused by interactions with the environment. We start with a Hilbert space $\mathcal{H}_T = \mathcal{H}_S \otimes \mathcal{H}_E$, which is decomposed into a system and environment subspace. For this derivation, we will use [18]. Our starting point is the von-Neumann equation:

$$\dot{\hat{\rho}}_T(t) = -\frac{i}{\hbar}[\hat{H}_T, \hat{\rho}_T(t)]. \tag{2.107}$$

From here onwards, we will set $\hbar = 1$ in this section. The quantity which we are interested in is $\hat{\rho}_S(t) = \mathrm{Tr}_E[\hat{\rho}_T(t)]$. We can decompose the Hamiltonian into a system, environment and interaction Hamiltonian:

$$\hat{H}_T = \hat{H}_S \otimes \hat{\mathbb{1}}_E + \hat{\mathbb{1}}_S \otimes \hat{H}_E + \alpha \hat{H}_I, \tag{2.108}$$

where α is the coupling strength between the system and the environment. The coupling strength α is assumed to be weak: $\alpha \ll 1$, to allow for a first-order treatment of the interaction effects. From now on, we will also drop the identity operators $\mathbb{1}_E$ or $\mathbb{1}_S$ if their presence is clear from context. We can write the interaction Hamiltonian as a sum of tensor products:

$$\hat{H}_I = \sum_j \hat{S}_j \otimes \hat{E}_j, \tag{2.109}$$

with S_j and E_j bounded linear operators on their respective Hilbert space $\mathcal{H}_S$ and $\mathcal{H}_E$. We will now use the interaction picture of quantum time evolution, which is a mix between the Schrödinger and Heisenberg pictures. In this picture, density matrices evolve according to the interaction Hamiltonian $\hat{H}_I$ and operators evolve according to the internal Hamiltonians $\hat{H}_S$ and $\hat{H}_E$. For any bounded operator $\hat{O}$ on the Hilbert space $\mathcal{H}_T$, its time-dependency is given by

$$\hat{O}(t) = e^{i(\hat{H}_S+\hat{H}_E)t}\hat{O}e^{-(\hat{H}_S+\hat{H}_E)t}. \tag{2.110}$$

For a density matrix $\hat{\rho}_T(t) : \mathcal{H}_T \to \mathcal{H}_T$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{\rho}_T(t) = -i\alpha[\hat{H}_I(t), \hat{\rho}(t)]. \tag{2.111}$$

Using perturbation theory, this equation can be rewritten, up to second order in α, as:

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{\rho}_T(t) = -i\alpha[\hat{H}_I(t), \hat{\rho}_T(0)] - \alpha^2 \int_0^t [\hat{H}_I(t), [\hat{H}_I(s), \hat{\rho}_T(t)]]ds + \mathcal{O}\left(\alpha^3\right). \tag{2.112}$$

We are only interested in the system density matrix $\hat{\rho}_S$, so we trace out the environment

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{\rho}_S(t) = \mathrm{Tr}_E\left[\frac{\mathrm{d}}{\mathrm{d}t}\hat{\rho}_T(t)\right]. \tag{2.113}$$

However, we have still not removed the dependency on $\hat{\rho}_T(t)$. We need to make a few more assumptions. Firstly, we assume that the total system is in a separable state initially, meaning: $\hat{\rho}_T(0) = \hat{\rho}_S(0) \otimes \hat{\rho}_E(0)$, and that the system is in the ground state initially. This means there are no entanglements or correlations between the system and its environment. This is a sensible assumption if the coupling strength is weak and any correlations which do arise decay on short timescales. Secondly, we assume that the initial state of the environment is thermal, meaning:

$$\hat{\rho}_E(0) = \frac{\exp\left(-\hat{H}_E/(k_B T)\right)}{\mathrm{Tr}\left[\exp\left(-\hat{H}_E/(k_B T)\right)\right]} \tag{2.114}$$

where k_B is the Boltzmann constant. Thirdly, we assume that the system is in the ground state at time $T = 0$, for simplicity. Using these assumptions, it can be shown that the first term of the right hand side of equation (2.112) is zero, leaving us with:

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{\rho}_S(t) = -\alpha^2 \int_0^t \mathrm{Tr}_E[\hat{H}_I(t), [\hat{H}_I(s), \hat{\rho}_T(s)]]ds. \tag{2.115}$$

To continue, we now need to make a rather strong assumption, namely that the system and environment are uncorrelated at all times and that the environment state is always thermal. Meaning that we can write $\hat{\rho}_T(t) = \hat{\rho}_S(t) \otimes \hat{\rho}_E(0)$. This assumption is still reasonable if any correlations between the system and the environment quickly decay, because the environment is far larger

than the system, and the coupling strength α is weak, so the system can't significantly affect the environment. We now have an equation which only depends on the system:

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{\rho}_S(t) = -\alpha^2 \int_0^t \mathrm{Tr}_E[\hat{H}_I(t), [\hat{H}_I(s), \hat{\rho}_S(s) \otimes \hat{\rho}_E(0)]]ds. \tag{2.116}$$

To simplify the above equation, we will use the Markovian approximation $\hat{\rho}_S(s) = \hat{\rho}_S(t)$, making the equation local in time. This approximation is justified because there are no correlations between the system and the environment. By a change of variables $s \to t - s$ and extending the upper limit of integration to ∞, we get the following:

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{\rho}_S(t) = -\alpha^2 \int_0^\infty ds\, \mathrm{Tr}_E[\hat{H}_I(t), [\hat{H}_I(s-t), \hat{\rho}_S(t) \otimes \hat{\rho}_E(0)]], \tag{2.117}$$

which is the Redfield equation [19]. We are justified in taking the integration limit to infinity because it can be shown that the integrand decays to zero exponentially at timescales beyond the correlation time scale.

Using the Redfield equation to evolve density matrices over time, the density matrices do not stay positive semi-definite as required in their definition 2.8.1 in some cases. To ensure that they stay positive semi-definite, we should apply the rotating wave approximation.

First, define the Liouvillian superoperator $\hat{\mathcal{L}}_S$ by

$$\hat{\mathcal{L}}_S \hat{A} := [\hat{H}_S, \hat{A}], \tag{2.118}$$

for all $\hat{A} \in B(\mathcal{H}_S)$ [10].

Let $\{|n\rangle\}$ be a set of eigenvectors of the system Hamiltonian $\hat{H}_S$. Then by the spectral theorem, this forms an orthonormal basis for $\mathcal{H}_S$. In turn, $\{|n\rangle \langle m|\}$ is a basis for $B(\mathcal{H}_S)$. It can then be easily verified that each basis element of $B(\mathcal{H}_S)$ is an eigenvector of the Liouvillian superoperator $\hat{\mathcal{L}}_S$. Remember that we decomposed the interaction Hamiltonian $\hat{H}_I$ in equation (2.109). We can decompose each system term contained in $\hat{H}_I$ into a sum of its eigenvectors.

$$\hat{S}_j = \sum_\omega \hat{S}_j(\omega) \tag{2.119}$$

Which have the eigenvalues

$$[\hat{H}_S, \hat{S}_j(\omega)] = -\omega \hat{S}_j(\omega). \tag{2.120}$$

It can be checked that the Hermitian conjugate $\hat{S}_j^\dagger(\omega)$ has eigenvalue $+\omega$. After some algebraic manipulation, we will encounter oscillatory terms like $e^{i(\omega-\omega')t}$, where we will make the assumption that only resonant ($\omega = \omega'$) terms affect the dynamics at the time scales relevant for the system. After discarding off-resonant terms, we are left with:

[10] Which denotes the set of bounded linear operators on $\mathcal{H}_S$.

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{\rho}_S(t) = \sum_{\omega,k,l}\left[\Gamma_{kl}(\omega)[\hat{S}_l(\omega)\hat{\rho}(t), \hat{S}_k^\dagger(\omega)] + (\Gamma_{lk}(\omega))^*\,[\hat{S}_l(\omega), \hat{\rho}(t)\hat{S}_k^\dagger(\omega)]\right], \tag{2.121}$$

where we use the expressions:

$$\Gamma_{kl}(\omega) := \int_0^\infty e^{i\omega s}\,\mathrm{Tr}_E[\hat{\tilde{E}}_k^\dagger(t)\hat{\tilde{E}}_l(t-s)\hat{\rho}_E(0)]ds, \tag{2.122}$$

$$\hat{\tilde{E}}_l(t) := e^{i\hat{H}_E t}\hat{E}_l e^{-i\hat{H}_E t}. \tag{2.123}$$

We can now divide the $\Gamma_{kl}(\omega)$ factor into a Hermitian and anti-Hermitian part:

$$\Gamma_{kl}(\omega) = \frac{1}{2}\gamma_{kl}(\omega) + i\pi_{kl}(\omega). \tag{2.124}$$

To account for the energy shift of the system due to interactions with the vacuum surrounding it, we also need to add a Lamb shift Hamiltonian term. The Lamb shift was first discovered in the energy levels of hydrogen using quantum electrodynamics, which is beyond the scope of this thesis.

Using quantum electrodynamics, the Lamb shift was later explained to originate from the interactions the electron of the hydrogen atom has with virtual photons, which are energy fluctuations of the vacuum electromagnetic field. These fluctuations are very small and short-lived, so the virtual photons disappear quickly after interacting with the electron.

$$\hat{H}_{Ls} = \sum_{\omega,k,l}\pi_{kl}\hat{S}_k^\dagger(\omega)\hat{S}_l(\omega) \tag{2.125}$$

Combining everything into one equation, we get:

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{\rho}_S(t) = -i[\hat{H} + \hat{H}_{Ls}, \hat{\rho}_S(t)] + \sum_{\omega,k,l}\gamma_{kl}(\omega)\left[\hat{S}_l(\omega)\hat{\rho}_S(t)\hat{S}_k^\dagger(\omega) - \frac{1}{2}\{\hat{S}_k^\dagger(\omega)\hat{S}_l(\omega), \hat{\rho}_S(t)\}\right]. \tag{2.126}$$

It can be shown that the matrix with coefficients $\Gamma_{kl}(\omega)$ is diagonalizable. Written in the diagonal form, we have the Lindblad master equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{\rho}_S(t) = -i[\hat{H} + \hat{H}_{Ls}, \hat{\rho}_S(t)] + \sum_{j,\omega}[\hat{L}_j(\omega)\hat{\rho}(t)\hat{L}_j^\dagger(\omega) - \frac{1}{2}\{\hat{L}_j^\dagger(\omega)\hat{L}_j(\omega), \hat{\rho}_S(t)\}]. \tag{2.127}$$

Where the $\hat{L}_j$ are often called jump operators, they can, for example, represent the decay of an electron from the excited state to the ground state. Because decay from the excited state to the ground state means that the energy of the system is not conserved, it cannot be modelled by the Schrödinger or Heisenberg equations of motion.

2.11 Overview

In this chapter, we have discussed the essential concepts of quantum mechanics, which will be used in this thesis. The foundation of quantum mechanics, of the bra-ket notation, linear operators and postulates has been treated in mathematical detail. After which, we introduced two-level quantum systems and their dynamics. And lastly, we introduced the Lindblad master equation and collapse operators, which will be required for formally treating decay and dephasing processes in quantum systems.

We will continue our mathematical treatment of quantum mechanics by resolving the issues caused by using non-compact operators. In Chapter 3, we will rigorously define position and momentum states and operators, and explain in which kind of spaces they reside.

In Section 4.3, we will discuss the problem of quantum optimal control. In Section 4.4, we will apply this theory to two-level systems and show that the π-pulse discussed in Section 2.6 is the optimal solution to an optimal control problem where you seek to minimise control pulse amplitude whilst under the constraint that the system is fully in the excited state at the end.

In Chapter 5 and 6, quantum optimal control algorithms are tested on two- and three-level systems. The collapse operators and Lindblad equation will also be used in Chapter 7 and 8.

Chapter 3

Rigged Hilbert Spaces and the Problem of non-Compact Operators

In this chapter, we will go into more depth regarding the issues with non-compact operators in quantum mechanics. Although non-compact operators like the position and momentum operators are fundamental to quantum mechanics, most textbooks, even those aimed at mathematicians, do not rigorously treat the case of non-compact operators and the complications arising from them. We will follow [20] by introducing the rigged Hilbert space formalism. Here, 'rigged' is meant in reference to the rigging of a ship; a less awkward term would be equipped Hilbert space.

Since the domain of the position and momentum operators is not the entire Hilbert space, we will take a subset of the Hilbert space to be the physically realisable space Φ, such that you can freely apply the position and momentum operators on states from this space, without producing a state which is not part of the Hilbert space.

The eigenkets of the position and momentum operators are not included in the Hilbert space; therefore, we should have an extension of the Hilbert space, which includes such eigenkets. These eigenkets can be represented as distributions, like the Dirac delta distribution. These eigenkets will turn out to be part of the continuous conjugate-linear dual space of Φ, which we will denote as $\Phi^\times$. The eigenbras will be part of the dual space Φ^*. The original Hilbert space is sandwiched between these two sets:

$$\Phi \subset \mathcal{H} \subset \Phi^\times, \tag{3.1}$$

this is referred to as the rigged Hilbert space formalism[1]. We will first introduce the mathematical concept of measure theory and distributions such that we can describe the eigenkets of non-compact operators.

[1] In a broader context, they are referred to as Gelfand triples [21]

3.1 Measure Theory

To better understand these issues, we should first start by understanding how to integrate elements of Hilbert spaces. To start with, we first introduce the σ-algebra and the concept of a measure. Using this, we can give a rigorous footing to the definition of L^2 spaces, which are one of the fundamental types of infinite-dimensional Hilbert spaces. The textbook of Stroock [22] can be used as a reference for further information.

Definition 3.1.1 (σ-Algebra)**.** *A σ-algebra on a set X is a set Σ containing subsets of X, which is closed under complements, countable unions and countable intersections.*

Definition 3.1.2 (Measure space)**.** *Let X be a set, and Σ a σ-algebra derived from X. The function $\mu : \Sigma \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ is a measure on (X, Σ) if:*

1. *$\mu(\emptyset) = 0$,*
2. *$\mu\left(\bigcup_{k=1}^{\infty} E_k\right) = \sum_{k=1}^{\infty} \mu(E_k)$, if the $\{E_k\}_{k=1}^{\infty}$ are pairwise disjoint.*

The triple (X, Σ, μ) forms a measure space. If for all $S \subset X$, $S \subseteq N \in \Sigma$ and $\mu(N) = 0$ implies that $S \in \Sigma$, then the measure space is complete. The Lebesgue measure λ^n on $\mathbb{R}^n$ is a complete translation invariant measure with $\lambda^n\left([0,1]^n\right) = 1$.

To define integration using the Lebesgue measure, we must first clarify which class of functions can be integrated with respect to a specific measure. These are the so-called measurable functions.

Definition 3.1.3 (Measurable functions)**.** *Let (X, Σ, μ) and (Y, Ξ, λ) be measure spaces. $f : X \to Y$ is measurable if for all $Z \in \Xi$,*

$$f^{-1}(Z) \in \Sigma. \tag{3.2}$$

Let us now move forward in defining Lebesgue integration. We will use the Lebesgue measure on any subset of $\mathbb{R}^n$ unless stated otherwise.

Definition 3.1.4 (Indicator function)**.** *Let $S \subseteq X$ be a measurable set, then its associated indicator function $1_S : X \to \{0, 1\}$ is:*

$$1_S(x) = \begin{cases} 1 & \text{if } x \in S \\ 0 & \text{if } x \notin S \end{cases}. \tag{3.3}$$

The integral of an indicator over a measurable set $Y \subseteq X$ is

$$\int_Y 1_S d\mu = \mu\left(S \cap Y\right). \tag{3.4}$$

This allows us to define simple functions.

Definition 3.1.5 (Simple Functions)**.** *A simple function $\phi : X \to \mathbb{R}$ is a linear combination of indicator functions:*

$$\phi = \sum_{k=1}^{N} a_k 1_{S_k}. \tag{3.5}$$

with $a_k \in \mathbb{R}$ and the sets S_k disjoint and measurable. The integral of a simple function over a measurable set $Y \subseteq X$ is

$$\int_Y \phi d\mu = \sum_{k=1}^{N} a_k \mu\left(S_k \cap Y\right). \tag{3.6}$$

Definition 3.1.6 (Lebesgue integration for non-negative functions)**.** *Let $f : X \to \mathbb{R}$ be a non-negative function. Then we define Lebesgue integration of f over a measurable set $Y \subseteq X$ as follows*

$$\int_{Y,Lebesgue} f d\mu = \sup\left\{\int_Y \phi d\mu : \phi(y) \leq f(y), \forall y \in Y \text{ and } \phi \text{ is simple}\right\}. \tag{3.7}$$

This can be extended to general function $f : X \to \mathbb{R}$ through separating the positive and negative parts of $f = f_+ - f_-$, where $f_\pm$ are both non-negative, assuming that the Lebesgue integral of $f_\pm$ are both finite.

Having now put our understanding of integrals on a firm mathematical footing, we can now properly define L^p spaces.

Definition 3.1.7 (L^p space)**.** *Let (X, Σ, μ) be a measure space and let $1 \leq p < \infty$. Then $L^p(X, \mu)$ is the set of all measurable function $f : X \to \mathbb{C}$ such that*

$$||f||_p := \left(\int_X |f|^p d\mu\right)^{1/p} < \infty. \tag{3.8}$$

We can now show that the Dirac delta distribution does not work as intended if we use the Lebesgue measure. Even ignoring that the Dirac delta distribution doesn't map into $\mathbb{R}$, but rather into $\mathbb{R} \cup \{\infty\}$, through the following proposition.

Proposition 3.1.8. *Let (X, Σ, μ) be a measure space and let $f, g : X \to \mathbb{C}$. If $f \overset{\text{a.e.}}{=} g$, then $\int f d\mu = \int g d\mu$.*

Here, $f \overset{\text{a.e.}}{=} g$ means that the equality $f(x) = g(x)$, with $x \in X$, only fails to hold on a set which have zero measure.

That shows that the Dirac delta distribution would integrate to zero if we were to assume it to be Lebesgue measurable. That means we need to expand our concept of a function by also looking at distributions. Because the Dirac delta distribution integrates to zero, it can't be used as a basis function like we would like to use for L^2 spaces. To address these issues, we will introduce rigged Hilbert spaces, which are Hilbert spaces augmented with a distribution space, which can serve as a kind of generalised basis for L^2.

3.2 Distributions

Distribution theory was developed to formalise the use of objects such as the Dirac delta distribution. Although the Dirac delta is often referred to as a function, it is not well-defined at $x = 0$ when viewed as a function. Instead, the Dirac delta distribution can be defined through its action on a space of test functions.

Let us quickly introduce some helpful notation. Let $K \subseteq X$, then $f \in C^k(K)$ if and only if $f \in C^k(X)$ and $\text{supp}(f) \subseteq K$. Let $C_c^k(X)$ denote the space of functions in $C^k(X)$ with compact support. Functions $f \in C_c^\infty(X)$ are referred to as test functions. Let us first define the convergence of test functions for measure spaces which are subsets of $\mathbb{R}^n$. This section is based on [23].

Definition 3.2.1 (Convergence of a sequence of test functions)**.** *Let X an open subset of $\mathbb{R}^n$. The sequence of test functions $\{\phi_j\}_{j\in\mathbb{N}} \subset C_c^\infty(X)$, is said to converge to $\phi \in C_c^\infty(X)$ as $j \to \infty$:*

$$\lim_{j\to\infty} \phi_j = \phi \in C_c^\infty(X) \tag{3.9}$$

if the following two conditions hold:

- *There exists a compact subset K of X such that $supp(\phi_j) \subset K$ for all $j \in \mathbb{N}$.*
- *For every multi-index $\vec{\beta} \in \mathbb{N}^n$, the sequence $\{D^{\vec{\beta}}\phi_j\}_{j\in\mathbb{N}}$ converges uniformly to $\partial^{\vec{\beta}}\phi$ on X, using the multi-index notation:*

$$D^{\vec{\beta}} := \partial_1^{\beta_1}\partial_2^{\beta_2}\cdots\partial_n^{\beta_n}. \tag{3.10}$$

Definition 3.2.2 (Distributions)**.** *A distributions on X is a linear functional $u : C_c^\infty(X) \to \mathbb{C}$, which is continuous in the sense that:*

$$\lim_{j\to\infty} u(\phi_j) = u(\phi), \tag{3.11}$$

for any sequence of test functions $\{\phi_j\}_{j\in\mathbb{N}} \subset C_c^\infty(X)$, which converges to ϕ on X.

The Schwarz space $\mathcal{S}(\mathbb{R}^n)$ is the function space of all functions whose derivatives decrease rapidly as they approach infinity, they will form an important component of the rigged Hilbert space.

Definition 3.2.3 (Schwarz space)**.** *Let $\mathbb{N}^n$ be the n-fold Cartesian product of $\mathbb{N}$. Then the Schwartz space is defined as:*

$$\mathcal{S}(\mathbb{R}^n) := \left\{ f \in C^\infty(\mathbb{R}^n) |\ \forall \vec{\alpha}, \vec{\beta} \in \mathbb{N}^n,\ ||f||_{\vec{\alpha},\vec{\beta}} < \infty \right\}. \tag{3.12}$$

With the semi-norm[2] $||\cdot||_{\vec{\alpha},\vec{\beta}}$:

$$||f||_{\vec{\alpha},\vec{\beta}} := sup_{x\in\mathbb{R}^n} \left| x^{\vec{\alpha}} \left(D^{\vec{\beta}} f \right)(x) \right|. \tag{3.13}$$

[2] A semi-norm is a norm which doesn't need to be positive definite, however, it is still non-negative.

using the multi-index notation:

$$x^{\vec{\alpha}} := x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}, \tag{3.14}$$

$$D^{\vec{\beta}} := \partial_1^{\beta_1} \partial_2^{\beta_2} \cdots \partial_n^{\beta_n}, \tag{3.15}$$

Its dual $\mathcal{S}^*$ is called the space of tempered distributions and can indeed be considered a subspace of distributions.

The space of test functions is a subset of the Schwartz space since the Schwartz space does not require functions to have compact support. The Fourier transform is an automorphism on Schwarz space [24][Theorem 3.3]. The Dirac delta distribution is the distribution δ which is defined as $\delta(f) = f(0)$ for any test function $f \in C^\infty(\mathbb{R}^n)$.

It can also be shown that $\mathcal{S}(\mathbb{R}^n)$ is dense in the Hilbert space $\mathcal{H} = L^2(\mathbb{R}^n)$. In Section 3.4, we will show that for $\mathcal{H} = L^2(\mathbb{R}^n)$, the Schwartz space $\mathcal{S}(\mathbb{R}^n)$ is the space of physically realisable states Φ and its conjugate-linear dual is the space $\Phi^\times$ containing all eigenvectors of non-compact operators.

3.3 Coordinate and Momentum Representations of Quantum states

Until now, we have not given physical meaning to specific quantum states, such as their location in space. To enable us to make such a measurement, we will introduce the coordinate operator $\hat{x}$, which will be a non-compact operator with a continuous eigenvalue spectrum, meaning we can't carry over all our results for compact operators such as the spectral Theorem 2.2.9. One key point is that the eigenkets of the position and momentum operators will not be included in the Hilbert space $\mathcal{H}$, but they will be included in $\Phi^\times$ in the rigged Hilbert space formalism. A good reference for more details on the position and momentum operators is [5].

Take as the Hilbert space $\mathcal{H} = L^2(\mathbb{R}^n)$, which is equipped with the Lebesgue measure on $\mathbb{R}^n$. Consider the position bra $\langle x_0|$ to be defined as the Dirac delta distribution centred at $x_0 \in \mathbb{R}^n$:

$$\langle x_0|\psi\rangle = |\psi\rangle\,(x_0), \tag{3.16}$$

since $|\psi\rangle \in L^2(\mathbb{R}^n)$, it is a function from $\mathbb{R}^n$ to $\mathbb{C}$, and we can thus assign a scalar to $|\psi\rangle\,(x_0)$. The corresponding ket $|x_0\rangle$ will act like a conjugate-linear Dirac delta distribution by setting:

$$\langle \psi|x\rangle := (\langle x|\psi\rangle)^*, \quad \forall\, |\psi\rangle \in \mathcal{H}. \tag{3.17}$$

The position bras $\langle x_0|$ and kets $|x_0\rangle$ are not included in the Hilbert space since they are distributions. Instead, they are part of the linear dual space Φ^* and the conjugate-linear dual space $\Phi^\times$ respectively.

The set of all position kets: $\{|x\rangle : x \in \mathbb{R}^n\}$ forms a complete (uncountable) orthonormal basis by the Gelfand-Maurin Theorem [20], which we will not discuss in further detail here. This means that we can write any state $|\psi\rangle$ as:

$$|\psi\rangle = \int \psi(x)\,|x\rangle\,d^n x, \tag{3.18}$$

with $\psi : \mathbb{R}^n \to \mathbb{C}$ often referred to as the wave function. Since $|\psi\rangle \in L^2(\mathbb{R}^n)$, we can easily translate it to a wave function by setting: $\psi(x) := \langle x|\psi\rangle$. The wave function $\psi : \mathbb{R}^n \to \mathbb{C}$ and the ket $|\psi\rangle \in \mathcal{H}$ are the same objects; however, the wave function is primarily viewed as a function and the ket is primarily viewed as a more abstract Hilbert space element. It depends on the context which perspective is more beneficial to take.

If the integration domain is not specified, it is assumed to be over the entire range of the integration variable. The probability density that the state is localised at a given point x in space after measurement is given by: $|\psi(x)|^2 dx$, as discussed in Postulate 2.3.2.

The position operator $\hat{x}_j : D(\hat{x}_j) \to \mathcal{H}$, corresponding to the jth basis vector of $\mathbb{R}^n$ is defined as the operator:

$$\hat{x}_j := \int x_j\,|x\rangle\,\langle x|\,d^n x, \tag{3.19}$$

where x_j is the jth coordinate of $x \in \mathbb{R}^n$. We can now show that the position kets $|x_0\rangle$ for $x_0 \in \mathbb{R}^n$ are eigenkets of the position operator $\hat{x}_j$:

$$\begin{aligned}
\hat{x}_j\,|x_0\rangle &= \int x_j\,|x\rangle\,\langle x|\,|x_0\rangle\,d^n x && (3.20)\\
&= \int x_j\,|x\rangle\,\delta(x - x_0) d^n x && (3.21)\\
&= (x_0)_j\,|x_0\rangle\,. && (3.22)
\end{aligned}$$

its associated eigenvalue is thus $(x_0)_j$.
When we apply the position operator $\hat{x}_j$ to an arbitrary state $|\psi\rangle$, we get:

$$\begin{aligned}
\hat{x}_j\,|\psi\rangle &= \int x_j\,|x\rangle\,\langle x|\psi\rangle\,d^n x, && (3.23)\\
&= \iint x_j\psi(y)\,|x\rangle\,\langle x|y\rangle\,d^n x d^n y, && (3.24)\\
&= \int x_j\psi(x)\,|x\rangle\,d^n x, && (3.25)
\end{aligned}$$

where we applied the Gelfand-Maurin theorem in the second line.

The domain of the operator $\hat{x}_j$ is $D(\hat{x}_j) = \{|\psi\rangle : \hat{x}_j\,|\psi\rangle \in \mathcal{H}\}$, which corresponds to the condition that the function $g(x) = x_j\psi(x)$ is square integrable. This condition does not always hold for every $\psi \in \mathcal{H}$. Take for example the wave function $\psi \in L^2(\mathbb{R})$, with $\psi(x) = \frac{1}{\sqrt{x^2+1}}$, which is square

integrable. However, $x\psi(x) = \frac{x}{\sqrt{x^2+1}}$ is obviously not.

To prevent issues, where $(\hat{x}_j)^k |\psi\rangle$ is no longer square integrable for some $k \in \mathbb{N}^+$, we should only accept states which decay exponentially to zero as $x \to \infty$. Later, we will show that it is the Schwartz space $\mathcal{S}(\mathbb{R}^n)$ which satisfies these requirements.

Similarly to $\{|x\rangle : x \in \mathbb{R}^n\}$, the momentum states, $\{|p\rangle : p \in \mathbb{R}^n\}$, also form a complete orthonormal basis by the same Gelfand-Maurin theorem. Momentum kets have a single well-defined momentum. In quantum mechanics, such objects correspond to plane waves, with a wavelength of $\frac{2\pi\hbar}{|p|}$. Through the Fourier transform, they can be defined as:

$$|p\rangle = \int e^{\frac{i}{\hbar}p\cdot x} |x\rangle \, d^n x, \tag{3.26}$$

where $\hbar$ is there as a scaling factor, to ensure that the exponent is dimensionless. In physics, it is referred to as the reduced Planck constant. The inverse Fourier transform is given as:

$$|x\rangle = \frac{1}{(2\pi\hbar)^n} \int e^{-\frac{i}{\hbar}p\cdot x} |p\rangle \, d^n p. \tag{3.27}$$

The factor $\frac{1}{(2\pi\hbar)^n}$ is a choice of convention for the definition of the Fourier transform; other references might use a factor of $\frac{1}{(2\pi\hbar)^{n/2}}$ in front of both equations.

The expressions for the inner product in coordinate and momentum representation are, respectively:

$$\langle\phi|\psi\rangle = \int \langle\phi|x\rangle \langle x|\psi\rangle \, d^n x, \tag{3.28}$$

$$\langle\phi|\psi\rangle = \frac{1}{(2\pi\hbar)^n} \int \langle\phi|p\rangle \langle p|\psi\rangle \, d^n p, \tag{3.29}$$

and they are equal by the Parseval–Plancherel identity [25]. The factor $\frac{1}{(2\pi\hbar)^n}$ comes from our chosen Fourier transform convention. Similarly, we have the following expressions for the identity operator $\hat{I}$:

$$\hat{I} = \int |x\rangle \langle x| \, d^n x, \tag{3.30}$$

$$\hat{I} = \frac{1}{(2\pi\hbar)^n} \int |p\rangle \langle p| \, d^n p. \tag{3.31}$$

Because both $|x\rangle$ and $|p\rangle$ form a complete basis on the Hilbert space, we can compute the inner product of $|x\rangle$ and $|p\rangle$ as follows:

$$\langle x|p\rangle = \langle x| \left(\int e^{ip\cdot x'/\hbar} |x'\rangle \, d^n x' \right) = e^{ip\cdot x/\hbar}. \tag{3.32}$$

Let's have a look at the momentum operators $\hat{p}_j$, for $i = 1, \cdots, n$, which are defined on the momentum basis as: $\hat{p}_j |p'\rangle = p'_j |p'\rangle$. We would also like to find the momentum operator in the coordinate representation, which can be deduced using the Fourier transform between $|x\rangle$ and $|p\rangle$:

$$\begin{aligned}
\langle x|\hat{p}_j|\psi\rangle &= \frac{1}{(2\pi\hbar)^n} \int \langle x|\hat{p}_j|p\rangle \langle p|\psi\rangle \, d^n p && (3.33)\\
&= \frac{1}{(2\pi\hbar)^n} \int p_j \langle x|p\rangle \langle p|\psi\rangle \, d^n p && (3.34)\\
&= \frac{1}{(2\pi\hbar)^n} \int p_j e^{ip\cdot x/\hbar} \langle p|\psi\rangle \, d^n p && (3.35)\\
&= \frac{\hbar}{i} \frac{1}{(2\pi\hbar)^n} \frac{\mathrm{d}}{\mathrm{d}x_j} \int e^{ip\cdot x/\hbar} \langle p|\psi\rangle \, d^n p && (3.36)\\
&= \frac{\hbar}{i} \frac{\mathrm{d}}{\mathrm{d}x_j} \langle x|\psi\rangle && (3.37)\\
&= \frac{\hbar}{i} \frac{\mathrm{d}}{\mathrm{d}x_j} \psi(x). && (3.38)
\end{aligned}$$

Where in the third line, we applied the relation: $p_j e^{ip\cdot x/\hbar} = \frac{\hbar}{i}\frac{\mathrm{d}}{\mathrm{d}x_j} e^{ip\cdot x/\hbar}$. This means we can also write the momentum operator using the derivative: $\hat{p}_j = \frac{\hbar}{i}\frac{\mathrm{d}}{\mathrm{d}x_j}$. The momentum operator is thus ill-defined for states whose wave functions in position space are not smooth. Therefore, we want to ensure that all wave functions are smooth. The Schwartz space $\mathcal{S}(\mathbb{R}^n)$ satisfies this property and is therefore still our candidate for the space of physically realisable states Φ.

3.4 Rigged Hilbert space

We are now finally at the point where we can define the rigged Hilbert space. As discussed before at the start of this chapter, it will be a triple of function spaces, which are the original Hilbert space $\mathcal{H}$, which we assume to be $L^2(\mathbb{R}^d)$ for simplicity in this section, the space of physically realizable states Φ, and its conjugate-linear dual, which will be a space of distributions which is a superset of $\mathcal{H}$ and contains the eigenvectors of unbounded operators such at the position and momentum operators. The rigged Hilbert space is an example of a Gelfand triple from functional analysis, which is further discussed in [26]. In this section, we largely follow [20]; however, the proof of Lemma 3.4.2 is my own. The Hilbert space is sandwiched between Φ and the conjugate-linear dual $\Phi^\times$:

$$\Phi \subset \mathcal{H} \subset \Phi^\times. \tag{3.39}$$

In the following section, the Hamiltonian operator $\hat{H}$ is used to define the Rigged Hilbert space. The Hamiltonian operator can measure the energy of a given quantum state and, through the Schrödinger equation, will generate the time evolution of the system. We will discuss it in more detail in Subsection 2.4.

Definition 3.4.1. *Define Φ to be the largest set that remains invariant under the application of any of the position operators $\hat{x}_j$, momentum operators $\hat{p}_j$ or the Hamiltonian $\hat{H}$ associated with the quantum system. Equivalently, this can be written down as the infinite intersection:*

$$\Phi = \bigcap_{k=0}^{\infty} \left\{ D\left(\prod_{j=1}^{k} \hat{X}_j \right) \;\middle|\; \forall j \leq k, \hat{X}_j \in \{\hat{x}_1, \cdots, \hat{x}_n, \hat{p}_1, \cdots, \hat{p}_n, \hat{H}\} \right\}, \tag{3.40}$$

where the domain of an operator is:

$$D(\hat{A}) := \left\{ |\psi\rangle \in L^2(\mathbb{R}^n) \middle| \langle\psi|\hat{A}^2|\psi\rangle < \infty \right\}. \tag{3.41}$$

You can view $\Phi \subset \mathcal{H}$ as the set of physically realisable states since for all states in Φ, all measurable quantities can be associated with well-defined values. All states are also associated with a smooth continuous function, whereas in $\mathcal{H}$, states can be discontinuous and are only required to be Lebesgue integrable.

Consider for example $f(x) = e^{-x}\theta(x)$ on $\mathcal{H} = L^2(\mathbb{R})$, where $\theta(x)$ is the Heaviside function[3]. The function $f : \mathbb{R} \to \mathbb{R}$ is normalizable, however, $\hat{p}f(x) \notin L^2(\mathbb{R}^n)$ because $f'(0)$ is not well-defined, where we use the expression $\hat{p} = \frac{\hbar}{i}\frac{\mathrm{d}}{\mathrm{d}x}$ for the momentum operator. In general, the requirement that $(\hat{p})^m f \in L^2(\mathbb{R}^n)$ for any $m \in \mathbb{N}$ ensures that f is smooth.

Lemma 3.4.2. *If $\hat{H}$ can be written as a multivariate polynomial of $\hat{x}_j, \hat{p}_k$ with complex coefficients, meaning that it can be written as:*

$$\hat{H} = \sum_{\vec{\alpha},\vec{\beta}} c_{\vec{\alpha},\vec{\beta}}\ \hat{e}_{\vec{\alpha},\vec{\beta}}. \tag{3.42}$$

With a finite number of $c_{\vec{\alpha},\vec{\beta}} \in \mathbb{C}$ which are non-zero. Here we borrow the multi-index notation of definition 3.2.3, and the basis operators $\hat{e}_{\vec{\alpha},\vec{\beta}}$ are defined as:

$$\hat{e}_{\vec{\alpha},\vec{\beta}} f = \hat{x}^{\vec{\alpha}}\hat{p}^{\vec{\beta}} f, \tag{3.43}$$

then Φ is the Schwartz space $\mathcal{S}(\mathbb{R}^n)$.

Proof. If $f \in \Phi$, then $f \in \mathcal{D}(\hat{x}^{\vec{\alpha}}\hat{p}^{\vec{\beta}})$ for any choice of $\vec{\alpha}, \vec{\beta}$, therefore, $f \in \mathcal{S}(\mathbb{R}^n)$. If $f \in \mathcal{S}(\mathbb{R}^n)$, then we have that $\hat{x}f(x), \hat{p}f(x)$ are also in $\mathcal{S}(\mathbb{R}^n)$. For the operator $\hat{H}$, we can write:

$$\left|\left|\hat{H}f\right|\right|_{\vec{\gamma},\vec{\delta}} = \left|\left| \sum_{\vec{\alpha},\vec{\beta}} c_{\vec{\alpha},\vec{\beta}} \hat{e}_{\vec{\alpha},\vec{\beta}} f \right|\right|_{\vec{\gamma},\vec{\delta}} \leq \sum_{\vec{\alpha},\vec{\beta}} \left|\left| c_{\vec{\alpha},\vec{\beta}} \hat{e}_{\vec{\alpha},\vec{\beta}} f \right|\right|_{\vec{\gamma},\vec{\delta}}. \tag{3.44}$$

Since every $e_{\vec{\alpha},\vec{\beta}} f \in \mathcal{S}(\mathbb{R}^n)$, its norm is bounded and because there are a finite number of summands, we can conclude that $||\hat{H}f||_{\vec{\gamma},\vec{\delta}} < \infty$, for any $\vec{\gamma}, \vec{\delta}$.

[3] $\theta(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$

Therefore, $\hat{H}f \in \mathcal{S}(\mathbb{R}^n)$ and $\mathcal{S}(\mathbb{R}^n)$ is invariant under application of $\hat{x}, \hat{p}, \hat{H}$. Note that if you set $\hat{H} = 0$, then $\Phi = \mathcal{S}(\mathbb{R}^n)$, by definition. Therefore, for general $\hat{H}$, $\Phi \subseteq \mathcal{S}(\mathbb{R}^n)$.

Since Φ is defined to be the largest invariant set under application of $\hat{x}, \hat{p}$ and $\hat{H}$, we must have that $\Phi = \mathcal{S}(\mathbb{R}^n)$.

□

The eigenbras are a part of Φ^* and the eigenkets are a part of $\Phi^\times$. If $\Phi = \mathcal{S}(\mathbb{R}^d)$, then Φ^* is the space of tempered distributions, and $\Phi^\times$ is the space of conjugate-linear tempered distributions, and we finally know which spaces constitute the rigged Hilbert space formalism.

When an operator $\hat{A} : \mathcal{H} \to \mathcal{H}$ is compact, it is defined on the entirety of the Hilbert space, and we do not have to employ the rigged Hilbert space formalism. However, when it is non-compact and can be written as a multivariate polynomial of position and momentum operators, like in Lemma 3.4.2, then we have to use the rigged Hilbert spaces formalism. In that case, we can still use the measurements Postulates 2.3.1 and 2.3.2, additional condition that the quantum states $|\psi\rangle$ are part of Φ and their eigenkets of the operator $\hat{A}$ are part of $\Phi^\times$.

3.5 Generalized Uncertainty Principle

One of the most fundamental features of quantum mechanics is the uncertainty principle, which states that certain pairs of observables cannot be simultaneously measured with full accuracy. For the Hilbert space $\mathcal{H} = L^2(\mathbb{R}^n)$, one such pair is position-momentum, for which the uncertainty principle asserts:

$$\sigma_{x_j}\sigma_{p_k} \geq \frac{\hbar}{2}\delta_{jk}, \tag{3.45}$$

where x_j and p_k represent the position and momentum along the jth and kth basis vector of $\mathbb{R}^n$, and σ_{x_j} and σ_{p_k} their standard deviations, δ_{jk} is the Kronecker delta[4]. This means that in quantum mechanics, the phase-space formalism of classical mechanics runs into trouble such we can't know a particle's position and momentum at the same time.

The standard deviation of an arbitrary operator $\hat{A}$ is defined as:

$$\sigma_A^2 = \langle\psi| \left(\hat{A} - \langle\hat{A}\rangle\right)^2 |\psi\rangle\rangle\,, \tag{3.46}$$

where $\langle\hat{A}\rangle := \langle\psi|\hat{A}|\psi\rangle$ is the expected value of $\hat{A}$. The wave functions $\psi(x)$ and $\tilde{\psi}(p)$ in the position and momentum space are Fourier transforms of each other, if the wavefunction $\psi(x)$ has a sharp spike, then $\tilde{\psi}(p)$ will naturally be more spread out, and the variance of our momentum measurements increases.

We will now briefly discuss the mathematical formalism behind the uncertainty principle for arbitrary Hermitian operators.

[4]It is defined such that $\delta_{jk} = 1$ if $j = k$, else $\delta_{jk} = 0$.

Definition 3.5.1 (Compatible operators)**.** *The pair of operators $\hat{A}, \hat{B}$ are called compatible if*

$$[\hat{A}, \hat{B}] := \hat{A}\hat{B} - \hat{B}\hat{A} = \hat{0} \tag{3.47}$$

Where $\hat{0}$ is the zero operator such that $\hat{0}\,|\psi\rangle = 0$ for all $|\psi\rangle \in \mathcal{H}$. Notice that any compatible pair of operators commute with each other: $\hat{A}\hat{B} = \hat{B}\hat{A}$, therefore, the bracket $[\hat{A}, \hat{B}]$ is often called the commutator. For a compatible pair of operators $\hat{A}, \hat{B}$, you can find a complete set of states which are eigenstates of both $\hat{A}$ and $\hat{B}$. Non-commutativity imposes a lower bound on the product of the standard deviations $\sigma_A \sigma_B$,

$$\sigma_A \sigma_B \geq \frac{\hbar}{2} |[\hat{A}, \hat{B}]|, \tag{3.48}$$

this the uncertainty principle for arbitrary pairs of operators. If we apply it to the position and momentum operator, we get:

$$[\hat{x}_j, \hat{p}_k]\,|\psi\rangle = \frac{i}{\hbar}\left[\hat{x}_j \frac{\mathrm{d}}{\mathrm{d}x_k}\,|\psi\rangle - \frac{\mathrm{d}}{\mathrm{d}x_k}\left(\hat{x}_j\,|\psi\rangle\right)\right] \tag{3.49}$$

$$= \frac{i}{\hbar}\int \left[x'_j \frac{\mathrm{d}}{\mathrm{d}x_k}\psi(x')\,|x'\rangle - \frac{\mathrm{d}}{\mathrm{d}x_k}\left(x'_j \psi(x')\,|x'\rangle\right)\right] dx' \tag{3.50}$$

$$= \frac{i}{\hbar}\int \left[x'_j \partial_k \psi(x')\,|x'\rangle - x'_j \partial_k \psi(x')\,|x'\rangle - \delta_{jk}\psi(x')\,|\psi\rangle\right] dx' \tag{3.51}$$

$$= \frac{\hbar}{i}\,|\psi\rangle\,. \tag{3.52}$$

Showing that equation (3.45) does hold indeed.

Chapter 4

Introduction to Optimal Control Theory

In this chapter, we will briefly introduce the basics of control theory before moving on to quantum control theory, which will use the framework with some adjustments because the Schrödinger equation is a complex PDE instead of a real PDE and because we will start working with the time-evolution of wave functions instead of the time-evolution of a (system of) particles.

We will be working with the modern formulation of (optimal) control theory, where we reduce the higher-order ODEs which are found in the formulation of the problem into a linear system of coupled first-order ODEs, which we will then solve and investigate using linear algebra. In this section, we will follow the book by van der Schaft and Meinsma [27].

4.1 Calculus of variations

We will start by introducing calculus of variations, which aims to optimise a function and is a simplified version of control theory, where we optimise over the outcomes stemming from a differential equation. Calculus of variations deals with the minimisation of expressions of the form:

$$\int_0^T F\big(t, x(t), \dot{x}(t)\big)dt, \tag{4.1}$$

for all functions $x : [0, T] \to \mathbb{R}^n$ which are sufficiently smooth. Where $F : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is some given function. Here $\dot{x}(t) := \frac{\mathrm{d}x}{\mathrm{d}t}(t)$ denotes the time derivative of the function. Calculus of variations has deep connections to physics through Hamilton's principle of least action, which states that a particle follows the path that minimises its action. The principle of least action allows you to derive the Euler-Lagrange equations of motion. We will start by discussing calculus of variations in its simplest form.

Definition 4.1.1 (Simplest problem in the calculus of variations)**.** *Given a final time $T > 0$ and a function $F : [0, T] \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, with $x_0, x_T \in \mathbb{R}^n$, the simplest problem in the calculus of*

variations is to minimise the cost functional J defined as:

$$J[x] = \int_0^T F\big(t, x(t), \dot{x}(t)\big)dt, \tag{4.2}$$

over all function $x : [0, T] \to \mathbb{R}^n$ which are sufficiently smooth that satisfy the boundary conditions $x(0) = x_0$ and $x(T) = x_T$.

The functional J is called the cost functional, and is not called a function, because it takes a function as its argument and then applies some operations like integration on it. The associated integrand function F is the Lagrangian. When necessary, it will be specified what it means to be sufficiently smooth; often, a function being $\mathcal{C}^1$ or $\mathcal{C}^2$ suffices. A function f being $\mathcal{C}^n$ means that it and its derivatives up to and including the nth order exist and are continuous.

The famous Euler-Lagrange equations give a necessary condition for a function to minimise the cost function.

Theorem 4.1.2 (Euler-Lagrange equation, necessary condition for optimality)**.** *Suppose that the Lagrangian F is $\mathcal{C}^1$. To solve the problem stated in Definition 4.1.1, it is a necessary condition for a $\mathcal{C}^1$ function $x_*(t)$ to satisfy the differential equation:*

$$\left(\frac{\partial}{\partial x} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial}{\partial \dot{x}}\right) F\big(t, x_*(t), \dot{x}_*(t)\big) = 0, \quad \forall t \in [0, T]. \tag{4.3}$$

Here, $\frac{\partial}{\partial x}$ represents a column vector of partial derivatives:

$$\begin{bmatrix} \frac{\partial}{\partial x_1} \\ \vdots \\ \frac{\partial}{\partial x} \end{bmatrix}, \tag{4.4}$$

this means that equation (4.3) represents n independent differential equations. In many applications, you will come across time-independent Lagrangians. For such Lagrangians, the Euler-Lagrange equations can be reduced to the Beltrami identity if $\dot{x}(t)$ is non-zero almost everywhere on its domain in the measure theoretic sense.:

$$F(x, \dot{x}) - \dot{x}^T \frac{\partial F}{\partial \dot{x}} = C_0, \tag{4.5}$$

where $C_0 \in \mathbb{R}$ is a constant determined by the boundary conditions. The term $\dot{x}^T$, transposes $\dot{x}$ into a row vector, which when multiplied by a column vector, turns into an inner product.

In quantum optimal control theory, solutions are constrained by differential equations like the Schrödinger equation. In calculus of variations, the equivalent of those constraints are integral constraints of the form:

$$\int_0^T M\big(t, x(t), \dot{x}(t)\big)dt = C_0, \tag{4.6}$$

with $M : [0,T] \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is some given function. Such problems can be solved using Lagrange multipliers, which we will now briefly explain.

Consider minimising a function $J : \mathbb{R}^n \to \mathbb{R}$ with the constraint that the solution is part of the kernel of $G : \mathbb{R}^n \to \mathbb{R}^k$:

$$\begin{cases} \min_{z \in \mathbb{R}^n} J(z), \\ \text{subject to } G(z) = 0. \end{cases} \tag{4.7}$$

The core idea behind the Lagrange multiplier method is to translate the constrained problem to an unconstrained problem with optimisation variables (z, λ), with $\lambda \in \mathbb{R}^k$, which has the cost function:

$$\mathcal{J}(z, \lambda) := \lambda^T G(z) + J(z). \tag{4.8}$$

Under the assumption that $\mathcal{J}$ is sufficiently smooth, the pair (z_*, λ_*) is a stationary solution of the unconstrained optimisation problem if and only if:

$$\frac{\partial \mathcal{J}}{\partial z} = 0, \quad \frac{\partial \mathcal{J}}{\partial \lambda} = 0, \tag{4.9}$$

which implies that $G(z_*) = 0$ and therefore, $\mathcal{J}(z_*, \lambda_*) = J(z_*)$. Under some mild assumptions, the first-order conditions on the unconstrained problems are equivalent to the conditions on the constrained problem, and we have thus found a solution to the constrained problem.

Theorem 4.1.3 (Euler-Lagrange equation for integral constrained optimisation)**.** *Suppose $F, M : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ are $\mathcal{C}^1$ functions and $x_* : \mathbb{R} \to \mathbb{R}$ is a $\mathcal{C}^2$ function, which is the minimizer of:*

$$J[x] = \int_0^T F\big(t, x(t), \dot{x}(t)\big) dt, \tag{4.10}$$

subject to the boundary conditions $x(0) = x_0; x(T) = x_T$, and the integral constraint:

$$\int_0^T M\Big(t, x(t), \dot{x}(t)\Big) dt = C_0, \quad \textit{with } C_0 \in \mathbb{R}. \tag{4.11}$$

Then either there exists a Lagrange multiplier $\mu_* \in \mathbb{R}$ such that

$$\left(\frac{\partial}{\partial x} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial}{\partial \dot{x}}\right)\Big(F\big(t, x_*(t), \dot{x}_*(t)\big) + \mu_* M\big(t, x_*(t), \dot{x}_*(t)\big)\Big) = 0, \tag{4.12}$$

for all $t \in [0,T]$ or M satisfies the Euler-Lagrange equations itself,

$$\left(\frac{\partial}{\partial x} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial}{\partial \dot{x}}\right) M\Big(t, x_*(t), \dot{x}_*(t)\Big) = 0, \quad \forall\ t \in [0,T]. \tag{4.13}$$

4.2 Optimal Control Theory

In calculus of variations, you are allowed to choose any function $x(t)$ as long as it is sufficiently smooth. However, in physical reality, systems always follow physical laws, and our control over them might not be complete. To take such constraints in account, we assume that the state $x : [0, T] \to \mathbb{R}^n$ follows a differential equation.

Definition 4.2.1 (Problem of Optimal control)**.** *Given a final time $T > 0$. The state function $x : [0, T] \to \mathbb{R}^n$ evolves according to*

$$\dot{x}(t) = f(x(t), u(t)), \quad x(0) = x_0, \tag{4.14}$$

with $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ determining the dynamics of the system and $u : [0, T] \to \mathbb{U} \subseteq \mathbb{R}^m$ being the control functions, which we are free to control. We aim to minimise the cost functional

$$J[u(\cdot)] := \int_0^T F(x(t), u(t))dt + G(x(T)), \tag{4.15}$$

whilst ensuring that the state satisfies equation (4.14)*. We refer to $G : \mathbb{R}^n \to \mathbb{R}$ as the final cost and $F : \mathbb{R}^n \times \mathbb{U} \to \mathbb{R}$ as the running cost.*

To ensure that equation (4.14) is satisfied, we can introduce the Lagrange multiplier function $p : [0, T] \to \mathbb{R}^n$. Let $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{U} \times \mathbb{R}^n \to \mathbb{R}$ be the running cost augmented with the Lagrange multiplier function p:

$$\mathcal{L}(x, \dot{x}, u, p) = p^T \Big(f(x, u) - \dot{x} \Big) + F(x, u). \tag{4.16}$$

The Lagrange multiplier function $p : [0, T] \to \mathbb{R}^n$ is created in analogy with the Lagrange multipliers used in calculus of variations in the previous section. We want to further investigate the stationary solutions (x_*, p_*, u_*) of the optimal problem, where the Lagrange multiplier is incorporated through the total cost:

$$\mathcal{J}[x(\cdot), p(\cdot), u(\cdot)] := \int_0^T \mathcal{L}(x(t), \dot{x}(t), u(t), p(t))dt + G(x(T)). \tag{4.17}$$

It can be shown through Beltrami's identity, that the so-called optimal control Hamiltonian $H : \mathbb{R}^n \times \mathbb{R}^n \times \mathcal{U} \to \mathbb{R}$,

$$H(x(t), p(t), u(t)) = p(t)^T f(x(t), u(t)) + F(x(t), u(t)), \tag{4.18}$$

is constant along any stationary trajectory $(x_*(\cdot), p_*(\cdot), u_*(\cdot))$. The optimal control Hamiltonian is also often referred to as the Pontryagin Hamiltonian. Here, p is the Hamiltonian co-state of x, in analogy to momentum being the co-state to position in Hamilton's equations of motion for classical mechanics, meaning that the pair (x, p) satisfies Hamilton's equations. This can be shown using Pontryagin's Minimum Principle[1], a core result of optimal control theory.

[1]It is often called the maximum or minimum principle in literature, depending on the sign convention used for the Hamiltonian. To avoid confusion, we call it Pontryagin's Minimum Principle.

Theorem 4.2.2 (Pontryagin's Minimum Principle)**.** *Let $\mathbb{U} = \mathbb{R}^m, x_0 \in \mathbb{R}^n$ and consider $f(x,u)$, $\frac{\partial f}{\partial x}(x,u)$, $F(x,u)$, $\frac{\partial F}{\partial x}(x,u)$, $G(x)$ and $G'(x)$ to be all continuous. Suppose $u_* : [0,T] \to \mathcal{U}$ is a solution of the optimal control problem and assume it is piecewise continuous. Let $x_* : [0,T] \to \mathbb{R}^n$ be the resulting optimal state. Then there exists an unique $p_* : [0,T] \to \mathbb{R}^n$ such that:*

$$\dot{x}_*(t) = \frac{\partial H}{\partial p}\Big(x_*(t), p_*(t), u_*(t)\Big), \quad x_*(0) = x_0, \tag{4.19}$$

$$\dot{p}_*(t) = -\frac{\partial H}{\partial x}(x_*(t), p_*(t), u_*(t)), \quad p_*(T) = \frac{\partial G}{\partial x}(x_*(T)), \tag{4.20}$$

$$H\Big(x_*(t), p_*(t), u_*(t)\Big) = min_{u\in\mathcal{U}} H\Big(x_*(t), p_*(t), u\Big), \quad \forall t \in [0,T], \tag{4.21}$$

where $u_*(t)$ is continuous. The optimal control Hamiltonian is constant along optimal trajectories:

$$H(x_*(t), p_*(t), u_*(t)) = C_0 \in \mathbb{R}, \quad \forall\ t \in [0,T]. \tag{4.22}$$

Pontryagin's minimum principle allows you to restate optimal control problems into a form which is easier to solve analytically or numerically. We will now move to apply the Pontryagin's minimum principle to quantum systems, which are governed by the Schrödinger equation.

4.3 Quantum Optimal Control

We have now finally arrived at the point where we are able to begin discussing quantum optimal control. In this section, we will follow [28]. We will be working with Hilbert spaces of the form $\mathcal{H} = \mathbb{C}^n$, unless otherwise specified. Throughout the thesis, we will consider Hamiltonians $\hat{H} : \mathcal{U} \to (\mathcal{H} \to \mathcal{H})$ of the form:

$$\hat{H}(u(t)) = \hat{H}_0 + \sum_{k=1}^{N_c} u_k(t)\hat{H}_k, \tag{4.23}$$

where $\hat{H}_0 : \mathcal{H} \to \mathcal{H}$ is the so-called drift Hamiltonian and represents the system's time dynamics with no control applied. For $k = 1, \cdots N_c$, the functions $u_k : [0,T] \to \mathbb{R}$ are the control functions and $\hat{H}_k : \mathcal{H} \to \mathcal{H}$ are the control Hamiltonians. For brevity, we write $u(t) = [u_1(t), \cdots, u_{N_c}(t)]^T$. We aim to minimise the cost functional J:

$$J[u(\cdot)] := \int_0^T F\big(\,|\psi(t)\rangle\,, u(t)\big)dt + G(|\psi(T)\rangle), \tag{4.24}$$

with $F : \mathcal{H} \times \mathcal{U} \to \mathbb{R}$ and $G : \mathcal{H} \to \mathbb{R}$ being the running and final cost, respectively. However, the ket $|\psi(\cdot)\rangle : [0,T] \to \mathcal{H}$ must still satisfy the constraints provided by the Schrödinger equation:

$$\frac{\mathrm{d}}{\mathrm{d}t}\,|\psi(t)\rangle = -\frac{i}{\hbar}\hat{H}(t)\,|\psi(t)\rangle\,, \quad \forall\ t \in [0,T], \tag{4.25}$$

with the normalized initial condition $|\psi(0)\rangle = |\psi_0\rangle$. From now on, we use $\hbar = 1$. Similarly to the classical case, we can introduce an adjoint quantum state $\langle\chi(\cdot)| : [0,T] \times \mathcal{H} \to \mathbb{R}$, which will act

as a Lagrange multiplier which ensures that $|\psi(\cdot)\rangle$ satisfies the Schrödinger equation.

We can now introduce the Lagrangian augmented with the Lagrange multiplier $\langle\chi(\cdot)|$ associated with the Schrödinger equation:

$$\mathcal{L}\left(|\psi(t)\rangle, |\dot{\psi}(t)\rangle, \langle\chi(t)|, u(t)\right) = F(|\psi(t)\rangle, u(t)) + \mathrm{Re}\Big\{\langle\chi(t)|\dot{\psi}(t)\rangle + i\,\langle\chi(t)|\hat{H}(u(t))|\psi(t)\rangle\Big\}. \tag{4.26}$$

However, we have to take care to only work with real variables, so each quantum state vector $|\psi(t)\rangle$ has to be split into its real part $\mathrm{Re}\{|\psi(t)\rangle\} \in \mathbb{R}^n$ and its imaginary part $\mathrm{Im}\{|\psi(t)\rangle\} \in \mathbb{R}^n$.

The conjugate variables of $\mathrm{Re}\{|\psi(\cdot)\rangle\} : [0,T] \to \mathbb{R}^n$ and $\mathrm{Im}\{|\psi(\cdot)\rangle\} : [0,T] \to \mathbb{R}^n$, are $\mathrm{Re}\{\langle\chi(\cdot)|\} : [0,T] \to \mathbb{R}^{1\times n}$ and $\mathrm{Im}\{\langle\chi(\cdot)|\} : [0,T] \to \mathbb{R}^{1\times n}$ respectively. For simplicity, we take $\mathbb{R}^{1\times n}$ to represent the dual space of $\mathbb{R}^n$, so that we can work intuitively with column and row vectors. They satisfy the equations:

$$\mathrm{Re}\{\langle\chi(t)|\} = \frac{\partial\mathcal{L}}{\partial\,\mathrm{Re}\Big\{|\dot{\psi}(t)\rangle\Big\}}(|\psi(t)\rangle, |\dot{\psi}(t)\rangle, \langle\chi(t)|, u(t)), \tag{4.27}$$

$$\mathrm{Im}\{\langle\chi(t)|\} = \frac{\partial\mathcal{L}}{\partial\,\mathrm{Im}\Big\{|\dot{\psi}(t)\rangle\Big\}}(|\psi(t)\rangle, |\dot{\psi}(t)\rangle, \langle\chi(t)|, u(t)). \tag{4.28}$$

This can be deduced using the following term from equation (4.26):

$$\mathrm{Re}\Big\{\langle\chi(t)|\dot{\psi}(t)\rangle\Big\} = \mathrm{Re}\{\langle\chi(t)|\}\,\mathrm{Re}\Big\{|\dot{\psi}(t)\rangle\Big\} + \mathrm{Im}\{\langle\chi(t)|\}\,\mathrm{Im}\Big\{|\dot{\psi}(t)\rangle\Big\}. \tag{4.29}$$

Note that in the equations (4.27) and (4.28), a ket derivative produces a bra state; this convention ensures that we don't have to take the transpose of vectors everywhere, but it also makes sense on a broader level. For example, consider the function $f : \mathbb{R}^n \to R$, with $f(x) = a^T x$ and $a \in \mathbb{R}^n$. In this case, it is most straightforward to use the convention that $\frac{\partial f}{\partial x} = a^T$.

Through the Euler-Lagrange equations for $\mathrm{Re}\{|\chi(\cdot)\rangle\}$ and $\mathrm{Im}\{|\chi(\cdot)\rangle\}$, we arrive at the real and imaginary parts of the Schrödinger equation separately:

$$\mathrm{Re}\Big\{|\dot{\psi}(t)\rangle\Big\} = \mathrm{Re}\Big\{-i\hat{H}(u(t))\,|\psi(t)\rangle\Big\}, \tag{4.30}$$

$$\mathrm{Im}\Big\{|\dot{\psi}(t)\rangle\Big\} = \mathrm{Im}\Big\{-i\hat{H}(u(t))\,|\psi(t)\rangle\Big\}. \tag{4.31}$$

Together, they ensure that the Schrödinger equation is satisfied. So even though we only try to conserve the real part of the Schrödinger equation in equation (4.26), we get the conservation of the imaginary part for free.

We can write down the cost functional for an optimal control problem as:

$$J\Big[\,|\psi(\cdot)\rangle\,,\langle\chi(\cdot)|\,,u(\cdot)\Big] = G\Big(\,|\psi(T)\rangle\,\Big) \tag{4.32}$$

$$+\int_0^T F\Big(\,|\psi(t)\rangle\,,u(t)\Big) + \mathrm{Re}\Big\{\langle\chi(t)|\dot{\psi}(t)\rangle + i\,\langle\chi(t)|\hat{H}(u(t))|\psi(t)\rangle\Big\}dt. \tag{4.33}$$

Which is the sum of the final and integrated running cost augmented by the Lagrange multipliers associated with the Schrödinger equation. We can then define the Pontryagin Hamiltonian $H_P : \mathcal{H}\times\mathcal{H}^*\times\mathcal{U}\to\mathbb{R}$ in analogy to equation (4.18):

$$H_P\Big(\,|\psi(t)\rangle\,,\langle\chi(t)|\,,u(t)\Big) = \mathrm{Re}\Big\{\langle\chi(t)|\dot{\psi}(t)\rangle\Big\} + \chi_0 F\Big(\,|\psi(t)\rangle\,,u(t)\Big). \tag{4.34}$$

Here $\chi_0\leq 0$ is the abnormal multiplier, in most cases it can be considered to be negative, and then subsequently normalised to -1. However, in some special cases, setting $\chi_0=0$ is required. For full generality, we will not set $\chi_0=-1$ here. Using the Schrödinger equation, it can also be written as:

$$H_P\Big(\,|\psi(t)\rangle\,,\langle\chi(t)|\,,u(t)\Big) = \mathrm{Im}\Big\{\langle\chi(t)|\hat{H}(u(t))|\psi(t)\rangle\Big\} + \chi_0 F\Big(\,|\psi(t)\rangle\,,u(t)\Big). \tag{4.35}$$

Now that we have a Hamiltonian, we would like to compute Hamilton's equations as well. However, we are now differentiating by complex wave functions, so care is required. Following [28], we define the complex derivative as:

$$\frac{\partial}{\partial\,|\psi\rangle} = \frac{1}{2}\left[\frac{\partial}{\partial\,\mathrm{Re}\{|\psi\rangle\}} - i\frac{\partial}{\partial\,\mathrm{Im}\{|\psi\rangle\}}\right]. \tag{4.36}$$

This introduces a factor of two into Hamilton's equations:

$$|\dot{\psi}_*(t)\rangle = 2\frac{\partial H_P}{\partial\,\langle\chi|}(|\psi_*(t)\rangle\,,\langle\chi_*(t)|\,,u_*(t)) = -i\hat{H}(u(t))\,|\psi_*(t)\rangle\,, \tag{4.37}$$

$$\langle\dot{\chi}_*(t)| = -2\frac{\partial H_P}{\partial\,|\psi\rangle}(|\psi_*(t)\rangle\,,\langle\chi_*(t)|\,,u_*(t)) = i\,\langle\chi_*(t)|\,\hat{H}(u(t)) - 2\chi_0\frac{\partial F}{\partial\,|\psi\rangle}(|\psi_*(t)\rangle\,,u_*(t)). \tag{4.38}$$

In addition, the adjoint state satisfies a boundary condition at the final time T:

$$|\chi_*(T)\rangle = 2\chi_0\frac{\partial G}{\partial\,|\psi\rangle}(|\psi_*(T)\rangle). \tag{4.39}$$

The final equation of the Pontryagin's minimum principle becomes:

$$H_P\big(\,|\psi_*(t)\rangle\,,\langle\chi_*(t)|\,,u_*(t)\big) = \min_{u\in\mathcal{U}}\,H_P\big(\,|\psi_*(t)\rangle\,,\langle\chi_*(t)|\,,u\big),\quad\forall\,t\in[0,T]. \tag{4.40}$$

A typical choices for the final and running costs are:

$$G(|\psi(T)\rangle) = 1 - |\langle\psi_{\text{target}}|\psi(T)\rangle|^2, \tag{4.41}$$

$$F(|\psi(t)\rangle, u(t)) = \frac{\lambda_1}{2}u(t)^T u(t), \tag{4.42}$$

with $\lambda_1 \geq 0$. We will now take the time to apply quantum optimal control to a simple toy problem, the two-level system.

4.4 Quantum Optimal control of two-level System

We will again study the two-level system, this time using the formalism of quantum optimal control theory. So the Hilbert space is $\mathcal{H} = \mathbb{C}^2$. We start with the Hamiltonian from equation (2.69):

$$\hat{H}(t) = -\frac{1}{2}\omega_0\hat{\sigma}_z - \mu E_{||}(t)\hat{\sigma}_x \tag{4.43}$$

Like in Section 2.7, we can apply the rotating wave approximation to it, under the assumption that the frequency ω of the electric field $E_{||}(t)$ matches the internal frequency ω, which results in the new Hamiltonian:

$$\hat{H}(t) = u(t)\hat{\sigma}_x, \tag{4.44}$$

where $u(t) = -\mu A(t)$, with $A(t)$ being the pulse amplitude of $E_{||}(t)$. Our goal is to minimise the time it takes to reach the $|\psi(T)\rangle = e^{i\theta}|1\rangle = \begin{pmatrix} 0 \\ e^{i\theta} \end{pmatrix}$ exactly, as fast as possible, the phase $\theta \in \mathbb{R}$ is arbitrary and not physically relevant, whilst also minimising for control cost amplitude:

$$G(|\psi(T)\rangle) = 1 - |\langle 1|\psi(T)\rangle|^2, \tag{4.45}$$

$$F(|\psi(t)\rangle, u(t)) = \frac{1}{2}\lambda_1(u(t))^2. \tag{4.46}$$

The running cost $F(|\psi(t)\rangle, u(t))$ is proportional to the instantaneous energy density of the electric field produced by the laser. The energy density of the electric field parallel to the dipole moment $\vec{\mu}$ is given by $\epsilon_0(E_{||}(t))^2$, with ϵ_0 being the vacuum permittivity constant. This means that the integral of this choice of running cost is proportional to the total energy of the pulse. Therefore, we will frequently look to minimise it in future chapters. Averaging over time due to the fast oscillation caused by the laser frequency ω, we get:

$$\langle U\rangle = \frac{1}{2}\epsilon_0(A(t))^2, \tag{4.47}$$

which means that $F(|\psi(t)\rangle, u(t)) = \mu^2\langle U\rangle$, therefore, we will use the control cost amplitude as a proxy for the energy in the electric field.

We can write the Pontryagin Hamiltonian $H_P : \mathcal{H} \times \mathcal{H}^* \times \mathcal{U} \to \mathbb{R}$ as:

$$H_P = \mathrm{Im}\Big\{ \langle \chi(t)|\hat{H}(t)|\psi(t)\rangle \Big\} + \frac{1}{2}\chi_0\lambda_1(u(t))^2, \tag{4.48}$$

$$= u(t)S_x(t) + \frac{1}{2}\chi_0\lambda_1(u(t))^2. \tag{4.49}$$

Where $\chi_0 \leq 0$ is the abnormal multiplier and we define:

$$S_j(t) := \mathrm{Im}\{\langle \chi(t)|\hat{\sigma}_j|\psi(t)\rangle\}, \quad j = x, y, z. \tag{4.50}$$

The resulting Hamilton's equations are then:

$$|\dot{\psi}(t)\rangle = -iu(t)\hat{\sigma}_x |\psi(t)\rangle , \tag{4.51}$$

$$|\dot{\chi}(t)\rangle = -iu(t)\hat{\sigma}_x |\chi(t)\rangle . \tag{4.52}$$

Using Pontragrayin's Minimum principle, we can state that

$$\frac{\partial H_P}{\partial u}\big(|\psi_*(t)\rangle , \langle \chi_*(t)| , u_*(t)\big) = S_x(t) + \chi_0\lambda_1 u_*(t) = 0. \tag{4.53}$$

Which will be useful later. We can rewrite these equations in terms of time derivatives of $S_{i=x,y,z}(t)$:

$$\dot{S}_j(t) = \mathrm{Im}\Big\{ \langle \dot{\chi}(t)|\hat{\sigma}_j|\psi(t)\rangle + \langle \chi(t)|\hat{\sigma}_j|\dot{\psi}(t)\rangle \Big\}, \tag{4.54}$$

$$= \mathrm{Re}\Big\{ \langle \chi(t)|\hat{H}(u(t))\hat{\sigma}_j|\psi(t)\rangle - \langle \chi(t)|\hat{\sigma}_j\hat{H}(u(t))|\psi(t)\rangle \Big\}, \tag{4.55}$$

$$= \mathrm{Re}\Big\{ \langle \chi(t)|[\hat{H}(u(t)), \hat{\sigma}_j]|\psi(t)\rangle \Big\}. \tag{4.56}$$

Which in our case translates into the linear differential equation:

$$\dot{\vec{S}}(t) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -2u(t) \\ 0 & 2u(t) & 0 \end{bmatrix} \cdot \vec{S}(t) \tag{4.57}$$

This means that $S_x(t)$ is a constant of motion, meaning that $u(t) = -\frac{1}{\chi_0\lambda_1}S_x(0) := u_0$. Plugging this into the Schrödinger equation, we get:

$$|\dot{\psi}(t)\rangle = -i \begin{bmatrix} 0 & u_0 \\ -u_0 & 0 \end{bmatrix} |\psi(t)\rangle . \tag{4.58}$$

If we start in the ground state, then the solution is

$$|\psi(t)\rangle = \begin{bmatrix} \cos(u_0 t) \\ \sin(u_0 t) \end{bmatrix}, \tag{4.59}$$

the requirement $|\psi(T)\rangle = |1\rangle$ turns into $u_0 T = \frac{\pi}{2} + k\pi$, with $k \in \mathbb{Z}$. To globally optimise for control amplitude cost, we pick $u_0 = \pm\frac{\pi}{2T}$. Reverting the rotating wave approximation, we get that the electric field amplitude

$$E_{||} = \frac{\pi}{2\mu T}\cos(\omega_0 t + \phi) \tag{4.60}$$

is the optimal control function. This is exactly the π-pulse solution which was presented in equation (2.68).

4.5 Quantum Controllability

We will now discuss some important concepts in quantum optimal control, which will aid in understanding the quantum control simulations which we discuss later. Some basic questions which we can ask are: Is the system controllable? Meaning that we can reach any arbitrary final state. Do there exist sub-optimal local minima of the cost function? We will find that under a very general set of conditions, the quantum control problem is surprisingly easy, as also expressed in [29]. However, the introduction of additional terms to the cost functions will complicate matters significantly.

Following [30], we can define the idea of controllability in quantum systems and highlight a theorem which will allow us to determine whether a system is controllable with relative ease.

Definition 4.5.1 (Projective state controllable quantum system)**.** *A quantum system is said to be projective state controllable if for all initial and target states $|\psi_0\rangle, |\psi_{target}\rangle$, there exists a time $T > 0$ and control $u : [0, T] \to \mathcal{U}$ such that $|\psi(T)\rangle = e^{i\phi}|\psi_{target}\rangle$, for some arbitrary phase $\phi \in [0, 2\pi)$.*

In quantum mechanics, the complex phase of the entire system is not relevant because it does not affect Schrödinger's equation or probability distribution for any observable. Therefore, it is easier not to care about the phase of the system at $t = T$. We can now state the following theorem.

Theorem 4.5.2 (Lie Algebra Condition for Quantum Controllability [30, Theorem 2.2])**.** *Let $\mathcal{H} = \mathbb{C}^n$, the quantum system governed by the Heisenberg equation of motion for the unitary propagator $\hat{U}(t)$:*

$$\frac{\mathrm{d}\hat{U}}{\mathrm{d}t} = -\frac{i}{\hbar}\left(\hat{H}_0 + \sum_{k=1}^{N_c} u_k(t)\hat{H}_k\right)\hat{U}(t), \quad \hat{U}(0) = \hat{I}^{n\times n}, \tag{4.61}$$

with $\hat{H}_l : \mathbb{C}^n \to \mathbb{C}^n$ for $l = 0, \cdots N_c$ as the Hamiltonian operators, which can simultaneously be considered as matrices in $\mathbb{C}^{n\times n}$ and $u_{k=1,\cdots N_c}(t)$ are real-valued control functions. This quantum system is projective state controllable if and only if the Lie algebra generated by the commutators of the set $\{-i\hat{H}_0, \cdots, -i\hat{H}_{N_c}\}$ is isomorphic to $\mathfrak{sp}(n/2)$ or $\mathfrak{su}(n)$ if n is even or to $\mathfrak{su}(n)$ if n is odd.

Here $\mathfrak{su}(n)$ denotes the group of skew-Hermitian $n \times n$ complex matrices and $\mathfrak{sp}(n/2)$ denotes the group of block matrices $\begin{bmatrix} A & B \\ C & -A^T \end{bmatrix}$ such that $A, B, C \in \mathbb{R}^{n/2\times n/2}$ and B, C are symmetric.

Typically, the conditions of Theorem 4.5.2 are easily met, and the system is projective state controllable; an example is given in Section 5.7.

4.6 Sub-Optimal Extrema of the Infidelity

Optimising the control functions solely to achieve maximal overlap between the final and target state is often remarkably easy to achieve. Gradient-based optimisation algorithms never seem to get stuck on sub-optimal local extrema for the majority of quantum optimal control (QOC) systems. This is a very attractive property of a large variety of QOC systems. However, before we move forward and depend blindly on this, we should determine for which QOC systems this holds specifically.

The overlap between the final state $|\psi(T)\rangle$ and target state $|\psi_{target}\rangle$ is often called the fidelity of the final state in the literature. We define the fidelity and infidelity as:

$$\mathcal{F} := \left|\langle\psi_{target}|\hat{U}(T,0)|\psi_0\rangle\right|^2, \quad \Delta\mathcal{F} = 1 - \mathcal{F}, \tag{4.62}$$

where $|\psi_0\rangle$ is the initial state of the system at $T = 0$ and is fixed. The time evolution operator $\hat{U}(t_2, t_1) : \mathcal{H} \to \mathcal{H}$ evolves the state from $t = t_1$ to $t = t_2$:

$$\hat{U}(t_2, t_1)\,|\psi(t_1)\rangle = \psi(t_2). \tag{4.63}$$

Following [31], we will give an analytic expression of the gradient of the infidelity cost function, and based on that, we can make statements about, for example, the existence of sub-optimal local infidelity minima. We will assume that the system is fully controllable since, otherwise, it is not guaranteed that a global optimum with zero infidelity exists.

Consider the system Hamiltonian with one control Hamiltonian:

$$\hat{H} = \hat{H}_0 + u(t)\hat{H}_1, \tag{4.64}$$

with initial state $|\psi_0\rangle$ and target state $|\psi_{target}\rangle$. We let the system evolve from $t = 0$ to $t = T$. Assume the cost function takes the form:

$$J[\psi(\cdot), \chi(\cdot), u(\cdot)] = 1 - \left|\langle\psi_{target}|\hat{U}(T,0)|\psi_0\rangle\right|^2 + \mathrm{Im}\left\{\int_0^{t_f} \langle\chi(t)\left|i\hbar\frac{\partial}{\partial t} - \hat{H}_0 - u(t)\hat{H}_1\right|\psi(t)\rangle\, dt\right\}. \tag{4.65}$$

The first term minimises the infidelity of the final state, and the second term acts as a Lagrange multiplier term such that the Schrödinger equation is satisfied. Additionally, we assume that $\langle\psi_{target}|\psi_0\rangle = 0$ to simplify some of our later calculations. We now wish to compute the stationary points of this cost function. Applying the Euler-Lagrange equations gives us the three equations:

$$i\hbar\frac{\partial}{\partial t}|\psi(t)\rangle = \left[\hat{H}_0 + u(t)\hat{H}_1\right]|\psi(t)\rangle, \quad |\psi(0)\rangle = |\psi_0\rangle, \tag{4.66}$$

$$i\hbar\frac{\partial}{\partial t}|\chi(t)\rangle = \left[\hat{H}_0 + u(t)\hat{H}_1\right]|\chi(t)\rangle, \quad |\chi(T)\rangle = -\frac{2}{\hbar}|f\rangle\langle f|\hat{U}(T,0)|i\rangle, \tag{4.67}$$

$$\frac{\delta J}{\delta u(t)} = \mathrm{Im}\Big\{\langle\psi(t)|\hat{H}_1|\chi(t)\rangle\Big\}. \tag{4.68}$$

Where $\frac{\delta J}{\delta u(t)}$ gives the infinitesimal change in J as you make an infinitesimal change in the value u at time t. Using the expression for $|\chi(T)\rangle$, the last equation can be rewritten to:

$$\frac{\delta J}{\delta u(t)} = -\frac{2}{\hbar}\,\mathrm{Im}\Big\{\langle\psi_0|\hat{U}^\dagger(t,0)\hat{H}_1\hat{U}(t,0)\hat{U}^\dagger(T,0)|\psi_{target}\rangle\,\langle\psi_{target}|\hat{U}(t_f,0)|\psi_0\rangle\Big\}, \tag{4.69}$$

$$= -\frac{2}{\hbar}\,\mathrm{Im}\Big\{\langle\psi_0|\hat{H}_1(t)|q\rangle\Big\} = 0, \tag{4.70}$$

with the second line using the following expressions:

$$\hat{H}_1(t) = \hat{U}^\dagger(t,0)\hat{H}_1\hat{U}(t,0), \tag{4.71}$$

$$|q\rangle = \hat{U}^\dagger(T,0)\,|\psi_{target}\rangle\,\langle\psi_{target}|\hat{U}(T,0)|\psi_0\rangle\,. \tag{4.72}$$

With this, we will now perform an analysis of the gradient at the global extrema and any potential non-global extrema.

Gradient at points such that $\mathcal{F} = 0$

Let us first look at the points of the quantum control landscape which we definitely wish to avoid, namely the points which minimise the fidelity of the final state: $\langle\psi_{target}|\psi(T)\rangle = 0$. This will serve as a warm-up. By assumption, we have $\langle\psi_{target}|\hat{U}(T,0)|\psi_0\rangle = 0$. Which implies that $|q\rangle = 0$. In turn, meaning that $\frac{\delta J}{\delta u(t)} = 0$ at all times.

This makes sense since if this were not true, you could pick a small perturbation $\Delta u(\cdot)$ such that $J[u(\cdot)+\Delta u(\cdot)] < 0$, whilst still satisfying the Schrödinger equation, meaning that you have negative fidelity, which is not possible. A similar argument will hold for control functions with $\mathcal{F} = 1$.

Gradient at points such that $\mathcal{F} = 1$

Since we consider the norm of the fidelity to be one, we thus have that the final state is $\hat{U}(T,0)\,|\psi_0\rangle = e^{i\phi}\,|\psi_{target}\rangle$, where $\phi \in \mathbb{R}$ is an arbitrary phase. This means that $|q\rangle = |\psi_0\rangle$, so the expression for $\frac{\delta J}{\delta u(t)} = 0$ turns into:

$$\frac{\delta J}{\delta u(t)} = -\frac{2}{\hbar}\,\mathrm{Im}\Big\{\langle\psi_0|\hat{H}_1(t)|\psi_0\rangle\Big\} = 0 \tag{4.73}$$

which holds at all times because $\hat{H}_1(t)$ is Hermitian and thus $\langle\psi_0|\hat{H}_1(t)|\psi_0\rangle$ is real at all times. This shows that points with zero infidelity are indeed extrema.

Gradient at points such that $0 < \mathcal{F} < 1$

Now we move onto the most interesting point, namely whether there exist sub-optimal local extrema. Luckily, we will see that they don't exist for a rather general class of systems.

We seek solutions such that $\frac{\delta J}{\delta u(t)} = 0$ for all times. Which in turn requires that $\frac{\partial^n}{\partial t^n}\left(\frac{\delta J}{\delta u(t)}\right) = 0$ at all times and all $n \in \mathbb{N}$. We can use the Heisenberg equation of motion:

$$\frac{\mathrm{d}\hat{A}}{\mathrm{d}t}(t) = \frac{i}{\hbar}\left[\hat{H}(t), \hat{A}(t)\right]. \tag{4.74}$$

Which, for $\hat{H}_1$ becomes:

$$\frac{\mathrm{d}\hat{H}_1}{\mathrm{d}t}(t) = \frac{i}{\hbar}\left[\hat{H}_0(t), \hat{H}_1(t)\right]. \tag{4.75}$$

Using these expressions, we can write down the first and second time derivative of $\frac{\delta J}{\delta u(t)}$ as:

$$\frac{\partial}{\partial t}\frac{\delta J}{\delta u(t)} = -\frac{2}{\hbar}\,\mathrm{Im}\left\{\left(\frac{i}{\hbar}\right)\langle\psi_0|\left[\hat{H}_0(t), \hat{H}_1(t)\right]|q\rangle\right\} = 0, \tag{4.76}$$

$$\frac{\partial^2}{\partial t^2}\frac{\delta J}{\delta u(t)} = -\frac{2}{\hbar}\,\mathrm{Im}\left\{\left(\frac{i}{\hbar}\right)^2\langle\psi_0|\left[\hat{H}(t), \left[\hat{H}_0(t), \hat{H}_1(t)\right]\right]|q\rangle\right\} = 0. \tag{4.77}$$

Where further derivatives also consist of nested commutators. These commutators depend on the set of commutators generated by $\hat{H}_0$ and $\hat{H}_1$. These same commutators of $\hat{H}_0$ and $\hat{H}_1$ are used to determine the controllability of the system in Theorem 4.5.2. We have assumed that the system is fully controllable. Meaning that the Lie algebra generated by these commutators is isomorphic to $\mathfrak{sp}(n/2)$ or $\mathfrak{su}(n)$. It can be shown that this implies that not all time derivatives of $\frac{\delta J}{\delta u(t)}$ are zero. Therefore, $\frac{\delta J}{\delta u(t)} = 0$ can't be satisfied at all times, so no sub-optimal extrema can exist if the system is controllable. This means that any reasonable gradient-based optimisation algorithm should always be able to find zero-infidelity states.

4.7 Quantum Control Landscape

For two-level quantum systems, with two smooth control fields $u_{1,2} \in L^1([0, T])$ and the final time T as a free variable, the system is considered strongly controllable and it can be shown that the kernel of $\Delta\mathcal{F}[u]$ is path-connected in the space of control functions [32].

This implies that if you wish to minimise a second quantity like the total control amplitude, you don't necessarily have to leave the zero-infidelity subspace, although it might be significantly more computationally expensive to converge to the optimal solution. The recent paper [33] shows, using numerical simulations, that there are paths in the low-infidelity subspace connecting different zero-infidelity points. An example of such a path is given in Figure 4.1.

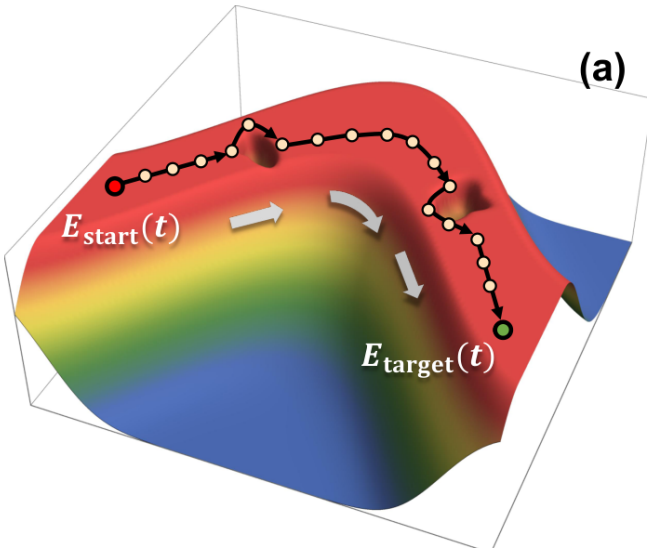

Figure 4.1: Schematic of the path-connectedness of the zero-infidelity subspace. Taken from [33].

For a N-level system, the Hessian at u_* is given by the matrix $\frac{\delta^2 \Delta \mathcal{F}}{\delta u(t) \delta u(t')}[u_*]$ has at most $2N-2$ eigenfunctions with positive eigenvalues. Meaning that if the number of control variables exceeds $2N-2$, then we are guaranteed to have a path-connected zero-infidelity subspace. The trace of the Hessian provides a measure of the robustness of the control pulse, as it is equal to the sum of all of the (positive) eigenvalues of the Hessian. When we employ the GRAPE algorithm, we typically use hundreds of control variables [34], meaning that the zero-infidelity subspace will not only be connected, but also have a very large number of dimensions.

4.8 Overview

In this chapter, we introduced the optimal control framework, which we will continue to use in this chapter. It is also the starting point from which we will describe the GRAPE and GOAT optimal control algorithms. The core result, which was discussed, is Pontryagin's Minimum Principle; however, even its quantum variant is not directly used in either the GRAPE or GOAT algorithm. Nevertheless, it is still an important jumping-off point; for example, the GRAPE algorithm also relies on adjoint states.

The theoretical results on controllability and the lack of non-global local extrema will be verified numerically throughout the following two chapters. The control landscape will play a key role in Section 5.4, however, we will have no space to further explore the control landscape.

Chapter 5

GRAPE Algorithm

Having now established the governing equations of quantum optimal control in the previous section, we will now attempt to numerically determine the optimal control fields. We will start by discussing the GRAPE quantum optimal control algorithm, which is conceptually the simplest.

In this chapter and Chapter 6, we will use arbitrary units for our numerical simulations of quantum systems. In Chapters 7 and 8, we will discuss more realistic models and will therefore use a more appropriate unit system.

The GRAPE algorithm was developed in 2005 [35] and it is widely used for QOC-problems. It models the control pulses as piecewise constant functions with n segments, and it performs a gradient descent algorithm[1] using the gradient of the fidelity function with respect to the pulse amplitude at each time interval. The primary innovation is the computation of these gradients. It is able to find high-fidelity solutions using this gradient descent algorithm easily.

The one downside compared to other algorithms is the limited ability to constrain the shape of the optimised pulse. You can either add constraints to the gradient descent algorithm, or you can optimise for a cost function which includes a fidelity term and a term which penalises any deviation of the control pulses from their desired shape. Our introduction to the GRAPE algorithm will follow [28].

5.1 Mathematical Model

We will now discuss the mathematical background of the GRAPE algorithm, assume that the Hilbert space is $\mathcal{H} = \mathbb{C}^{N_d}$ and that the control Hamiltonian is similar to the control Hamiltonian in Section 4.3, except the general control function $u_k(t) : [0, T] \to \mathbb{R}$ are replaced by piecewise constant functions $f_{pc} : [0, T] \times \mathbb{R}^n \to \mathbb{R}$ parametrised by $\overline{u}_{(k)} \in \mathbb{R}^n$:

[1] Contrary to the full name of the GRAPE algorithm, our objective will be to minimise infidelity, which is equivalent to maximising the fidelity. However, when solutions reach near-perfect fidelity on the range of $\mathcal{F} = 1 - 10^{-12}$, working with the infidelity becomes easier.

$$\hat{H}(t;\overline{u}_{(1)},\cdots,\overline{u}_{(N_c)}) = \hat{H}_0 + \sum_{k=1}^{N_c} f_{pc}(t;\overline{u}_{(k)})\hat{H}_k, \tag{5.1}$$

where f_{pc} generates a piecewise constant function with n segments of size $\Delta t = \frac{T}{n}$ with values given by $\overline{u}_{(k)}$. The segment pieces are:

$$S_l = \begin{cases} [(l-1)\Delta t,\ l\Delta t) & \text{for } l = 1,\cdots,n-1, \\ [(n-1)\Delta t, T] & \text{for } l = n, \end{cases} \tag{5.2}$$

We can now define the piecewise constant function f_{pc} as:

$$f_{pc}(t;\overline{u}) = \sum_{l=1}^{n} u_l 1_{S_l}(t), \tag{5.3}$$

where the indicator functions $1_{S_l} : [0,T] \to \{0,1\}$ are defined in Definition 3.1.4.

The control Hamiltonians are $\hat{H}_k : \mathbb{C}^{N_d} \to \mathbb{C}^{N_d}$. For the following discussion, N_d is arbitrary, and we will set $N_c = 1$ for simplicity, and we will assume that $\hat{H}_0$ and $\hat{H}_1$ do not commute. However, you can also perform the GRAPE algorithm with multiple control functions at the same time.

The GRAPE algorithm can also be used to perform quantum optimal control when working with open quantum systems, where density matrices represent the state of the system. However, for simplicity, we will perform the mathematical derivation of the GRAPE algorithm in the bra-ket formalism, see the original paper [35] for the density matrix formalism.

We have now transformed the infinite-dimensional control problem into a finite-dimensional problem to optimise the control inputs $\overline{u} \in \mathbb{R}^n$. We will now work to find the gradient of the infidelity function with respect to the control inputs $\overline{u}$.

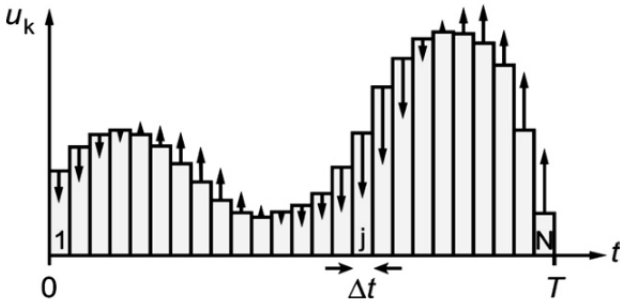

Figure 5.1: Illustration of the GRAPE algorithm, where the piecewise constant function illustrates the initial pulse and the attached arrows signify one iteration of the gradient descent algorithm. Taken from [35].

Define the evolution operator $\hat{U}_l : \mathbb{R} \times \mathbb{C}^{N_d} \to \mathbb{C}^{N_d}$, which evolves the state from $t = (l-1)\Delta t$ to $t = l\Delta t$ as:

$$\hat{U}_l(u_l) := \hat{U}\big(l\Delta t, (l-1)\Delta t\big) = \exp\left(-i\Delta t\left(\hat{H}_0 + u_l\hat{H}_1\right)\right), \tag{5.4}$$

for $l = 1, \cdots, n$. Additionally define,

$$|\psi_l\rangle := \hat{U}_l(u_l)\cdots\hat{U}_1(u_1)\,|\psi_0\rangle = |\psi(l\Delta t)\rangle\,, \quad \text{for } l = 1, \cdots n, \tag{5.5}$$

with $|\psi_0\rangle$ being the initial state at $t = 0$. The infidelity function $\Delta\mathcal{F} : \mathbb{R}^n \to [0, 1]$ is defined as:

$$\Delta\mathcal{F}(\overline{u}) = G(\hat{U}_n(u_n)\cdots\hat{U}_1(u_1)\,|\psi_0\rangle) = G(|\psi_n\rangle), \tag{5.6}$$

where $G : \mathbb{C}^n \to \mathbb{R}$ give the infidelity of the control pulse and is typically given as:

$$G(|\psi_n\rangle) = 1 - |\,\langle\psi_{target}|\psi_n\rangle\,|^2, \tag{5.7}$$

where the ket $|\psi_{target}\rangle$ is the target state of the quantum control optimisation. The infidelity function has the fastest decrease in the $-\frac{\partial(\Delta\mathcal{F})}{\partial u_l}$ direction. Therefore, we update u_l according to the gradient descent algorithm:

$$u_l' = u_l - \epsilon\frac{\partial(\Delta\mathcal{F})}{\partial u_l}(\overline{u}). \tag{5.8}$$

Where the parameter $\epsilon > 0$ is tuned to be small enough to ensure convergence, but still large enough for speedy convergence. However, the derivative $\frac{\partial(\Delta\mathcal{F})}{\partial u_l}(\overline{u})$ is difficult to compute, and its efficient computation forms the core innovation of GRAPE. We will now show its derivation.

Using the chain rule, we can write:

$$\frac{\partial(\Delta\mathcal{F})}{\partial u_l}(\overline{u}) = \frac{\partial G}{\partial\,|\psi_n\rangle}(|\psi_n\rangle)\frac{\partial\,|\psi_n\rangle}{\partial u_l}(\overline{u}) + \frac{\partial\,\langle\psi_n|}{\partial u_l}(\overline{u})\frac{\partial G}{\partial\,\langle\psi_n|}(|\psi_n\rangle), \tag{5.9}$$

$$= \frac{\partial G}{\partial\,|\psi_n\rangle}(|\psi_n\rangle)\hat{U}_n(u_n)\cdots\frac{\partial\hat{U}_l}{\partial u_l}(u_l)\cdots\hat{U}_1(u_1)\,|\psi_0\rangle \tag{5.10}$$

$$+\,\langle\psi_0|\,\hat{U}_1^\dagger(u_1)\cdots\frac{\partial\hat{U}_l^\dagger}{\partial u_l}(u_l)\cdots\hat{U}_n(u_n)\frac{\partial G}{\partial\,\langle\psi_n|}(|\psi_n\rangle), \tag{5.11}$$

Now we should compute $\frac{\partial\hat{U}_l}{\partial u_l}(u_l)$, which is not straightforward since $\hat{H}_0$ and $\hat{H}_1$ do not compute. However, that can be achieved using Wilcox's formula [36] for a matrix operator $\hat{A} : \mathbb{R} \to (\mathbb{C}^{N_d} \to \mathbb{C}^{N_d})$ depending on a real parameter θ:

$$\frac{\partial(e^{t\hat{A}(\theta)})}{\partial\theta} = e^{t\hat{A}(\theta)}\int_0^t e^{-t'\hat{A}(\theta)}\frac{\partial\hat{A}}{\partial\theta}(\theta)e^{t'\hat{A}(\theta)}dt'. \tag{5.12}$$

Plugging $\hat{A}(\theta) = -i(\hat{H}_0 + \theta\hat{H}_1)$ into the above equation, it can be shown that,

$$\frac{\partial\hat{U}_l}{\partial u_l}(u_l) = -i\Delta t\hat{U}_l(u_l)\overline{\hat{H}_1}(u_l). \tag{5.13}$$

with $\overline{\hat{H}_1}(u_l)$ being the time-average of $\hat{H}_1$ in the Heisenberg representation:

$$\overline{\hat{H}_1}(u_l) = \frac{1}{\Delta t}\int_0^{\Delta t} e^{it'(\hat{H}_0+u_l\hat{H}_1)}\hat{H}_1 e^{-it'(\hat{H}_0+u_l\hat{H}_1)}dt'. \tag{5.14}$$

If we let $\Delta t \to 0$, then $\overline{\hat{H}_1}(u_l) \to \hat{H}_1$. When doing numerical work, later in this thesis, we often set Δt pretty small such that the control pulses resemble smooth functions. Therefore, we approximate $\overline{\hat{H}_1}(u_l)$ to be equal to $\hat{H}_1$. Additionally, this approximation only leads to an error in the computation of the gradient, not in the computation of the fidelity itself. In our numerical work, it does not seem to affect the convergence of the gradient descent algorithm. So we can say:

$$\frac{\partial \hat{U}_l}{\partial u_l}(u_l) \simeq -i\Delta t \hat{U}_l(u_l)\hat{H}_1. \tag{5.15}$$

From the initial conditions from equation (4.39), we have that the adjoint state satisfies the following boundary condition at the final time T:

$$\langle\chi_n| = \langle\chi(n\Delta t)| = \langle\chi(T)| = 2\chi_0 \frac{\partial G}{\partial\langle\psi_n|}(|\psi_n\rangle). \tag{5.16}$$

We can then define the adjoint state at $t = l\Delta t$ through backward propagation in time:

$$\langle\chi_l| = \langle\chi_n|\hat{U}_n(u_n)\cdots\hat{U}_{l+1}(u_{l+1}), \quad \text{for } l = 0, \cdots n-1. \tag{5.17}$$

Then using equation (5.9), we can state:

$$\frac{\partial(\Delta\mathcal{F})}{\partial u_l}(\overline{u}) = -i\Delta t(\langle\chi_l|\hat{H}_1|\psi_{l-1}\rangle + \langle\psi_{l-1}|\hat{H}_1|\chi_l\rangle), \tag{5.18}$$

$$= 2\Delta t\,\mathrm{Im}\Big\{\langle\chi_l|\hat{H}_1|\psi_{l-1}\rangle\Big\}. \tag{5.19}$$

We have thus finally arrived at the desired expression

$$u_l' = u_l - \epsilon\,\mathrm{Im}\Big\{\langle\chi_l|\hat{H}_1|\psi_{l-1}\rangle\Big\}, \tag{5.20}$$

where the $2\Delta t$ factor got absorbed into the arbitrary parameter ϵ. We can augment the cost function with any term $\mathcal{C}$, where $\frac{\partial\mathcal{C}}{\partial u_l}$ is only a function of $\overline{u}$ itself. Like, for example, $\mathcal{C}(\overline{u}) = \frac{1}{2}\sum_{l=1}^{n} u_l^2$. Additional terms, which also depend on $\psi(t)$, like a running cost, would require serious adjustments of any computer package which you are using to run GRAPE. However, you can extend the derivation above to include a running cost.

5.2 Software Implementation

The numerical simulation work done in this thesis is all done using the QuTiP Python package [37] and its offshoot packages Qutip-qtrl [38] and Qutip-qoc [39], which specialise in quantum optimal control. There are multiple layers of code required to discretise the problem at hand and apply

numerical optimisation to it.

An overview of the Qutip-qtrl is given in Figure 5.2. In the Qutip-qtrl codebase, everything relevant to a numerical quantum optimal control optimisation is gathered in the Optimizer object. It contains the parameters and termination conditions provided to the optimisation algorithm and also contains the dynamics object, which deals with the fidelity-related, timeslot-related and propagation computations.

Usually, we don't have to interact with most parts of the code case. We have made large changes to the FidelityComputer part, to allow for cost functions which are not only dependent on infidelity but also on the control pulses themselves. An adjusted version of the Optimizer object was also written to accommodate the various control cost functions, which will be discussed in the following sections.

For more specifics on the adjusted codebase and the IPYNB notebooks used to create the plots in this thesis, check out the Github repository:
`https://github.com/RoyRodenburg/MasterProjectQuantumOptimalControl`.

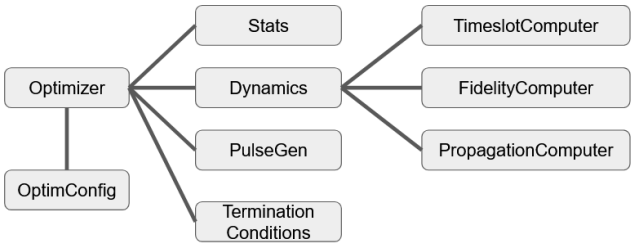

Figure 5.2: Overview of the dependencies of the Optimizer object in the qutip-qtrl package, adopted from [40].

We will now briefly discuss the generic setup, which we use when we try to numerically solve a quantum optimal control problems with the GRAPE algorithm.

We start with the GRAPE Hamiltonian

$$\hat{H}(t;\overline{u}_{(1)},\cdots,\overline{u}_{(N_c)}) = \hat{H}_0 + \sum_{k=1}^{N_c} f(t;\overline{u}_{(k)})\hat{H}_k. \tag{5.21}$$

The infidelity, which is to be minimised is defined as:

$$\Delta\mathcal{F}(\overline{u}) := 1 - |\,\langle\psi_{target}|\hat{U}_n(u_n)\cdots\hat{U}_1(u_1)|\psi_0\rangle\,|^2, \tag{5.22}$$

The gradient descent algorithm will work using the gradient-based BFGS optimisation algorithm, which is discussed in appendix A. The BFGS algorithm uses the gradients provided through the

GRAPE algorithm and uses a numerical estimation of the second derivative to speed up the gradient descent.

We will do our first GRAPE optimisation on one of the most basic quantum system: the two-level system $\mathcal{H} = \mathbb{C}^2$ with the Hamiltonian:

$$\hat{H}(u(t)) = \hat{H}_0 + u(t)\hat{H}_1, \tag{5.23}$$

with the Hamiltonians $\hat{H}_{0,1} : \mathbb{C}^2 \to \mathbb{C}^2$ being given using Pauli matrices:

$$\hat{H}_0 = -\frac{1}{2}\hat{\sigma}_z = \begin{bmatrix} -1/2 & 0 \\ 0 & 1/2 \end{bmatrix}, \quad \hat{H}_1 = \hat{\sigma}_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \tag{5.24}$$

At time $t = 0$, the system is in the ground state $|\psi_0\rangle = |0\rangle$ and we set the target state as $|\psi_{\text{target}}\rangle = |1\rangle$.

This system is very similar to the system in Section 2.6, except that it only has the real part of the laser field. In quantum optimal control, we work with real control fields; therefore, we require two separate control fields, which represent the real and imaginary parts of the laser field. However, adding a second control field slightly complicates matters, so we will start with just one control field. In Section 5.3, we will use two control fields, to recreate a π-pulse using the GRAPE algorithm.

The termination conditions for the GRAPE algorithm are: a minimal infidelity $\Delta\mathcal{F}$ of 10^{-15}, a minimal infidelity gradient $\left|\frac{\partial(\Delta\mathcal{F})}{\partial u_l}\right|$ of 10^{-15} and the maximal number of gradient descent steps is set at 400. The number of timeslots is 200, although the optimisation is not very sensitive to this parameter as long as it is not too small. If the optimisation is set up correctly, it should finish rather quickly, so setting very strict termination conditions is not an issue.

As an initial guess, we use the uniform probability distribution on $[0, 1]$ to assign a value to each element of $\overline{u}$.

Adam's Method

The default integration method used to solve the Schrödinger equation in the QuTiP package is the implicit Adam's method, using the Scipy implementation. The implicit Adam's method is a generalisation of Euler's method, providing additional accuracy and stability. When we have the differential equation $y' = f(t, y)$, then the next sequence element is:

$$y_{n+1} = y_n + \int_{t_n}^{t_{n+1}} f(t, y)dt. \tag{5.25}$$

Euler's method approximates f as a constant function, assuming a constant step size Δt, this leads to:

$$y_{n+1} = y_n + \Delta t f(t_n, y_n). \tag{5.26}$$

Instead, the implicit Adam's method uses an estimation of the average function value of f over the time interval $[t_n, t_{n+1}]$ to gain better accuracy. The second-order implicit Adam's method uses the following update equation:

$$y_{n+1} = y_n + \frac{\Delta t}{2}(f(t_n, y_n) + f(t_{n+1}, y_{n+1})). \tag{5.27}$$

This update equation is called implicit, because y_{n+1} is included on both sides. To efficiently solve such an equation numerically, we can use so-called predictor-corrector pairs: first, we get a predicted $\tilde{y}_{n+1}$ using the explicit Adam's method:

$$\tilde{y}_{n+1} = y_n + \frac{\Delta t}{2}(3y_n - y_{n+1}) \tag{5.28}$$

and then plugging this into the implicit Adam's method, we get the corrected y_{n+1}:

$$y_{n+1} = y_n + \frac{\Delta t}{2}(f(t_n, y_n) + f(t_{n+1}, \tilde{y}_{n+1})). \tag{5.29}$$

This method leads to a great amount of accuracy and stability for non-stiff problems. Scipy uses this type of Adam's method, with variable order: if the error between the predictor and corrector grows, the stepsize decreases, and the order of the Adam's method also decreases.

5.3 π-pulse Recreation

We will now work to recreate the π-pulse using the GRAPE algorithm, which will smoothly transfer the system from the ground state to the excited state using one control pulse, as discussed in Section 2.7.

$$\hat{H}(t; \overline{u}) = -\frac{1}{2}\hat{\sigma}_z + f_{pc}(t; \overline{u})\hat{\sigma}_x. \tag{5.30}$$

To recreate the π-pulse, we will impose a cost on the control amplitude:

$$\mathcal{C}_1(\overline{u}) = \frac{\Delta t}{2}\sum_{k=1}^{n} u_k^2, \tag{5.31}$$

the factor $\Delta t = \frac{T}{n}$ is added such that the cost $\mathcal{C}_1(\overline{u})$ stays approximately the same as the amount of pulse segments n is increased. The total cost to be minimised becomes:

$$\mathcal{C}(\overline{u}) = \Delta\mathcal{F}(\overline{u}) + \lambda_1\mathcal{C}_1(\overline{u}), \tag{5.32}$$

with cost parameter $\lambda_1 \geq 0$. The GRAPE algorithm works with a gradient descent algorithm, which requires the derivative $\frac{\partial\mathcal{C}}{\partial u_k}$. Because we have added the cost term $\mathcal{C}_1(\overline{u})$, we should also provide its derivative to the GRAPE algorithm to work:

$$\frac{\partial \mathcal{C}}{\partial u_k} = \frac{\partial(\Delta \mathcal{F})}{\partial u_k} + \lambda_1 \Delta t u_k. \tag{5.33}$$

If the optimised pulses would recreate a π-pulse, than $u(t)$ would be a sine function with amplitude $A = \frac{\pi}{2T}$ and have an angular frequency of 1 as discussed at the end of Section 2.6.

In Figure 5.3, we can see that adding a non-zero control amplitude cost term λ_1 drastically lowered the total control amplitude. Additionally, it also simplified the state transfer process, leading to a smoother transition from the ground state to the excited state.

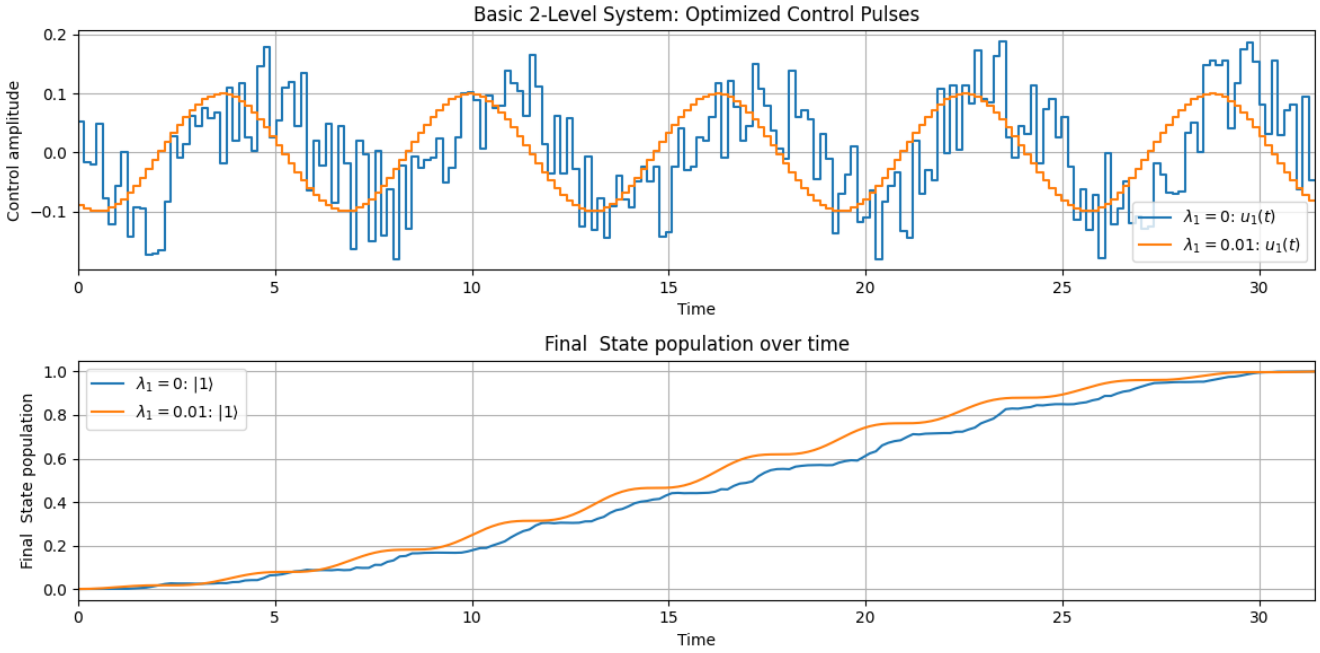

Figure 5.3: The results of the optimisation are shown above. In the upper plot, it can be seen that not optimising for control amplitude produces a rather jagged curve; the fidelity error of the co-optimised pulse is rather small. The fidelity errors are $< 1 \cdot 10^{-16}$ and $5.0 \cdot 10^{-5}$ for the optimisation with $\lambda_1 = 0$ and $\lambda_1 = 0.01$ respectively. The amplitude cost term $\mathcal{C}_1(\overline{u}_1, \overline{u}_2)$ are 0.128 and 0.078 respectively. The final time T is 10π.

As will become evident later, for a wide range of choices of λ_1, the optimised control pulses will remain largely similar. However, when the optimal control problems become too complicated, small values of λ_1 will get stuck at local minima of the infidelity, with final control pulses which have near-zero infidelity, but unnecessarily large control inputs. However, when λ_1 is set to be too large, the algorithm might choose to only optimise for control amplitude, preventing the system from reaching the target state.

The optimised control pulse, which corresponds to the π-pulse solution, which we discussed at the end of Section 2.6. There, we assumed that the control pulse would have an angular frequency of

$E_1 - E_0 = 1$, which corresponds to our solution. The constant amplitude of the sine waves of $u(t)$ can be explained through the introduction of $\mathcal{C}_1$, since a constant amplitude would minimise the total squared control amplitude. From Section 1.7, it is predicted that $A = \frac{\pi}{2T}$, which leads to a control amplitude of $A = \frac{\pi}{T} = 0.1$, which corresponds to what we see in the figure.

The introduction of a non-zero cost parameter λ_1 means that you have to balance between minimising infidelity and control amplitude cost, which leads to a higher infidelity, compared to the case that $\lambda_1 = 0$. It can be shown that there is a quadratic relationship between the infidelity and λ_1, which we will discuss in the next section.

We can also use two control pulses for the GRAPE optimisation:

$$\hat{H}(t; \overline{u}_1, \overline{u}_2) = -\frac{1}{2}\hat{\sigma}_z + f_{pc}(t; \overline{u}_1)\hat{\sigma}_x + f_{pc}(t; \overline{u}_2)\hat{\sigma}_y. \tag{5.34}$$

with the control cost becoming:

$$\mathcal{C}_1(\overline{u}_1, \overline{u}_2) = \frac{\Delta t}{2} \sum_{i=1}^{2} \sum_{k=1}^{n} u_{i,k}^2, \tag{5.35}$$

and the total cost function being:

$$\mathcal{C}(\overline{u}_1, \overline{u}_2) = \Delta\mathcal{F}(\overline{u}_1, \overline{u}_2) + \lambda_1 \mathcal{C}_1(\overline{u}_1, \overline{u}_2), \quad . \tag{5.36}$$

Optimising this cost function using the GRAPE algorithm, we get the following in Figure 5.4. You can see that the state transfer is a lot smoother compared to Figure 5.3, when a cost parameter $\lambda_1 = 0.01$ is imposed. Because with two pulses, at least one control pulse is non-zero at all times. The control pulses are both sine waves, which are $\frac{\pi}{2}$ out of phase with each other. Note that the control amplitude of each pulse is halved compared to the situation with only one pulse.

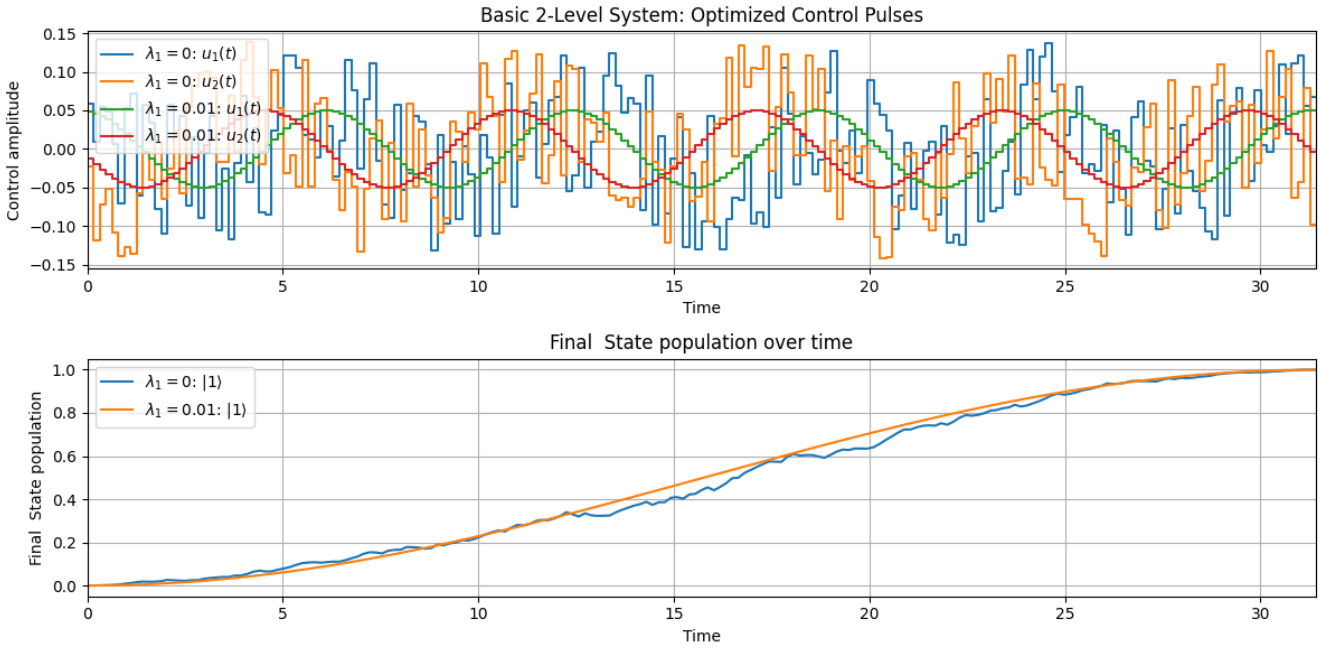

Figure 5.4: The results of the optimisation are shown above. In the upper plot, it can be seen that not optimising for control amplitude produces a rather jagged curve; the fidelity error of the co-optimised pulse is rather small. The fidelity errors are $1.6 \cdot 10^{-15}$ and $1.2 \cdot 10^{-5}$ for the optimisation with $\lambda_1 = 0$ and $\lambda_1 = 0.01$ respectively. The amplitude cost term $\mathcal{C}_1(\overline{u}_1, \overline{u}_2)$ are 0.0766 and 0.0196 respectively. The final time T is 10π.

5.4 Power Law Relationship between Infidelity and Cost Parameter

When you optimise for a linear combination of infidelity and control pulse amplitude, like in equation (5.32), you will no longer get that the optimised control pulse achieves zero infidelity. Rather, for small choices of λ_1, the infidelity will be small, but still non-zero. We would like to quantify the relationship between the cost parameter λ_1 and the infidelity.

Suppose that u_* is an optimal solution of the optimal control problem, such that $\Delta\mathcal{F}(\overline{u}_*) = 0$ and the cost function $\mathcal{C}_1(\overline{u})$ is also separately locally optimised with the constraint that the infidelity is zero.

However, we know that the zero-infidelity subspace of the control landscape is not zero-dimensional and can be very complex, as is discussed in Section 4.7. Therefore, we have to careful to not assume that the optimum we are looking for is isolated, usually there is a high-dimensional subspace of the control landscape which all have the same resulting cost.

To start, we will write the fidelity function $\Delta\mathcal{F}$ as a multi-dimensional Taylor expansion of the variable $\delta\overline{u} = \overline{u} - \overline{u}_*$:

$$\Delta\mathcal{F}(\delta u) \approx \frac{1}{2}\delta\overline{u}^T \mathrm{Hess}_{u_*}(\Delta\mathcal{F})\delta\overline{u}. \tag{5.37}$$

This Taylor expansion is valid as long as $|\delta\overline{u}|$ is sufficiently small. The matrix $H := \mathrm{Hess}_{u_*}(\Delta\mathcal{F})$ is the Hessian matrix of $\Delta\mathcal{F}$, where the matrix elements are:

$$H_{jk} = \frac{\partial^2\Delta\mathcal{F}}{\partial u_j \partial u_k}(u_*), \quad \text{for } j,k \leq n. \tag{5.38}$$

We know that the Hessian matrix H is positive semi-definite, because otherwise, we would be able to achieve a negative infidelity, which is not allowed.

For this derivation, we will only use values of $\delta\overline{u}$ such that $\delta\overline{u} \cdot \overline{u}_* < 0$, because only then $\mathcal{C}_1$ is lowered. This choice of δu must lead to an increase in infidelity; otherwise, our assumption that we locally optimised for the cost function under a zero-infidelity constraint is false. Therefore, we will assume that H is positive definite and thus invertible in the following text.

We will now write the cost function $\mathcal{C}_1$ as Taylor expansion of $\delta\overline{u}$:

$$\mathcal{C}_1(\delta\overline{u}) = \frac{1}{2}(\delta\overline{u} + \overline{u}_*) \cdot (\delta\overline{u} + \overline{u}_*) = \frac{1}{2}\delta\overline{u} \cdot \delta\overline{u} + \delta\overline{u} \cdot \overline{u}_* + \frac{1}{2}\overline{u}_* \cdot \overline{u}_*. \tag{5.39}$$

Combining these two Taylor expansions, we can write the Taylor expansion of the total cost function $\mathcal{C}(\delta\overline{u})$:

$$\mathcal{C}(\delta\overline{u}) \approx \frac{1}{2}\delta\overline{u}^T \left[H + \lambda_1 I\right] \delta\overline{u} + \lambda_1 \overline{u}_* \cdot \delta\overline{u} + \frac{1}{2}\lambda_1 \overline{u}_* \cdot \overline{u}_*. \tag{5.40}$$

When we optimise for this total cost function, the derivative with respect to $\delta\overline{u}$ should be the zero vector:

$$\frac{\partial \mathcal{C}}{\partial(\delta\overline{u})} = [H + \lambda_1 I]\,\delta\overline{u} + \lambda_1 \overline{u}_* = \overline{0}. \tag{5.41}$$

Because H is positive definite, it follows that $H + \lambda_1 I$ is also positive definite for $\lambda_1 > 0$, and thus also invertible. Therefore, we can write:

$$\delta\overline{u} = -\lambda_1 \left(H + \lambda_1 I\right)^{-1} \overline{u}_*, \tag{5.42}$$

$$\delta\overline{u}^T = -\lambda_1 \overline{u}_*^T \left(H + \lambda_1 I\right)^{-1}, \tag{5.43}$$

here the second equation stem from the fact that $H + \lambda_1 I$ is a symmetric matrix. We can write this matrix inverse using an perturbative expansion for small λ_1:

$$\left(H + \lambda_1 I\right)^{-1} \approx H^{-1} - \lambda_1 H^{-2} + \mathcal{O}\left(\lambda_1^2\right). \tag{5.44}$$

Plugging this into the infidelity function, we get:

$$\Delta\mathcal{F}(\delta\overline{u}) = \frac{1}{2}\lambda_1^2 \overline{u}_*^T \left(H + \lambda_1 I\right)^{-1} H \left(H + \lambda_1 I\right)^{-1} \overline{u}_*, \tag{5.45}$$

$$= \frac{1}{2}\lambda_1^2 \overline{u}_*^T \left[H^{-1} - 2\lambda_1 H^{-2} + \mathcal{O}\left(\lambda^2\right)\right] \overline{u}_*, \tag{5.46}$$

In Figure 5.5, the relationship between the cost parameter λ_1 and the infidelity is depicted, which shows that the power law relationship is indeed quadratic at first order.

This result might generalise to general control functions $u(t)$ in the framework of quantum optimal control, which we discussed in Section 4.3. However, because the control functions are no longer piecewise constant, but instead smooth functions $u : [0, T] \to \mathbb{R}$, the linear algebra techniques used in this section are no longer directly applicable. It is unclear whether letting $n \to \infty$ for piecewise constant control functions is sufficient to approximate smooth control functions in this context.

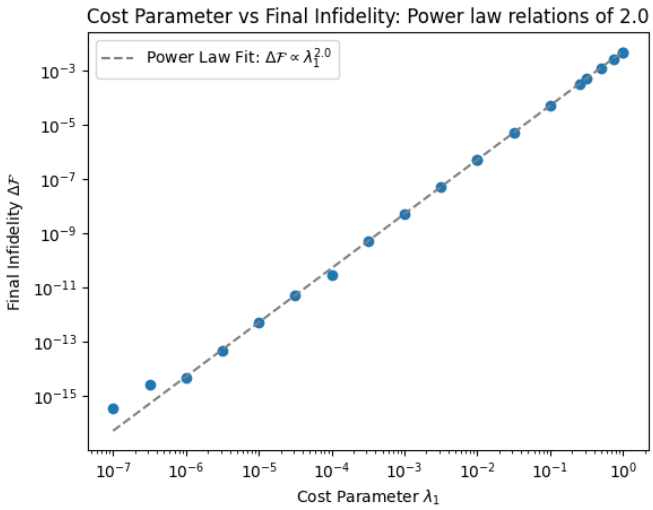

Figure 5.5: The infidelity of control pulses optimised with varying cost parameter values λ_1 is plotted. The same quantum control problem setup as in Section 5.3 is used. The figure shows that the power law relationship $\Delta\mathcal{F} \propto \lambda_1^2$, holds clearly, until around $\Delta\mathcal{F} = 10^{-15}$, which makes sense, since at point we are getting close to the precision of the Python float datatype, which can only distinguish numbers which are more than 10^{-16} apart.

5.5 Minimizing Rapid Changes in Control Amplitude

In real-world experiments, there are often constraints on the control pulse. In the previous section, we already discussed minimising the control pulse amplitude to ensure that the energy of the pulse is as small as possible whilst achieving the desired result. Another common experimental constraint is that the control pulse is not allowed to vary too quickly over time. We can minimise the rapid changes in the control amplitude by adding an additional term $\mathcal{C}_2 : \mathbb{R}^n \to \mathbb{R}^{\geq 0}$ to the cost function, which quantifies the total squared sum of the derivative:

$$\mathcal{C}_2(\overline{u}) = \frac{1}{2\Delta t} \sum_{k=1}^{n-1} (u_{k+1} - u_k)^2, \tag{5.47}$$

where the factor $\frac{1}{\Delta t}$ is added such that $\mathcal{C}_2(\overline{u})$ stays approximately constant as n is varied. Note that here we are penalising changes in control pulse amplitude, we are not directly penalising the existence of high-frequency terms: a low-frequency wave with a large amplitude might have a higher cost $\mathcal{C}_2$ than a high-frequency wave with low amplitude. However, high-frequency terms are indirectly suppressed this way.

The GOAT algorithm discussed in Chapter 6.1 will prove to be far better suited for constraining high-frequency terms, so we will not attempt more techniques to penalise high-frequency terms in the GRAPE algorithm.

The resulting total cost function $\mathcal{C} : \mathbb{R}^n \to \mathbb{R}^{\geq 0}$ is:

$$\mathcal{C}(\overline{u}) = \Delta\mathcal{F}(|\psi(T)\rangle) + \lambda_1\mathcal{C}_1(\overline{u}) + \lambda_2\mathcal{C}_2(\overline{u}), \tag{5.48}$$

with cost parameters $\lambda_{1,2} \geq 0$. Because the GRAPE algorithm uses a gradient descent algorithm, it also requires the derivative of the cost function $\mathcal{C}_2$:

$$\frac{\partial\mathcal{C}_2}{\partial u_k} = \frac{1}{\Delta t}(2u_k - u_{k+1} - u_{k-1}), \quad \text{for } k = 2, \cdots, n-1. \tag{5.49}$$

We have to be careful with the start and end of the control pulse: u_1 and u_n. We have three options. The simplest option is to not only compute the one-sided derivative at the endpoints:

$$\frac{\partial\mathcal{C}_2}{\partial u_1} = \frac{1}{\Delta t}(u_1 - u_2), \quad \frac{\partial\mathcal{C}_2}{\partial u_n} = \frac{1}{\Delta t}(u_n - u_{n-1}). \tag{5.50}$$

We can also assume that the pulse is periodic:

$$\frac{\partial\mathcal{C}_2}{\partial u_1} = \frac{1}{\Delta t}(2u_1 - u_2 - u_n), \quad \frac{\partial\mathcal{C}_2}{\partial u_n} = \frac{1}{\Delta t}(2u_n - u_{n-1} - u_1). \tag{5.51}$$

Or finally, can ensure a smooth turn-on and turn-off of the pulse by setting:

$$\frac{\partial\mathcal{C}_2}{\partial u_1} = \frac{1}{\Delta t}(2u_1 - u_2), \quad \frac{\partial\mathcal{C}_2}{\partial u_n} = \frac{1}{\Delta t}(2u_n - u_{n-1}), \tag{5.52}$$

which penalises any non-zero values of $u_{1,n}$. This is equivalent to setting the missing neighbouring control segment of $u_{1,n}$ to zero. This results in a relatively smooth turn-on and off.

In Figure 5.6, the effects of imposing the cost term $\lambda_2 = 0.1$ are displayed on the generic GRAPE setup discussed in Section 5.2. The resulting pulse is still smooth and looks similar to the pulse optimised using $\lambda_1 = 0.1$, except that it is shifted upwards and thinned out. This shows that the $\mathcal{C}_2$ had the desired effect. Both pulses have a small infidelity on the order of $\sim 10^{-4}$, which for practical purposes is more than good enough.

As highlighted in Figure 5.7, the one-sided derivative and periodic endpoint conventions produce the same results, which happened because the optimal pulse with the one-sided derivative convention already happened to be periodic. It could also happen that the optimal pulse starts and ends at zero as well, and that all three conventions produce the same results.

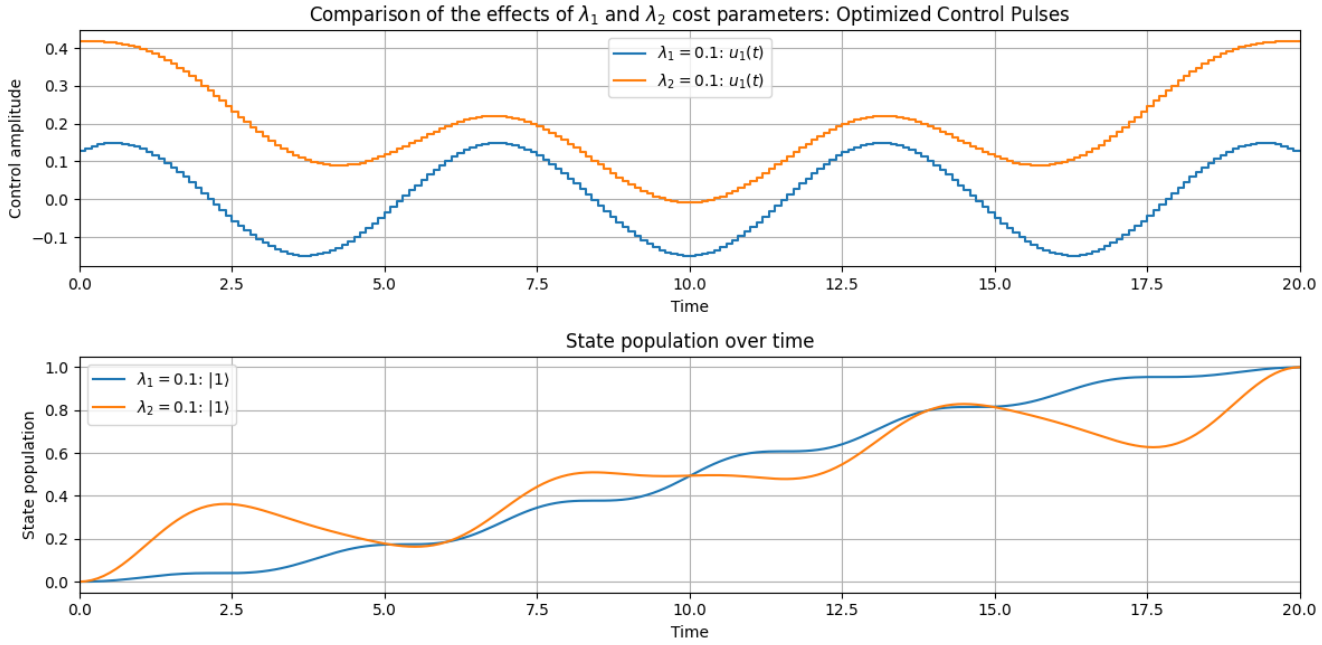

Figure 5.6: The GRAPE algorithm was applied to the quantum optimal control problem discussed in Section 5.2 for different values of $\lambda_{1,2}$. The blue pulse was optimised for infidelity and amplitude cost $\mathcal{C}_1$, the orange pulse was optimised for infidelity and derivative cost $\mathcal{C}_2$. The control values of the orange pulse are larger, but their variability is smaller, as expected. The corresponding infidelities are $1.1 \cdot 10^{-4}$ and $5.6 \cdot 10^{-5}$, the amplitude costs $\mathcal{C}_1$ are 0.116 and 0.501, and the derivative costs $\mathcal{C}_2$ are 0.106 and 0.063. The one-sided derivative convention is used for the endpoints.

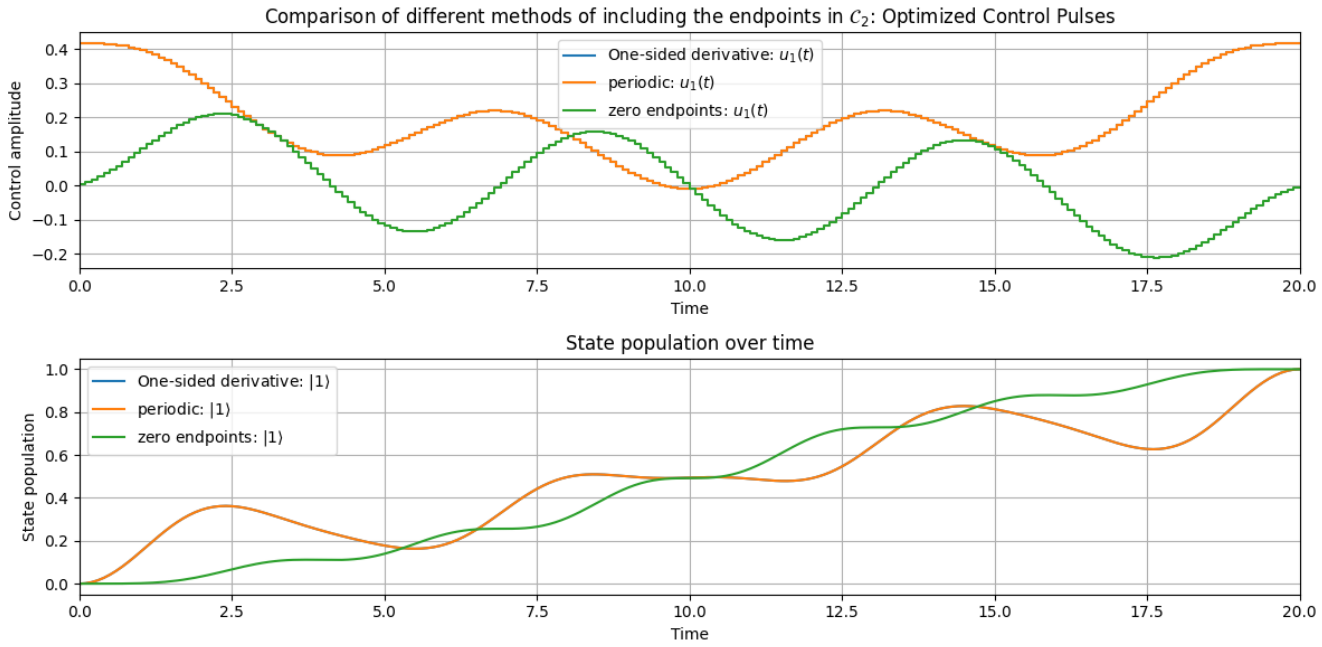

Figure 5.7: The GRAPE algorithm was applied to the quantum optimal control problem discussed in Section 5.2 with three different endpoint conventions. They have the same cost parameters: $\lambda_1 = 0$ and $\lambda_2 = 0.1$. You can see that the zero endpoints convention indeed produces a pulse which starts and ends at zero, and the periodic convention also produces a pulse which is periodic. The one-sided derivative convention produces a pulse which entirely overlaps with the one-sided pulse.

In Figure 5.8, a power law relationship between the cost parameter λ_2 and the final infidelity is displayed. It can be seen that over a range of about six orders of magnitude, a quadratic relationship holds very well. The analytic proof from Section 5.4 could potentially be adjusted to work with $\mathcal{C}_2$ instead of $\mathcal{C}_1$, by writing $\mathcal{C}_2(\overline{u}) = \overline{u}^T A \overline{u}$ for some matrix $A \in \mathbb{R}^{n \times n}$, however, such a derivation is outside the scope of this text.

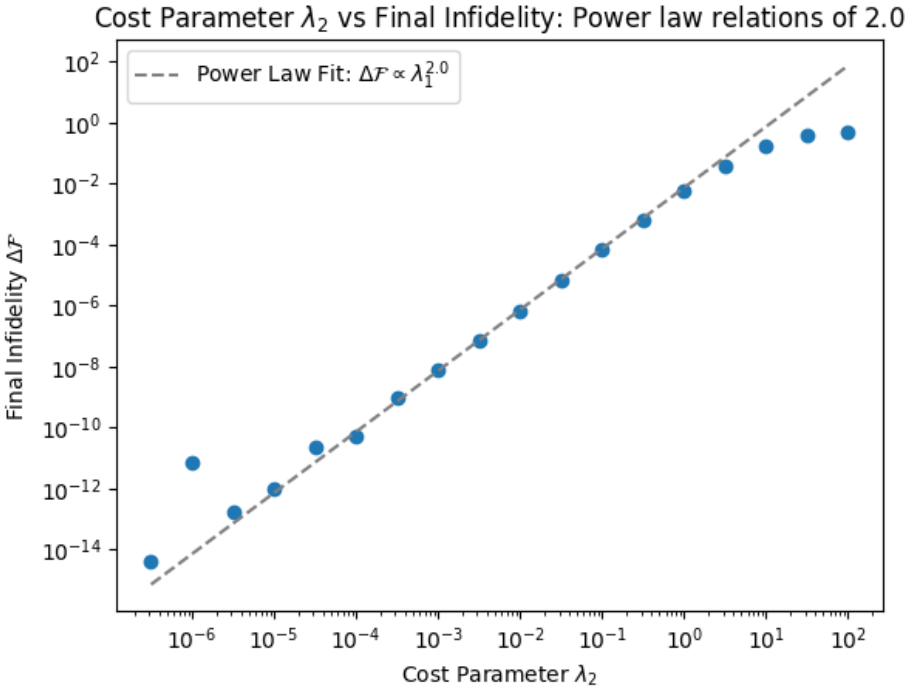

Figure 5.8: The same quantum control problem as described in Section 5.3 is used, with varying values of λ_2. The figure shows that the power law relationship $\Delta\mathcal{F} \propto \lambda_2^2$ holds clearly over approximately six orders of magnitude of λ_2.

5.6 Shape Functions

One of the other experimental constraints is that a control pulse requires a smooth turn-on and off. To ensure this constraint is satisfied for control pulses produced by the GRAPE algorithm, we can use shape functions $S : [0, T] \to \mathbb{R}^{>0}$, which can be incorporated into $\mathcal{C}_1$ as follows:

$$\mathcal{C}_1(\overline{u}) = \frac{\Delta t}{2} \sum_{k=1}^{n} \frac{u_k^2}{S(k\Delta t)}. \tag{5.53}$$

The shape function should be close to 1 in the middle of the interval $[0, 1]$, but should decrease when we get closer to $t = 0$ or $t = T$, such as to penalise large control amplitudes at $t = 0, T$. That means pulses benefit in terms of cost by having control amplitudes close to zero at the start and end. Our typical choice for the shape function is:

$$f_1(t) = \begin{cases} 1 + (\beta - 1)\left(\frac{t - t_{step}}{t_{step}}\right)^n & \text{if } t < t_{step} \\ 1 & \text{if } t_{step} \leq t \leq T - t_{step} \\ 1 + (\beta - 1)\left(\frac{T - t - t_{step}}{t_{step}}\right)^n & \text{if } t \geq T - t_{step} \end{cases} \tag{5.54}$$

with $t_{step} = \frac{T}{5}$ and $\beta = 0.1$. Other options are $S(t) = \sin^2\left(\pi \frac{t}{T}\right)$ or $S(t) = \exp\{-32(\frac{t}{T} - 1/2)^2\}$ [30]. For the former function, we can avoid division by zero issues by adding a small constant ϵ. In Figure 5.9, the plots of the two shape functions which we are going to use are displayed.

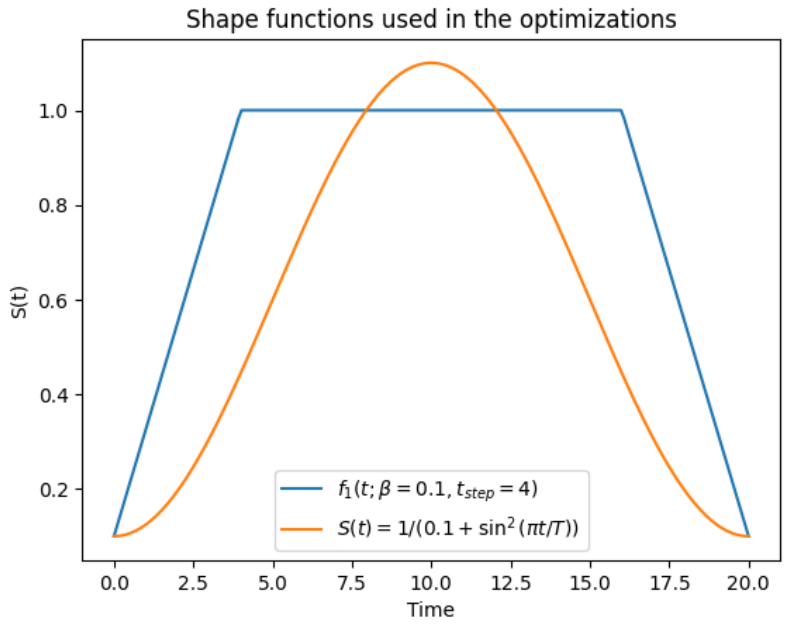

Figure 5.9: Plots of the two shape functions which are used for GRAPE optimisation

In Figure 5.10, the GRAPE optimisation results using various shape functions is displayed. You can see that the shape functions indeed have the desired effect and smooth turn-on and off is achieved.

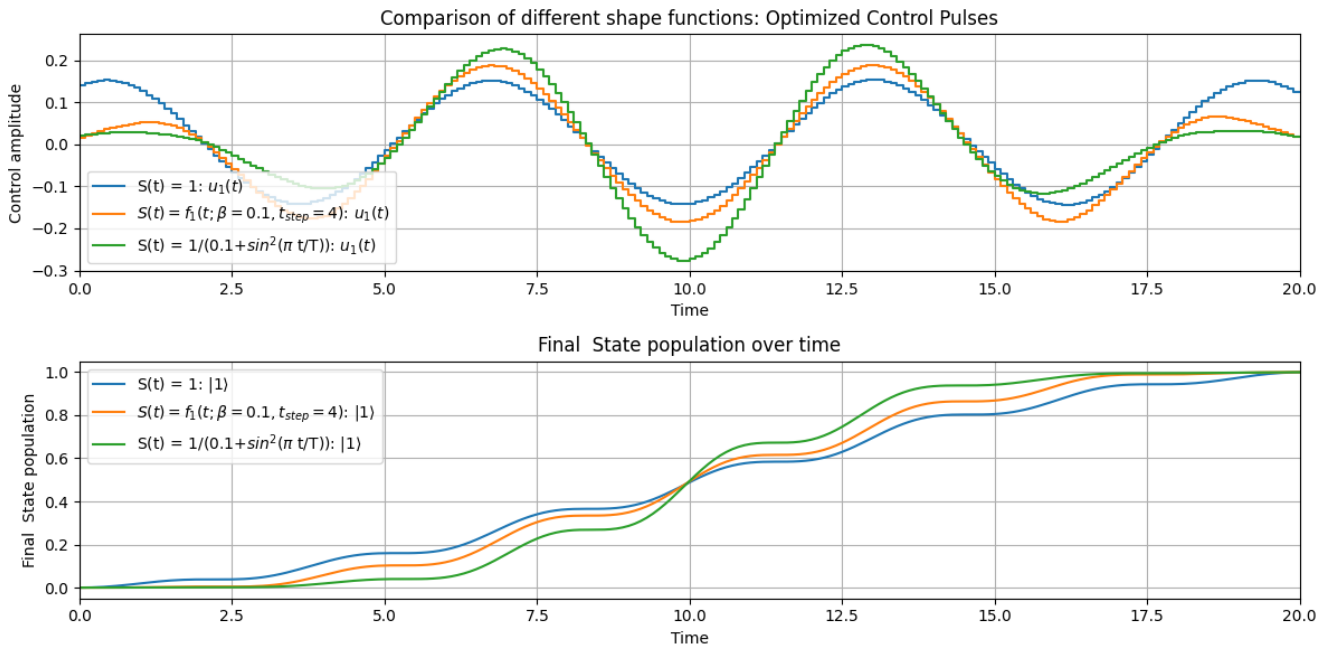

Figure 5.10: The GRAPE algorithm was applied to the quantum optimal control problem discussed in Section 5.2 for three different shape functions. For each optimisation, $\lambda_1 = 0.1$ was used. For the shape function $f_1(t)$, we used the parameters $\beta = 0.1$ and $t_{step} = 4$. Compared to the $S(t) = 1$ case, the other choices of the shape functions do achieve their aim of ensuring a smooth turn-on and off.

5.7 Controllability Example

We will now briefly showcase an example of a system which satisfies the conditions of the quantum controllability Theorem 4.5.2. The following four-level system has just one control Hamiltonian:

$$\hat{H}_0 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}, \quad \hat{H}_1 = \begin{bmatrix} 0 & 1 & 1/2 & 1/4 \\ 1 & 0 & 1 & 1/2 \\ 1/2 & 1 & 0 & 1 \\ 1/4 & 1/2 & 1 & 0 \end{bmatrix}. \tag{5.55}$$

It can be shown that the Lie Algebra generated by $\{-i\hat{H}_0, -i\hat{H}_1\}$ satisfies the conditions of the quantum controllability theorem; therefore, it should be able to reach any state given enough time. We use the GRAPE algorithm to find the control pulse which brings the system from the ground state to the highest excited state $|4\rangle$ at the final time $T = 16$.

If you pick λ_1 to be too small, it will get stuck too soon in a nearby zero-fidelity solution, instead of exploring more of the control landscape and finding another zero-infidelity solution with significantly lower control amplitude cost $\mathcal{C}_1(\overline{u})$. Therefore, the state transfer process is more chaotic-looking than necessary, as depicted in Figure 5.11.

However, if you pick λ_1 to be too large, it will converge to the global minimum of the cost function, but the infidelity might be significant, and the state transfer process will look wonky as seen in Figure 5.12.

We can use a trick to ensure that we get the best possible convergence, first we pick $\lambda = 0.1$, as depicted in Figure 5.12 and then we use the resulting control pulse as the initial guess for a GRAPE optimisation, with a significantly smaller $\lambda_1 = 10^{-4}$, the final result, as depicted in Figure 5.13, is a very symmetric-looking state transfer process with low infidelity and a smooth control pulse. With this method, you can find a local minimum of the control cost function under the constraint of zero infidelity.

For this method to be guaranteed to work, we need the fact that there are no sub-optimal local minima of the infidelity, as discussed in Section 4.6. Because that fact ensures that you can always find a zero-infidelity solution using a gradient descent algorithm, of course, that might not be the zero-infidelity solution which is nearest to the initial guess, but when the infidelity of the initial guess is small, like in the case with $\lambda = 0.1$, then the zero-infidelity solution produced by the gradient descent algorithm is likely rather close by.

You can see that all these control pulses have some degree of periodicity. The internal frequencies of the system are given by the difference in energy levels of the system, since the only energy differences are $1, 2$ and 3, we should see periodic behaviour with periods of $T = 2\pi, \pi, \frac{2}{3}\pi$. Which makes sense because the time differences between the pulse maxima with $u \sim 0.3$ are 6.1 and 6.5, which seems to agree with a periodicity of $T = 2\pi$ which is suggested by the energy levels.

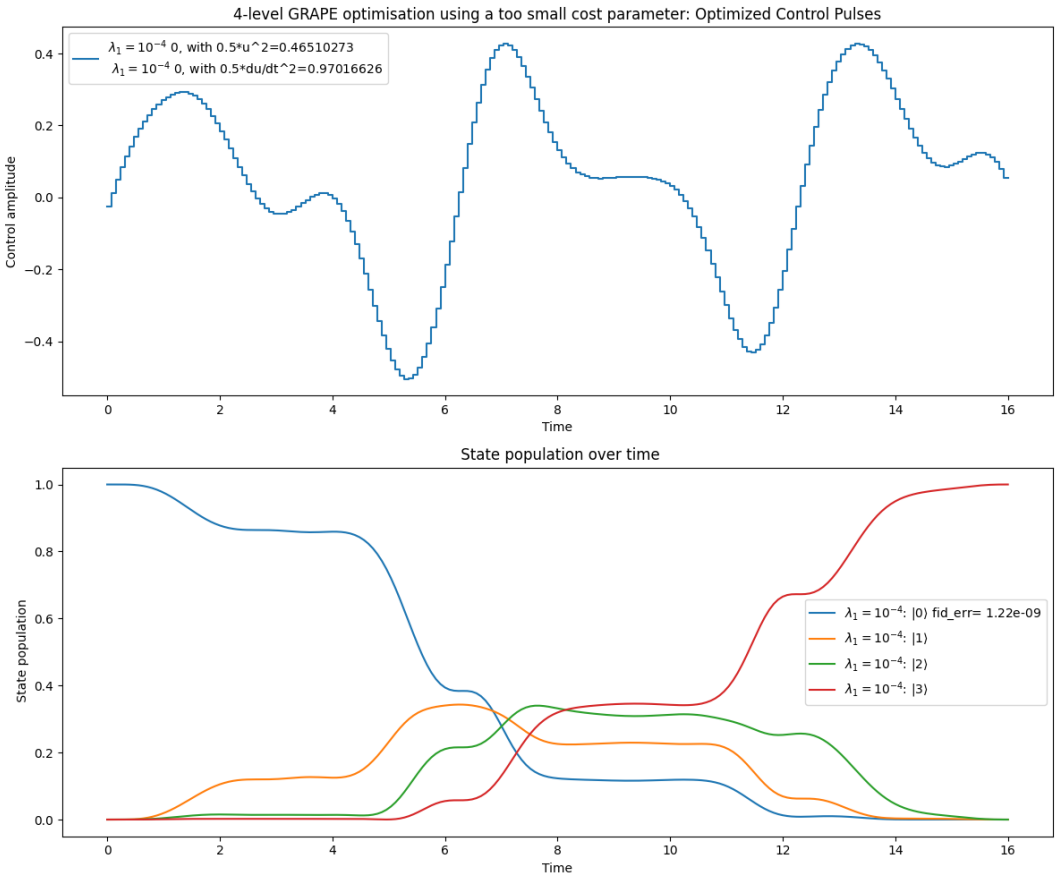

Figure 5.11: The GRAPE algorithm was performed using the control Hamiltonian (5.55) and $\lambda_1 = 10^{-4}$. The cost parameter turns out to be too small since there exists another zero-infidelity solution with lower control amplitude cost, which can be found using repeated GRAPE optimisation as shown in Figure 5.13.

This suggests that there might be a minimal required time for the system to transfer the system from the ground state to the highest excited state. In appendix B. It can be seen that around the final time $T = 2$, the optimisation starts to fail, even if you don't impose a cost parameter λ_1 on the optimisation and you leave the GRAPE algorithm unconstrained to optimise for infidelity.

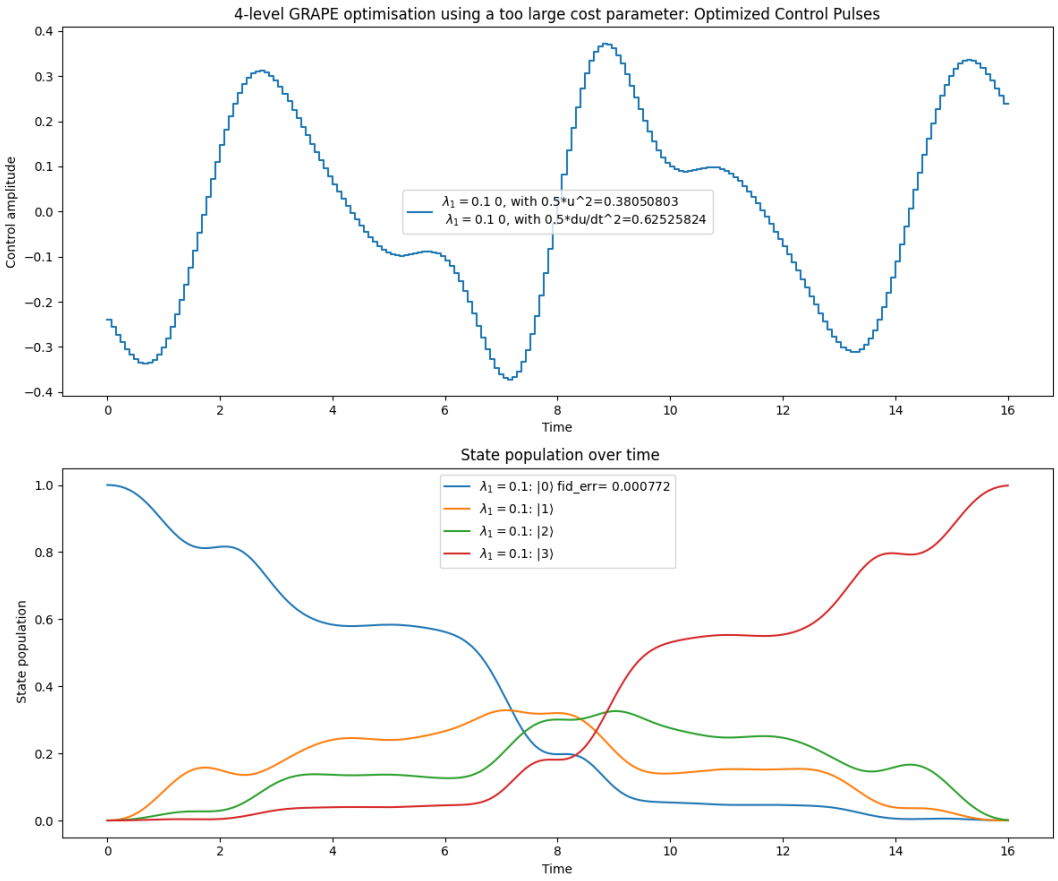

Figure 5.12: The GRAPE algorithm was performed using the control Hamiltonian (5.55) and $\lambda_1 = 0.1$. The control amplitude cost $\mathcal{C}_1(\overline{u})$ is significantly lower than in Figure 5.11. However, the state transfer process looks slightly asymmetric, suggesting that there is some trade-off between control amplitude cost and infidelity.

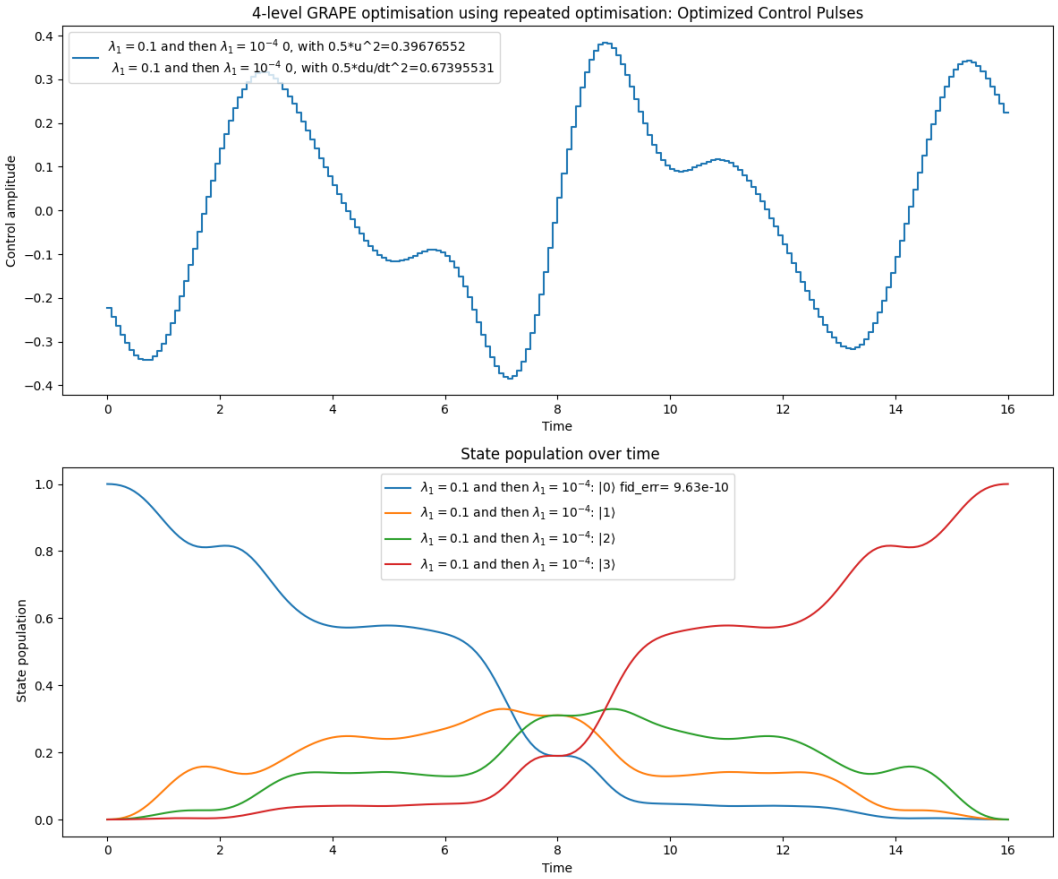

Figure 5.13: The GRAPE algorithm was performed using the control Hamiltonian (5.55) and $\lambda_1 = 0.1$. Then another GRAPE optimisation is performed, but now with the previous optimised control pulse as the initial guess and $\lambda_1 = 10^{-4}$. The control pulse also looks smooth, and the state transfer process is also highly symmetric along the $T = 8$ vertical line and has very small infidelity; those are indications that the cost function has been successfully optimised (at least on a local level), because an arbitrary control pulse would be extremely unlikely to contain such characteristics as well. However, the state transfer still looks very similar to the state transfer in Figure 5.12, which suggests that the second GRAPE optimisation with $\lambda_1 = 10^{-4}$ did not change the pulse too much.

Chapter 6

GOAT Algorithm

6.1 Gradient Optimization of Analytic Controls (GOAT) Algorithm

The GOAT algorithm is a gradient-based optimisation method which allows you to optimise the parameters of analytic control pulses. This means that you have more control over the shape of the optimised control pulse than through GRAPE, and the number of optimised parameters can be far lower. The method was first introduced in [41], which we will also follow for our presentation of the mathematics behind the GOAT algorithm.

Let the Hilbert space be $\mathcal{H} = \mathbb{C}^{N_d}$, and assume that the Hamiltonian $\hat{H} : [0,T] \times \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_{N_c}} \to (\mathbb{C}^{N_d} \to \mathbb{C}^{N_d})$ to be of the form:

$$\hat{H}(t; \overline{\alpha}_1, \cdots \overline{\alpha}_{N_c}) = \hat{H}_0 + \sum_{i=k}^{N_c} f_k(t; \overline{\alpha}_k)\hat{H}_k \tag{6.1}$$

with parametrised control functions $f_k : [0,T] \times \mathbb{R}^{d_k} \to \mathbb{R}$. For simplicity, we will again set $N_c = 1$. However, $d_1 \in \mathbb{N}$ will be left unrestricted. This control Hamiltonian differs from the Hamiltonian used for the GRAPE algorithm, because instead of the control functions being piecewise constant, they are now smooth functions which are parametrised by $\overline{\alpha}$. An example of a possible parametrised control function is $f : [0,T] \times \mathbb{R}^3 \to \mathbb{R}$, which is a sine wave parametrised by $\overline{\alpha}$:

$$f(t; \overline{\alpha}) = \alpha_1 \sin(\alpha_2 t + \alpha). \tag{6.2}$$

The GOAT algorithm works to minimise the infidelity:

$$\Delta\mathcal{F}(\overline{\alpha}) := 1 - |\, \langle \psi_{\text{target}} | \hat{U}(T; \overline{\alpha}) | \psi_0 \rangle \,|^2, \tag{6.3}$$

where $|\psi_{target}\rangle$ and $|\psi_0\rangle$ are the target and initial states, respectively. The operator $\hat{U}(T; \overline{\alpha})$ is the evolution operator, which is given by:

$$\hat{U}(T; \overline{\alpha}) = \mathbb{T}\left[\exp\left(\int_0^T -\frac{i}{\hbar}\hat{H}(t; \overline{\alpha})dt\right)\right]. \tag{6.4}$$

In the above expression, $\mathbb{T}$ is the time-ordering operator and plays a crucial role. To compute $\hat{U}(T;\overline{\alpha})$, you have to use a power series expansion for the matrix exponential. As the length of this series grows, the approximation improves[1]. However, because when the Hamiltonian is evaluated at different times, it doesn't commute with itself:

$$\left[\hat{H}(t_1), \hat{H}(t_2)\right] \neq 0, \tag{6.5}$$

the order of the Hamiltonian operators matters. To resolve this, we introduce the time-ordering operator, which ensures that operators are chronologically ordered: the operator corresponding with $t = 0$ occurs first, and the operator corresponding with $t = T$ is at the end of the multiplication.

Since we are working in a finite-dimensional Hilbert space, we can view operators as matrices, for which the necessary addition, integration, and exponentiation operations are well-defined.

The GOAT algorithm contains two parts. An efficient computation of $\partial_{\alpha_k}(\Delta\mathcal{F})(\overline{\alpha})$, and a generic gradient descent optimisation method over the parameter space, generic in the sense that we can use the same numerical optimisation method for both GRAPE and GOAT. The gradient descent algorithm updates the parameters as follows:

$$\alpha_k' = \alpha_k - \epsilon \frac{\partial(\Delta\mathcal{F})}{\partial\alpha_k}(\overline{\alpha}), \tag{6.6}$$

for $k = 1, \cdots, d_1$. Define $A := \langle\psi_{\text{target}}|\hat{U}(T;\overline{\alpha})|\psi(0)\rangle$, then we have:

$$\partial_{\alpha_k}(\Delta\mathcal{F})(\overline{\alpha}) = -[\langle\psi_{\text{target}}|\partial_{\alpha_k}\hat{U}(T;\overline{\alpha})|\psi(0)\rangle\, A^* + \langle\psi(0)|\partial_{\alpha_k}\hat{U}^\dagger(T;\overline{\alpha})|\psi_{\text{target}}\rangle\, A] \tag{6.7}$$

$$= -2\,\text{Re}\Big\{A^*\,\langle\psi_{\text{target}}|\partial_{\alpha_k}\hat{U}(T;\overline{\alpha})|\psi(0)\rangle\Big\}. \tag{6.8}$$

There is, however, no closed-form expression for $\partial_{\overline{\alpha}}\hat{U}(t;\overline{\alpha})$. Because $\hat{U}(t;\overline{\alpha})$ is the time evolution operator and we are working in the Schrödinger picture, the operator $\hat{U}(t;\overline{\alpha})$ evolves according to the Schrödinger equation, this can be shown by plugging in the identity: $|\psi(t)\rangle = \hat{U}(t;\overline{\alpha})\,|\psi(0)\rangle$ and then dropping the ket $|\psi(0)\rangle$, leading to:

$$\partial_t\hat{U}(t;\overline{\alpha}) = -\frac{i}{\hbar}\hat{H}(t;\overline{\alpha})\hat{U}(t;\overline{\alpha}). \tag{6.9}$$

However, it can be expressed as the solution of the differential equation:

$$\partial_t \begin{pmatrix} \hat{U} \\ \partial_{\overline{\alpha}}\hat{U} \end{pmatrix} = -\frac{i}{\hbar} \begin{pmatrix} \hat{H} & 0 \\ \partial_{\overline{\alpha}}\hat{H} & \hat{H} \end{pmatrix} \begin{pmatrix} \hat{U} \\ \partial_{\overline{\alpha}}\hat{U} \end{pmatrix} \tag{6.10}$$

The equation for $\partial_t\partial_{\overline{\alpha}}\hat{U}(t;\overline{\alpha})$ is given by

[1]For brevity, we will omit the proof of the convergence of the power series expansion.

$$\partial_t \partial_{\overline{\alpha}} \hat{U}(t;\overline{\alpha}) = \partial_{\overline{\alpha}} \partial_t \hat{U}(t;\overline{\alpha}) = \partial_{\overline{\alpha}} \left(-\frac{i}{\hbar} \hat{H}(t;\overline{\alpha}) \hat{U}(t;\overline{\alpha}) \right) \tag{6.11}$$

$$= -\frac{i}{\hbar} \left[\partial_{\overline{\alpha}} \hat{H}(t;\overline{\alpha}) \hat{U}(t;\overline{\alpha}) + \hat{H}(t;\overline{\alpha}) \partial_{\overline{\alpha}} \hat{U}(t;\overline{\alpha}) \right] \tag{6.12}$$

Additionally, $\partial_{\overline{\alpha}} \hat{H}(t;\overline{\alpha}) = \partial_{\overline{\alpha}} f_1(t;\overline{\alpha}) \hat{H}_1$ can be computed by applying the chain rule to equation (6.1). Using the differential equation (6.10), we can then find a solution for $\partial_{\overline{\alpha}} \hat{U}(t;\overline{\alpha})$, which we can use for the derivative of the fidelity (6.7), which in turn can be used for the gradient descent algorithm.

Summarising, the main steps of the GOAT algorithms are:

1. Pick an initial guess $\overline{\alpha}_0$
2. Initiate a gradient-driven search to minimize $\Delta\mathcal{F}(\overline{\alpha})$, which requires the computation of $\hat{U}$ and $\partial_{\overline{\alpha}} \hat{U}$ for specific values of $\overline{\alpha}$.
3. The values of $\hat{U}$ and $\partial_{\overline{\alpha}} \hat{U}$ are computed by a numerical forward integration of equation (6.10). For which numerical integration methods like Runge-Kutta methods with adaptive step sizes may be used.

Constraints on the parameters $\overline{\alpha}$ can also be easily added to this problem through the gradient descent algorithm.

6.2 GOAT Optimisation of Two-Level System

We start with a simple example of a sine wave control function, which will bring a simple two-level system from the ground state to the excited state. The Hamiltonians are given by:

$$\hat{H}_0 = -\frac{1}{2}\hat{\sigma}_z = \begin{bmatrix} -\frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}, \quad \hat{H}_1 = \hat{\sigma}_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \tag{6.13}$$

With the sine wave control function being $f : \mathcal{U} \to ([0,T] \to \mathbb{R})$:

$$f(t;\overline{\alpha}) = \alpha_1 \sin(\alpha_2 t + \alpha_3), \tag{6.14}$$

$$\mathcal{U} := [-1,1] \times [-10,10] \times [-\pi,\pi]. \tag{6.15}$$

with $\mathcal{U}$ being the set of allowed parameters for the sine wave. These constraints help the convergence of the algorithm by limiting the parameter space to only control pulses whose amplitude and frequency are small enough to smoothly transfer the system from the ground state to the excited state.

In preparation for using the GOAT algorithm on open quantum systems with jump operators which represent coherence decay, we will move to density matrices to describe the quantum state

of the system. We start in the state $\hat{\rho}_0 = |0\rangle\langle 0|$ and our target state is $\hat{\rho}_{target} = |1\rangle\langle 1|$. This means that we also have to change the way we compute the infidelity of the state transfer. We have to move to the trace difference method:

$$\mathcal{F} = \frac{1}{2}\,\mathrm{Tr}\left[(\hat{\rho}_{target} - \hat{\rho}(T))^{\dagger}(\hat{\rho}_{target} - \hat{\rho}(T))\right] := \frac{1}{2}||\hat{\rho}_{target} - \hat{\rho}(T)||^2_{\mathrm{td}}. \tag{6.16}$$

We can show that this trace is positive real, by setting $\hat{\sigma} := \hat{\rho}_{target} - \hat{\rho}(T) = \sum_{i,j}\sigma_{ij}\,|i\rangle\langle j|$. Writing it out, we get:

$$\mathcal{F} = \frac{1}{2}\,\mathrm{Tr}\left[\left(\sum_{k,l}\sigma_{kl}\,|k\rangle\langle l|\right)^{\dagger}\sum_{i,j}\sigma_{ij}\,|i\rangle\langle j|\right], \tag{6.17}$$

$$= \frac{1}{2}\,\mathrm{Tr}\left[\sum_{k,l}\sigma^*_{kl}\,|l\rangle\langle k|\sum_{i,j}\sigma_{ij}\,|i\rangle\langle j|\right], \tag{6.18}$$

$$= \frac{1}{2}\,\mathrm{Tr}\left[\sum_{l,i,j}\sigma^*_{il}\sigma_{ij}\,|l\rangle\langle j|\right], \tag{6.19}$$

$$= \frac{1}{2}\sum_{i,j}\sigma^*_{ij}\sigma_{ij} = \frac{1}{2}\sum_{i,j}|\sigma_{ij}|^2 \geq 0. \tag{6.20}$$

Therefore, the trace-difference method is equivalent to the squared sum of the matrix elements of $\hat{\sigma} := \hat{\rho}_{target} - \hat{\rho}(T)$, just written in a coordinate-free way.

Unless otherwise specified, we let the integrator pick its own time steps We set an infidelity target of 10^{-10}. For the gradient descent algorithm, we used the BFGS algorithm, and we set the number of gradient descent iterations to 100.

We have heavily constrained the number of parameters in the optimal control problem, reducing it to three parameters instead of ≥ 100 parameters in the GRAPE algorithm for the piecewise constant pulses. On the one hand, this makes the problem more tractable since you have fewer parameters and more control over the resulting optimised control pulse.

However, we have become very restricted in the control pulses which we can choose, and therefore, Theorem 4.5.2 on quantum controllability becomes invalid. This means that we can no longer determine beforehand whether any zero-infidelity solutions will exist. Additionally, it is no longer true that there exist no suboptimal local minima of infidelity. That means that the gradient descent algorithm can get stuck in minima whose infidelity is not zero.

To combat this issue, we employ non-local optimisation algorithms. The QUTIP-qoc package supports both the Basin-hopping and Dual Annealing algorithms. We will use the former. The basin-hopping algorithm first performs a local optimisation. It records these solutions and then selects a perturbed version of the previous optimisation result as the initial guess, and then performs the gradient descent algorithm again. It does for multiple iterations, and the globally optimised

result is the locally optimised result, which has the lowest objective cost. In our setup, we set the number of basin-hopping iterations to 100.

We can also include a cost function for the parameters, such that we also penalise excessively large choices of parameters, and find the zero-infidelity optimisation solution which has the smallest possible choice of parameters.

$$\mathcal{C}_1(\overline{\alpha}) = \sum_{k=1}^{d_1} \frac{1}{2}\lambda_{1,k}\frac{\alpha_k^2}{\alpha_{k,max}^2}. \tag{6.21}$$

In this optimisation setup, we pick $\overline{\lambda}_1 = (\frac{1}{10}, 0, 0)$, since we only want to impose a cost on the amplitude of the sine wave, not its frequency or phase. We already know that the optimal frequency of the sine wave should be $\alpha_2 = 1$, and the starting phase of the sine wave is not of importance to us.

We will start with one control Hamiltonian:

$$\hat{H}(t;\overline{\alpha}) = -\frac{1}{2}\hat{\sigma}_z + f(t;\overline{\alpha})\hat{\sigma}_x. \tag{6.22}$$

The results of the GOAT optimisation are presented in figure 6.1. We can see that the state transfer process is not one smooth line, but has many plateaus where no state transfer occurs. To remedy this, we introduce a second control Hamiltonian $\hat{H}_2 = \hat{\sigma}_y$, which will help to smooth these plateaus out and achieve a more efficient state transfer.

We will now work with two control Hamiltonians:

$$\hat{H}(t;\overline{\alpha}_1, \overline{\alpha}_2) = -\frac{1}{2}\hat{\sigma}_z + f(t;\overline{\alpha}_1)\hat{\sigma}_x + f(t;\overline{\alpha}_2)\hat{\sigma}_y. \tag{6.23}$$

Based upon our discussion on π-pulses in Subsection 2.7, we expect that:

$$f(t;\overline{\alpha}_1) + if(t;\overline{\alpha}_2) = Ae^{i\omega(t+\phi)}, \tag{6.24}$$

with $\omega = 1$ and $A = \frac{\pi}{2T} = 0.0625$. It can be seen that this corresponds very well with the optimised control parameters presented in Figure 6.2. The infidelity of the control pulse predicted by theory is surprisingly large: $\Delta\mathcal{F} \approx 2 \cdot 10^{-2}$; however, this error starts to decrease when you increase the final time T, suggesting that this infidelity comes from the fact that the rotating wave approximation does not exactly hold.

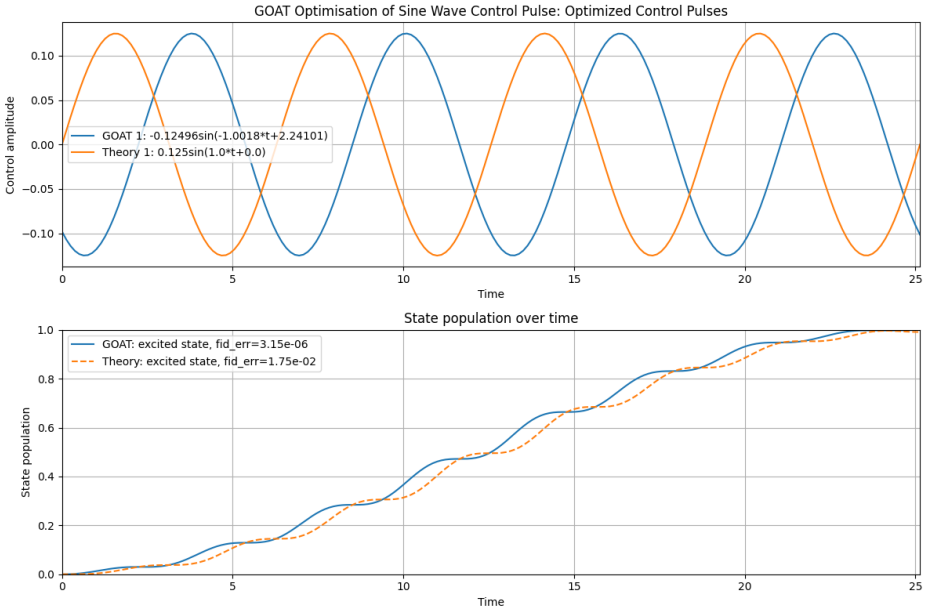

Figure 6.1: We use the control cost parameter vector: $\overline{\lambda}_1 = \left(\frac{1}{10}, 0, 0\right)$ and set the final time $T = 8\pi$. The optimised control parameters are: $(-0.12496, -1.0018, 2.24101)$. For comparison, we also plot the sine wave predicted by π-pulse theory, which produces a similar state transfer. The infidelities of both control pulses are small: $\sim 3 \cdot 10^{-6}$ and $\sim 2 \cdot 10^{-2}$, respectively.

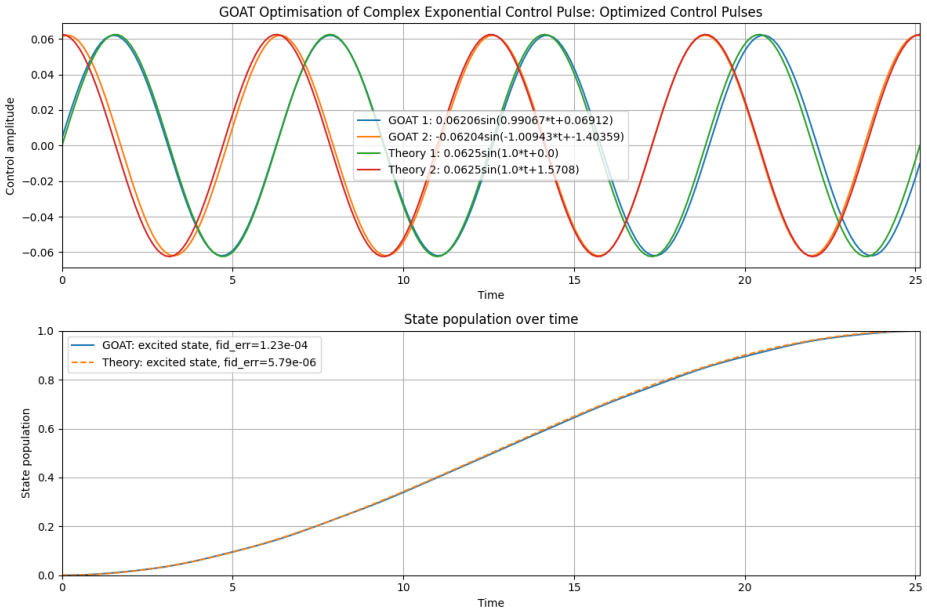

Figure 6.2: We use the control cost parameter vectors: $\overline{\lambda}_1 = \overline{\lambda}_2 = \left(\frac{1}{100}, 0, 0\right)$ and set $T = 8\pi$. The optimised control parameters are very close to the expected values, and the phase difference of $\frac{\pi}{2}$ between the two sine waves is also present as predicted. The infidelity is also very low at $\Delta\mathcal{F} \approx 10^{-4}$. In orange, the two waves are plotted with the parameters predicted in equation (6.24). The shifted phase has no significant effect on the dynamics.

Chapter 7

Double-Pulse Excitation Experiments

7.1 Introduction

In this chapter, we will work on simulations of double-pulse excitation experiments. In such experiments, a single molecule, which we model as a two-level quantum system, is hit by two pulses, where the second pulse is delayed by some delay time Δt compared to the first pulse.

Our primary sources of experimental measurements will be: [42] and [43]. Based on this data, unknown parameters of the system can be inferred. In those papers, those parameters were inferred by fitting possible simulation curves by hand. However, the main focus of these papers was to show that it is experimentally possible to manipulate and observe the quantum states of single molecules on femtosecond time scales and at room temperature.

We will use the GOAT algorithm to provide a gradient-based optimisation of the fit to enable us to determine the range of parameters which are compatible with the experimental measurements. We will consider multiple extensions of the two-level quantum system discussed in the original papers and compare the quality of the resulting fits. Because we use the GOAT algorithm to provide the fits, we can easily fit the data using multiple different models.

The experimental setup consists of single molecules of terrylenediimide (TDI) dispersed in a film of polymethyl methacrylate (PMMA). The laser will then be near-resonant to a purely electronic transition of the TDI. Because it is resonant to one electronic transition, we will initially model it as a two-level system; however, we will later also describe it using a three-level system.

In this chapter and the following chapter, we will use femtosecond units, one femtosecond is equal to 10^{-15} seconds.

The experimental laser pulse consists of two Gaussian pulses, with a time delay between them. We also have control over the relative phase of the two pulses. Generally, they will have the same phase. The pulse width was fixed in the experiment as $\tau_p = 75$ fs, our convention for the definition of pulse width is given in equation (7.2).

In this text, we will only discuss the basics of the experiment and will not delve into the details of the experiment, which can be found in [42] and [43].

In the experiments, single TDI molecules were excited by phase-locked double pulse sequences from their ground state to an excited state. Here, phase-locked means that the two pulses have the same frequency, and thus their phase difference remains constant over time; we assume this phase difference is zero, unless otherwise specified. We will model this process using a two-level quantum system. However, in reality, there are many energy levels close together, but only a few will be close to being resonant with the frequency of the laser pulse.

One experimental complication is that the orientation of the TDI molecule in space is randomised, meaning that we don't know the orientation of its electronic dipole moment, which means we can't predict the parallel electric field amplitude $E_{||}$, which determines the strength of the interaction between the laser pulse and the molecule. This means that the interaction strength parametrised by the Rabi frequency $\omega_{R,max}$ of the laser pulse will become a fitting parameter in the next chapter. However, since the plots of excited state population versus delay time, the so-called delay traces, were performed on a single molecule, the Rabi frequency is the same for each data point in the delay trace.

Typical time delays range between 0 and 600 fs for the experimental data, so we will follow that in the simulation work. There is also population decay in the system. The excited state lifetime is 3.5 ns, which is an order of magnitude larger than the timescales of the system dynamics and is therefore ignored. This excited state decay emits photons, which are measured in the experiment. These photon emissions are linearly correlated with the excited state population after interaction with the laser pulse.

One distinctive aspect of this experiment is that it was carried out at room temperature, which was unique at the time. At such relatively high temperatures, the coherence of a quantum state quickly decays, causing the system to completely lose the memory of the first pulse by the time that the second pulse typically arrives. However, thanks to the use of the femtosecond-length pulses, this experiment successfully measured the coherence decay lifetime of a single molecule.

This coherence decay is caused by electronic dephasing, which is caused by interactions with the environment. It causes a loss of coherence between the ground and excited state, because their energy levels are disturbed by the environment. These disturbance means that the phase relationship between the two states becomes uncertain, which is modelled by an exponential decay of the u, v variables in Equation (2.87) or by adding a collapse operator to the quantum mechanical system. This coherence decay means that subsequent pulses are less efficient at transferring the system from one state to the other, because these pulses' effect depends on being in-phase with the system.

7.2 Mathematical Model

The first model, which we will discuss, is a two-level system with a decay parameter for coherence loss. It was also used as the theoretical model in the experimental papers. In later sections, we will expand this model. The system Hamiltonian $\hat{H}(t) = \hat{H}_0 + \hat{H}_1(t)$ is given by:

$$\hat{H}_0 = -\frac{\Delta}{2}\hat{\sigma}_z, \quad \hat{H}_1(t) = -\frac{1}{2}\Omega(t)\hat{\sigma}_x, \tag{7.1}$$

where Δ is the detuning parameter, which indicates the frequency offset between the laser pulse and the transition frequency, the Hamiltonian corresponds to the one discussed in Section 2.7. The function $\Omega : \mathbb{R} \to \mathbb{R}$ is the pulse envelope function and consists of the sum of two Gaussian pulses, it is given as:

$$\Omega(t) = \omega_{R,max}\left[e^{-t^2/2\tau_p^2} + e^{-(t-\Delta t)^2/2\tau_p^2}\right], \tag{7.2}$$

where τ_p is the pulse width, and Δt is the time delay between the first and second pulse. The parameter $\omega_{R,max}$ is the peak Rabi frequency achieved by the pulse when the detuning Δ is zero, when $\Delta \neq 0$, then the Rabi frequency shifts, as given in equation (2.89).

In our simulations, we will set $\tau_p = 31$ fs, corresponding with the pulses used in the experiments, which are described as having a FWHM (Full Width at Half Maximum) of 75 fs.

Because of interactions with the environment, there is a loss of coherence of the two-level system due to electronic dephasing, which is represented with the dephasing collapse operator:

$$\hat{c}_{dephasing} = \sqrt{\frac{1}{T_2}}\hat{\sigma}_z, \tag{7.3}$$

where T_2 is the coherence lifetime, typical values of the coherence lifetime are 400 to 800 fs, and one goal of the experiments is to determine this lifetime by fitting the simulations with the delay traces produced by the experiments. Collapse operators were introduced in Section 2.10.

The system starts in the ground state and is then hit by the two laser pulses, with a time delay Δt in between. We can extend the rotating wave approximation equations (2.87), to include electronic dephasing as follows by including an exponential decay of the coherences $\rho_{1,2}(t)$ with a rate of $\frac{1}{T_2}$:

$$\frac{\mathrm{d}}{\mathrm{d}t}\vec{\rho}(t) = \begin{bmatrix} -\frac{1}{T_2} & -\Delta & 0 \\ +\Delta & -\frac{1}{T_2} & \mu A(t) \\ 0 & -\mu A(t) & 0 \end{bmatrix}\vec{\rho}(t). \tag{7.4}$$

7.3 Simulations

We will now showcase some simulations of the model described in this chapter to further explain the effects of the detuning and dephasing parameters. We will start with a simple π-pulse. All the

simulations in this section are done using QuTiP-based Python code, with the Hamiltonian given in equation (7.1).

Gaussian π-pulse

We will start by constructing a simple $|0\rangle \rightarrow |1\rangle$ transfer using a single Gaussian pulse. The detuning Δ of the system is set to zero, and electronic dephasing is not included in the simulations. We know that the pulse area should follow equation (2.66):

$$\int_{-\infty}^{\infty} \frac{1}{2}\Omega(t)dt = \pi \tag{7.5}$$

Using that the integral of a Gaussian pulse is:

$$\int_{-\infty}^{\infty} e^{-t^2/2\tau_p^2}dt = \sqrt{2\pi}\tau_p, \tag{7.6}$$

we can show that the peak Rabi frequency $\omega_{R,max}$ of the pulse to ensure a smooth transfer to excited state should be:

$$\omega_{R,max} = \sqrt{\frac{\pi}{2}}\frac{1}{\tau_p}. \tag{7.7}$$

Which is verified in Figure 7.1. The first subfigure shows the time evolution of the expectation values of the Pauli matrices, which relate to u, v, w in the optical Bloch equation (2.87). The second subfigure shows that the system is smoothly fully transferred to the excited state $|1\rangle$. The third subfigure shows the pulse amplitude of the Gaussian pulse over time. Restricting the simulation to $\pm$200 fs around the peak does not meaningfully affect the system dynamics.

Final excited state population versus Detuning

For a constant pulse with height $\omega_{R,0}$, our discussion on the rotating wave approximation in Section 2.7 provides an analytic solution for the maximal excited state population given the detuning (2.90):

$$P_{ext}(\Delta) = \frac{\omega_{R,0}^2}{\omega_{R,0}^2 + \Delta^2}. \tag{7.8}$$

This equation agrees with our simulations, as depicted in Figure 7.2. It must be emphasized that we plot the maximal excited state population achieved during the time interval and not the excited state population at final time T, because the frequency of population inversion also changes depending on the detuning Δ as given in equation (2.89):

$$\omega_{R,\Delta} = \sqrt{\omega_{R,0}^2 + \Delta^2}. \tag{7.9}$$

Figure 7.2 depicts what the maximal excited state population achieved during a constant pulse is. The simulations completely agree with equation (7.8). However, the Rabi frequency itself shifts because of the detuning according to equation (7.9), the excited state population at $T = 100$ fs is lower than the maximum, unless $\omega_{R,\Delta}$ is a multiple of $\frac{\pi}{100}$.

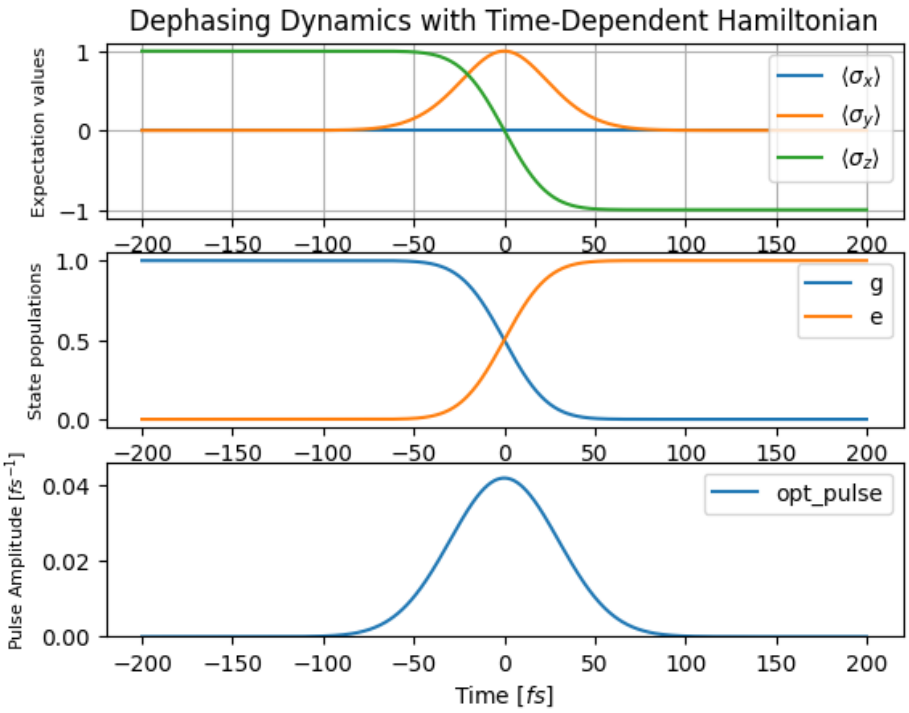

Figure 7.1: Example of a Gaussian pulse which almost perfectly transfers the state to the excited state. The maximum Rabi frequency is $\omega_{R,max} = \sqrt{\frac{\pi}{2}}\frac{1}{\tau_p} \approx 0.042$ fs^{-1}, with $\tau_p = 30$ fs. The final infidelity is $\Delta\mathcal{F} \approx 4 \cdot 10^{-5}$.

However, we have no such analytic solution for Gaussian pulses when the length of the Gaussian pulse is similar to the characteristic timescale for detuning, which is $T_{detuning} = \frac{2\pi}{\Delta}$. Because the pulse is no longer constant, it becomes harder to quantify the effect of the pulse becoming out of phase with the system. Another complication is that the Rabi frequency $\omega_{R,\Delta}$ becomes time-dependent since the Rabi frequency $\omega_{R,0}$ has a Gaussian profile.

In Figure 7.3, we show the effect of detuning on the maximal excited state population for a Gaussian π-pulse and we have a graph of equation (7.8) as well, to show where this approximation is valid, however, whereas it is easy to ascribe a singular pulse amplitude to a constant pulse, this is more difficult for a Gaussian pulse. For a π-pulse, the peak amplitude is $\sqrt{\frac{\pi}{2}}\frac{1}{\tau_p}$, as given in equation (7.7). As an ansatz, we can pick $\frac{1}{\tau_p}$ to represent the 'average' pulse amplitude of a π-pulse. Leading to the following formula for the excited state population:

$$P_{exc}(\Delta) = \frac{1}{1 + \Delta^2\tau_p^2}. \tag{7.10}$$

It can be seen in the Figure 7.3 that it provides a good approximation for small values of detuning. In Figure 7.4, it is shown that it is a good approximation up to $\Delta \approx 0.03$ fs^{-1}, which is sensible because at that point the detuning gets to a similar scale as $\frac{1}{\tau_p}$ and the detuning means that the latter half of the Gaussian pulse becomes out of phase.

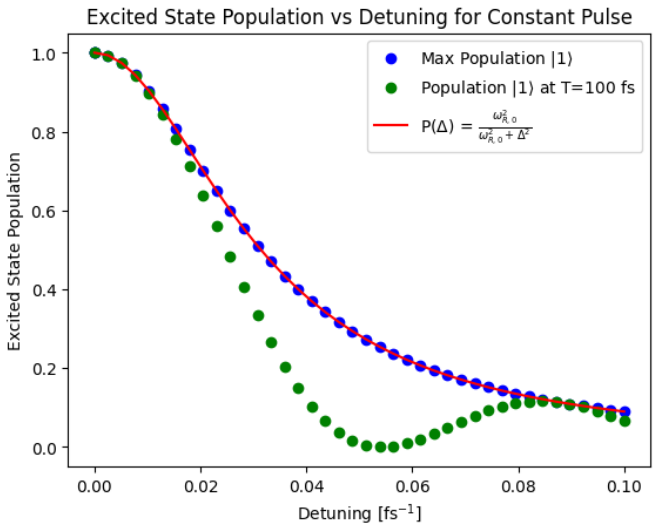

Figure 7.2: The simulation is the same as before, except that the pulse is now constant, $\Omega(t) = \frac{\pi}{T}$, with $T = 100$ fs. We see that the maximal excited state population achieved during this 100 fs interval aligns perfectly with the theoretical predictions. Note that the excited state population at the final time $T = 100$ fs is lower because the population inversion frequency depends on the detuning Δ. The green and blue curves meet at $\Delta \sim 0.088$ fs^{-1}. For that detuning, $P(\Delta) = 3\omega_{R,0}$, meaning that peaks for at $T' = \frac{100}{3}$ fs and then peaks for a second time at $T = 100$ fs.

Coincidentally, the Taylor series of $f(x) = \frac{1}{1+x^2}$ centred at $x = 0$ also has a radius of convergence of 1, which might be related to the divergence between the experimental data and equation (7.10) around $\Delta = \frac{1}{\tau_p}$.

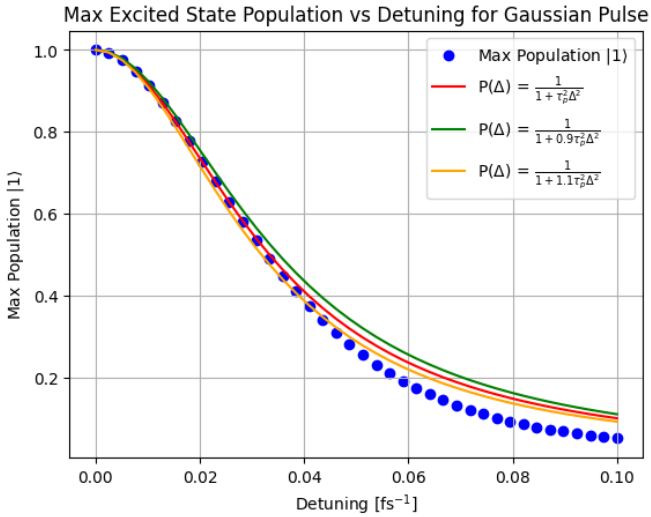

Figure 7.3: The experimental data are shown in blue, together with three fitted curves. The middle fit, shown in red, is best for small detunings $\Delta < 0.04$ fs^{-1}. The other two fits demonstrate that varying the fitting parameter does not improve the fit. The fitting errors of each curve are shown in Figure 7.4.

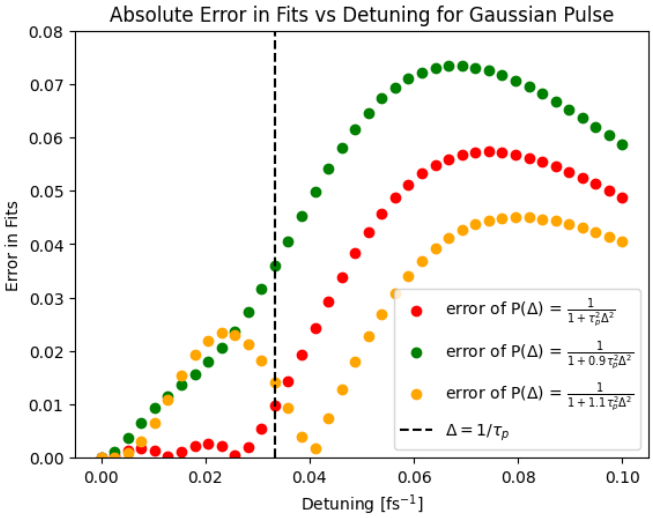

Figure 7.4: The absolute errors of each fit are displayed. The error of the red fit is consistently low up to around $\Delta = \frac{1}{\tau_p}$, where the fit starts to fail. The other fits fail at smaller detunings. The orange fit happens to be better for high detunings because its function values are smaller than those of the red fit.

Electronic Dephasing

We will now examine the effect of including electronic dephasing into our simulations. Electronic dephasing leads to coherence loss between the ground and excited state and is modelled using the following collapse operator:

$$\hat{c}_{dephasing} = \sqrt{\frac{1}{T_2}}\hat{\sigma}_z, \tag{7.11}$$

where T_2 is the coherence lifetime, electronic dephasing leads to control pulses becoming less effective at changing the state populations of the system, because the phase relations between the ground and excited states are smeared out. For example, Rabi oscillations will get damped exponentially with a rate of $\frac{1}{2T_2}$, as depicted in Figure 7.5.

Typical values of T_2 in the experimental data are on the order of 100s of femtoseconds, which is significantly larger than the pulse width τ_p of our Gaussian pulses, meaning that during the Gaussian pulse, electronic dephasing will not be significant. In the next section, we will discuss the effect of electronic dephasing when a delayed double Gaussian pulse is used.

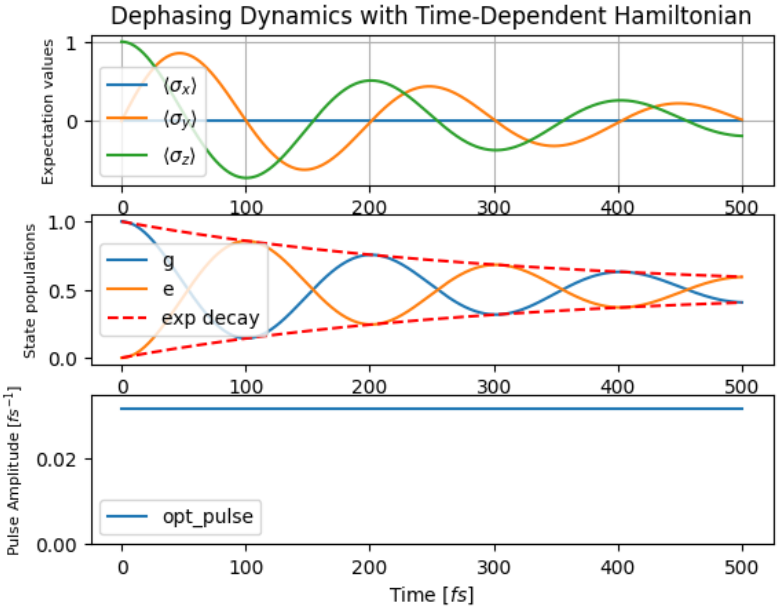

Figure 7.5: A constant pulse $\Omega(t) = \frac{\pi}{100}$ interacts with a two-level system with zero detuning. The electronic dephasing lifetime is $T_2 = 150$ fs. As can be seen in the middle plot, the amplitude of the Rabi oscillations is well described by exponential decay through the function: $f(t) = \frac{1}{2}\left[1 \pm \exp(-t/2T_2)\right]$.

7.4 Delay Traces

In experiments, so-called pump and probe pulses are often used, with the excited state population after the final pulse being recorded. By varying the delay between the pump and probe pulse, you

can create a graph of the excited state population as a function of this delay. Such experiments allow you to determine unknown parameters of the system, and they will be our focus in the next chapter on parameter identification.

We parametrise the delayed double Gaussian pulse as follows:

$$\Omega(t) = \omega_{R,max}\left[e^{-t^2/2\tau_p^2} + e^{-(t-\Delta t)^2/2\tau_p^2}\right], \tag{7.12}$$

with $\omega_{R,max}$ the peak Rabi frequency and $\Delta t > 0$ being the delay between the two pulses. The first pulse is not strongly affected by the electronic dephasing, but for long enough decay times, the second pulse will be very diminished in efficiency as the coherences of the system will have decayed, as shown in Figure 7.5. It depends on the state of the system after the first pulse, how that will affect the final state of the system: If the first pulse puts the system almost completely in the excited state, then having a long delay for the second pulse, such that its effect is diminished, ensures that the system remains mostly in the excited state.

On the flip side, if the first pulse only puts the system half in the excited state, then a very short delay for the second pulse ensures that its effect on the system is large enough to finish the job and put the system mostly in the excited state.

One common way to visualise such effects is through delay traces, which graph the final excited state population as a function of the delay Δt between the two pulses. These graphs are our primary source of experimental data and will become the centrepiece of the next chapter on parameter identification.

Let's start with the delay trace of a system, as described in Section 7.2, with parameters $\Delta = 0.03$ fs^{-1}, $\tau_p = 30$ fs and $\omega_{R,max} = \frac{1}{2}\sqrt{\frac{\pi}{2}}\frac{1}{\tau_p} \approx 0.021$ fs^{-1}, meaning that each Gaussian pulse has an area of $\frac{\pi}{2}$ and thus puts the system halfway to the excited state in the situation of zero detuning. The delay trace of this system is depicted in Figure 7.6.

The period between peaks corresponds to $\frac{2\pi}{\Delta} \approx 210$ fs. One curious aspect is that the peaks at $\Delta t = 0$ fs are smaller than their respective peaks at $\Delta t \approx 200, 400$ fs. This is because the pulse amplitude is larger at $T = 0$ fs as the two pulses completely overlap. Using equation (7.10), we get that the predicted maximal excited state population is ~ 0.55, which corresponds with the simulation data.

When the two pulses no longer overlap, the situation becomes more complicated, because the peak pulse amplitude is now lower, equation (7.8) suggests that the effects of detuning should become even more significant. However, because the second pulse arrives when the laser and the system are momentarily in phase together again, this means that the effects of the detuning are severely lessened and the pulse transfer becomes more efficient.

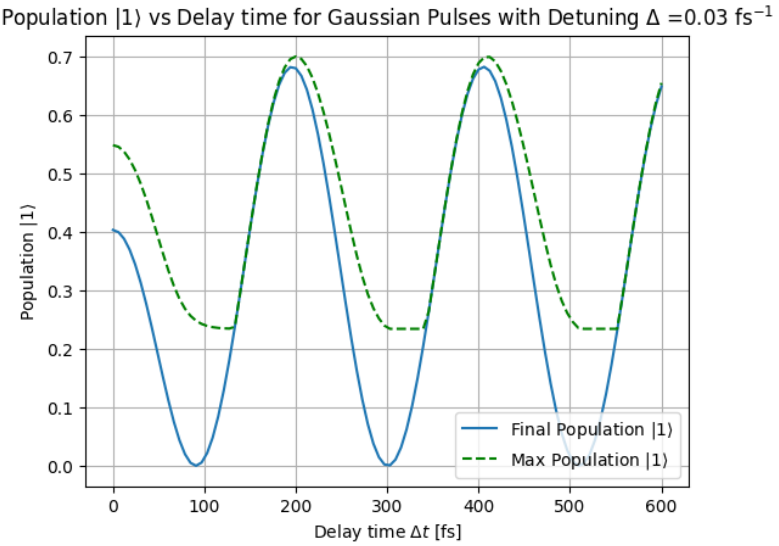

Figure 7.6: Delay trace of the system described in Section 7.2, with parameters $\Delta = 0.03$ fs^{-1}, $\tau_p = 30$ fs and $\omega_{R,max} = \frac{1}{2}\sqrt{\frac{\pi}{2}}\frac{1}{\tau_p} \approx 0.021$ fs^{-1}. Each Gaussian pulse has an area of $\frac{\pi}{2}$. The trace shows oscillations with a period of $\frac{2\pi}{\Delta} \approx 210$ fs.

Delay Trace: Electronic Dephasing

Here we have a system as described in Section 7.2, with electronic dephasing lifetime $T_2 = 300$ fs and detuning $\Delta = 0$ fs^{-1}. You can see that it neatly follows an exponential decay after the delay between the two pulses becomes large enough such that they can be considered separate.

For small values of Δt, the assumption that the dephasing primarily happens during the delay between the two pulses falls apart, because the dephasing which occurs during the pulses becomes relevant as well.

7.5 Conclusion

In this chapter, we began by examining the effects of parameters such as detuning and electronic dephasing lifetime on two-level systems dynamics, by first considering the case of a constant pulse before moving to Gaussian pulses. Afterwards, we moved to the delay traces which are produced by double-pulse excitation experiments, which will form the foundation of the next chapter, where we will aim to identify the values of parameters like the detuning and the electronic dephasing lifetime based upon the delay traces themselves by introducing a parameter identification algorithm, which employs the GOAT algorithm.

One aspect which we have not discussed in this chapter, but which will become important, is experimental noise and modelling uncertainties. In the next chapter, these will play a far larger role.

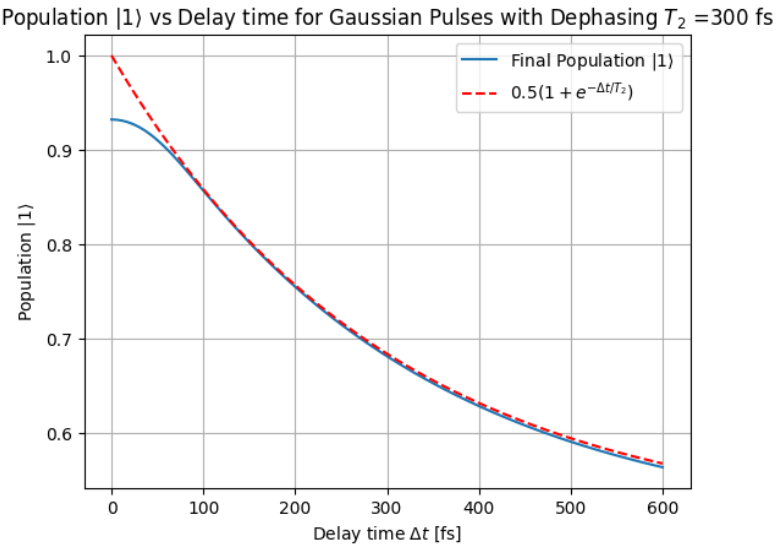

Figure 7.7: Delay trace of the system as described in Section 7.2, with parameters $T_2 = 300$ fs, $\Delta = 0$ fs^{-1}, $\tau_p = 30$ fs and $\omega_{R,max} = \frac{1}{2}\sqrt{\frac{\pi}{2}}\frac{1}{\tau_p} \approx 0.021$ fs^{-1}. Each Gaussian pulse has an area of $\frac{\pi}{2}$. The trace confirms that the exponential decay rate is indeed 300 fs.

Chapter 8

Parameter Identification Using GOAT Algorithm

We will now look into parameter identification of quantum systems using the GOAT algorithm. Parameter identification is the process of using experimental data to determine the system Hamiltonian and its related decay parameters. This process might seem unconnected to optimal control; however, they employ similar methods to achieve different goals. Quantum control aims to find the system Hamiltonian which takes the system to some pre-specified state. We can therefore use a quantum control algorithm like GOAT to determine which pulses can reproduce the experimental data. In the paper [44], the well-posedness of Hamiltonian Identification problems and the uniqueness of their solutions are discussed.

In 2003, [45] discussed the idea of optimal identification, where the goal is to design experiments in such a way that the data is as sensitive as possible to the parameter you wish to identify as well. Because of experimental noise and/or different sets of system parameters producing similar experimental data, situations might occur where a parameter can not be restricted to a small possible interval, in Figure 8.4, such a situation is depicted.

Oftentimes, the aim in optimal identification is to find the control function which maximises the quantum Fisher information, which is a metric of the sensitivity of the quantum state to a parameter. The formula for the Fisher information for a control function $u : [0, T] \to \mathbb{R}$ and a Hamiltonian $\hat{H}(t; \alpha)$, with $\alpha \in \mathbb{R}$, has the following equation, for the Fisher information centred at α_0:

$$F[u](\alpha_0) = \lim_{\delta\alpha \to 0} \frac{4}{(\delta\alpha)^2} \left|\left|\hat{\rho}_{\alpha_0}(T; u) - \hat{\rho}_{\alpha_0+\delta\alpha}(T; u)\right|\right|_{\text{td}} . \tag{8.1}$$

Where T is the final time and we use the trace-difference norm given in equation (6.16). In this chapter, we will not directly work with quantum Fisher information; however, based on our simulations, we can numerically estimate the quantum Fisher information.

We will take pre-existing experimental data, so the quantum Fisher information is not maximised for every parameter, and we might find that we can not conclusively determine a parameter based

upon the available data. However, this in itself is valuable information, because if you try to determine the parameters of a system by fitting a simulated curve to the experimental data yourself, it is difficult to quantify how sensitive this fit is to perturbations of the parameter, and you might draw false conclusions.

We will use the GOAT algorithm for parameter identification; without it, it would be impossible to do parameter identification for the experimental data we have, because in the experiment, Gaussian pulses were used, and we have their peak Rabi frequency $\omega_{R,max}$ as a fitting parameter. With the GRAPE algorithm, it is rather complicated to ensure that the optimised control pulse has a Gaussian profile. To the knowledge of the author, no applications of the GOAT algorithm for parameter identification are present in the literature.

The usage of the GOAT algorithm has some possible benefits, such as being able to perform a gradient-based optimisation of the fit between the simulated and experimental data. However, due to the relatively few control parameters, which through the control pulse parameterisation are all mutually dependent, the gradient-based algorithm often fails to converge to a global optimum. Therefore, global optimisation algorithms such as the basin-hopping algorithm have to be used; this requirement partly negates the benefits of having access to gradient information. Especially in Section 8.4, such issues with convergence arise.

Another potential benefit is that the GOAT algorithm employs parameterised control pulses, meaning that it can be used for parameter identification after the experiment has been performed. In contrast, in the optimal identification method, the quantum Fisher information of the control pulse is first optimised using an algorithm. Then the optimised control pulse is implemented in an experiment, which means that this method can not be used retroactively, and it might be difficult to ensure that this optimised control pulse can be implemented in an experiment. However, the potential downside of using the GOAT algorithm without also designing the experiment to have a high quantum Fisher information is that even in conditions where experimental noise is negligible and the underlying model is known, a parameter still can not be determined with high accuracy; this issue is illustrated in Section 8.3.

However, even with these difficulties, the parameter identification algorithm does allow for a very systemic way of finding good fits to delay traces, because a large amount of the human bias involved with fitting a curve by hand are removed, additionally, it also becomes easier to quantify to what degree a fit and its associated parameters are unique in producing a fit with such an infidelity. In the Sections 8.5 and 8.6, fits are presented that fit within the error bars caused by experimental noise.

8.1 Mathematical Model

We will use the GOAT algorithm to identify the parameters which best fit the experimental delay traces reported in [42] and [43]. That means that we will run multiple quantum optimal control problems simultaneously, one for each data point of the delay trace, and minimising the sum of

their cost functions.

We will consider the following three-level quantum system as a model to explain the experimental data. However, by setting restrictive bounds to some control parameters, we can effectively turn it into a two-level quantum system.

We have the control pulse basis Hamiltonians:

$$\hat{H}_{12} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \hat{H}_{13} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}. \tag{8.2}$$

In this model, we consider the laser to be near-resonant with both the $|1\rangle \rightarrow |2\rangle$- and $|1\rangle \rightarrow |3\rangle$-transitions, so the same control pulse can drive two different transitions, the Hamiltonians parametrising the detuning of the laser pulse compared to the $|1\rangle \rightarrow |2\rangle$ and $|1\rangle \rightarrow |3\rangle$-transitions are:

$$\hat{H}_{\text{detuning,1}} = \begin{bmatrix} -\frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} \end{bmatrix}, \quad \hat{H}_{\text{detuning,2}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \tag{8.3}$$

where the second detuning Hamiltonian is used to create an energy level difference between the second and third levels and is expected to be small compared to the energy level splitting between the first level and the other levels.

The control function for the delayed double Gaussian pulse is:

$$f(t; \omega_{R,max}, \Delta t) = \omega_{R,max} \left[\exp\left(-\frac{(t - T/4)^2}{2\tau_p^2} \right) + \exp\left(-\frac{(t - T/4 - \Delta t)^2}{2\tau_p^2} \right) \right]. \tag{8.4}$$

We assume that this pulse is resonant with both the $|1\rangle \leftrightarrow |2\rangle$- and $|1\rangle \leftrightarrow |3\rangle$-transitions to some degree, which will be regulated by their respective Rabi frequencies ω_{12} and ω_{13}. Ideally, this should also be related to the energy level splittings between the second and third levels. However, in the following sections, this relationship between detuning and Rabi frequency will not be enforced.

The first pulse peaks at $t = \frac{T}{4}$, with $T = 1000$ fs, and this is done such that our simulations can run from $t = 0$ to $t = T$.

All of these Hamiltonians combine into the following system Hamiltonian:

$$\hat{H}(t, \Delta t; \omega_{12}, \omega_{13}, \Delta_1, \Delta_2) = f(t; \omega_{12}, \Delta t)\hat{H}_{12} + f(t; \omega_{13}, \Delta t) + \Delta_1 \hat{H}_{\text{detuning,1}} + \Delta_2 \hat{H}_{\text{detuning,2}}. \tag{8.5}$$

With $(\omega_{12}, \omega_{13}\Delta_1, \Delta_2)$ thus being the control parameters of the problem for which the GOAT algorithm will optimise.

The collapse operators related to electronic dephasing and vibrational decay are:

$$\hat{c}_{12} = \sqrt{\gamma_2} \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \hat{c}_{13} = \sqrt{\gamma_2} \begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \hat{c}_{vib} = \sqrt{\frac{1}{\tau_{vib}}} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}. \tag{8.6}$$

The vibrational decay collapse operator $\hat{c}_{vib}$ leads to an exponential decay from $|3\rangle$ to $|2\rangle$; vibrational decay is possible because the excited states are assumed to be close in energy, suggesting they can be considered to be a part of the same vibrational energy band.

In our model, we don't care whether the state is in $|2\rangle$ or in $|3\rangle$; we assume that both lead to the same emission after they decay to the ground state $|1\rangle$. We don't attach a cost function to the control parameters, since we are only trying to identify the original physical parameters. We will put bounds on the control parameters to restrict the search space of the optimisation algorithm; however, the bounds should be chosen such that the optimal solution lies within the bounds.

The combined cost function is:

$$\mathcal{C} = \sum_{\Delta t \in A} (P_{|0\rangle}(\Delta t) - |\langle 0|\psi(T;\Delta t)\rangle|)^2, \tag{8.7}$$

with the set A containing all of the decay times which we include in our parameter identification algorithm, and $P_{|0\rangle}(\Delta t) := |\langle 0|\psi_{data}(T;\Delta t)\rangle|$ the ground state population inferred from the experimental data. This cost function enables us to quantify the quality of the fit of the simulated delay trace with the experimental data.

However, the decay parameters γ_2 and τ_{vib} can't be part of the GOAT optimisation, because we don't have access to the derivative $\frac{\partial(\Delta\mathcal{F})}{\partial\gamma_2}$ through the GOAT algorithm. However, in [46], a GRAPE-style algorithm is discussed, which does allow for gradient-based decay parameter optimisation. This suggests that it could perhaps be possible to extend the GOAT algorithm to allow for decay parameter optimisation; however, no such extension is available in the literature.

Therefore, we will use a simple grid search method to find the optimal combination of decay parameters. For every combination (γ_2, τ_{vib}), we will then perform a GOAT optimisation to find the best possible fit by changing the remaining parameters. This can then be used to produce a heat map, which shows which decay parameter combinations best fit the data. We will put constraints on the control parameters $(\omega_{12}, \omega_{13}, \Delta_1, \Delta_2)$, to help the basin-hopping algorithm find a reasonable optimum quicker.

This grid search method is rather resource-intensive, since we have to perform $N_{\tau_{vib}} \cdot N_{\gamma_2} \cdot N_{\Delta t}$ different GOAT optimisations. Therefore, we used the Nieuwpoort cluster of the University of Groningen to perform these optimisations in parallel. Using the cluster, the most resource-intensive of these plots still took less than a day using two CPUs with fourteen cores each.

8.2 Benchmarking the Parameter Identification Algorithm

We will first apply the parameter identification (PI) algorithm to fit simulated data. This will firstly let us check that the algorithm actually identifies the correct parameter, and secondly, it lets us see the sensitivity of the fit to certain parameters. It could be that increasing a parameter by 100% barely affects the fit, whilst changing another parameter by 10% would produce significant effects.

We can then look at whether the PI algorithm produces the same pattern of parameter sensitivity on the experimental data. Any mismatch could either be caused by the noise of the experimental data washing out the fingerprints of a certain parameter or a sign that our model has too many degrees of freedom, meaning that we can't decide which combination of parameters is the most likely.

One nice thing about the PI algorithm is that it takes away the human skill when fitting by hand, allowing us to also find optimal fits with more complicated models. It also allows us to more quantitatively decide which models explain the experimental data the best.

Our benchmark model follows the model discussed at the start of this chapter with the following parameters:

$$\omega_{12} = 0.01 \text{ fs}^{-1}, \quad \omega_{13} = 0.056 \text{ fs}^{-1}, \tag{8.8}$$

$$\Delta_1 = 0 \text{ fs}^{-1}, \quad \Delta_2 = 0.0048 \text{ fs}^{-1}, \tag{8.9}$$

$$\gamma_2 = 0.008 \text{ fs}^{-1}, \quad \tau_{vib} = 500 \text{ fs}. \tag{8.10}$$

In Figure 8.1, the delay trace of the benchmark model is shown. One of its interesting characteristics is the initial dip in excited-state population, which was selected because it is similar to left side of the experimental delay trace which is examined in Section 8.6.

We will now use the parameter identification algorithm to find the parameters of the benchmark system. We can choose how many parallel optimisations we want to perform: how many data points we want to take from the delay trace to sum to create the total objective function. Below we perform the PI algorithm with seven, ten and fourteen data points.

We can see that in Figure 8.2, the optimal parameters which the algorithm found are slightly off; however, the correct decay parameters are still part of the low-infidelity region, in the heatmaps generated from fitting with ten and fourteen data points, depicted in Figures 8.3 and 8.4, the correct parameters are identified. However, due to the grid search method with a grid of 41 by 41 elements, the algorithm only has a limited amount of choices for τ_{vib} and γ_2 that it can consider.

In all three heat maps, it is clear that there is a very sensitive dependence on the vibrational decay time τ_{vib}, but very much not so on the electronic dephasing γ_2. Because of the absence of experimental noise, the algorithm still identifies the correct value of γ_2. However, if this had been a real experiment, it would have needed to have been designed in such a way as to produce

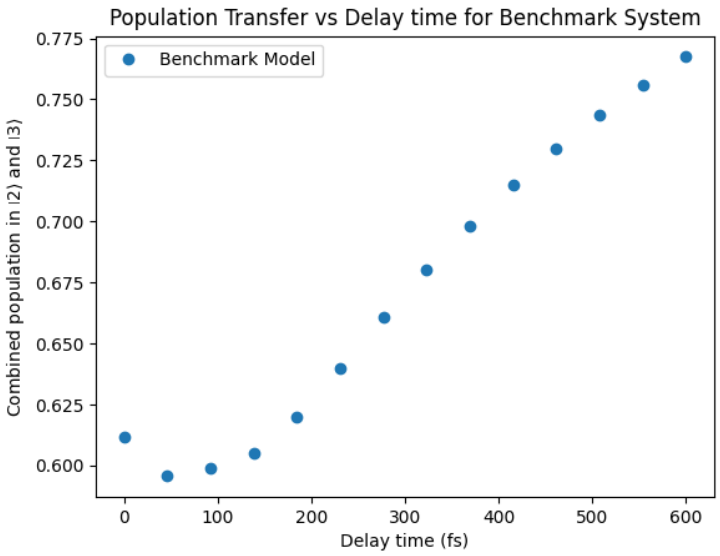

Figure 8.1: Delay trace of the benchmark model using the parameters specified in equations (8.8)-(8.10). Data points are sampled at 50 fs increments.

a higher quantum Fisher information with respect to γ_2 to ensure that we could identify γ_2 with reasonable precision. These plots suggest that having ten or more data points ensures that the correct parameters are identified. In the following sections, we will use fourteen data points to ensure no issues occur from not sampling enough data points of the delay trace.
In Figure 8.5, we can see that the fits really don't strongly depend on γ_2, and that all three fits almost perfectly agree, meaning that the GOAT algorithm found the correct combination of the remaining fitting parameter to recreate the target curve. The differing choices of γ_2 are primarily accounted for by shifting ω_{12} and ω_{23} as displayed in Table 8.1. This shows that the parameter identification algorithm is capable of compensating for a different γ_2 by changing the other parameters. However, this seems not possible for τ_{vib}.

γ_2	τ_{vib} [fs]	Δ_1	Δ_2	ω_{12}	ω_{13}	$\Delta\mathcal{F}$
0.008	500	0.00000	0.00479	0.01000	0.05600	0
0.006	500	-0.00003	0.00001	0.01332	0.05358	2.77×10^{-6}
0.008	500	0.00034	-0.00515	0.00985	0.05604	1.19×10^{-9}
0.010	485	-0.00126	0.02210	0.00750	0.05805	1.25×10^{-7}

Table 8.1: The first row displays the benchmark parameters, the last three rows show the optimal fits identified by the parameter identification algorithm, when γ_2 is fixed at 0.006 fs^{-1}, 0.008 fs^{-1} or 0.010 fs^{-1} respectively. All quantities have units of fs^{-1} except for τ_{vib}. It can be seen that the fit with $\gamma_2 = 0.008$ fs^{-1} has fitting parameters which correspond with the benchmark, however, that is not the case for the other two fits.

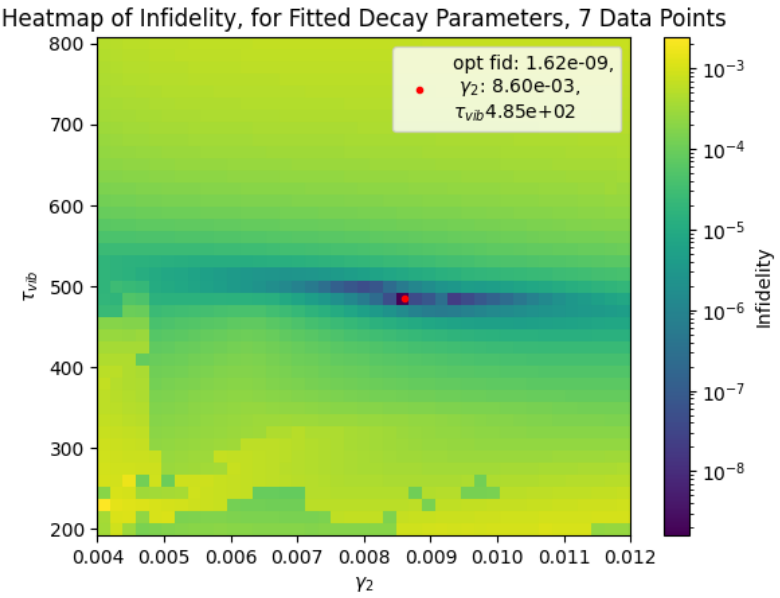

Figure 8.2: Heat map of Infidelity, the optimal infidelity is 1.62×10^{-9}, with decay parameters which are close to the benchmark parameters themselves.

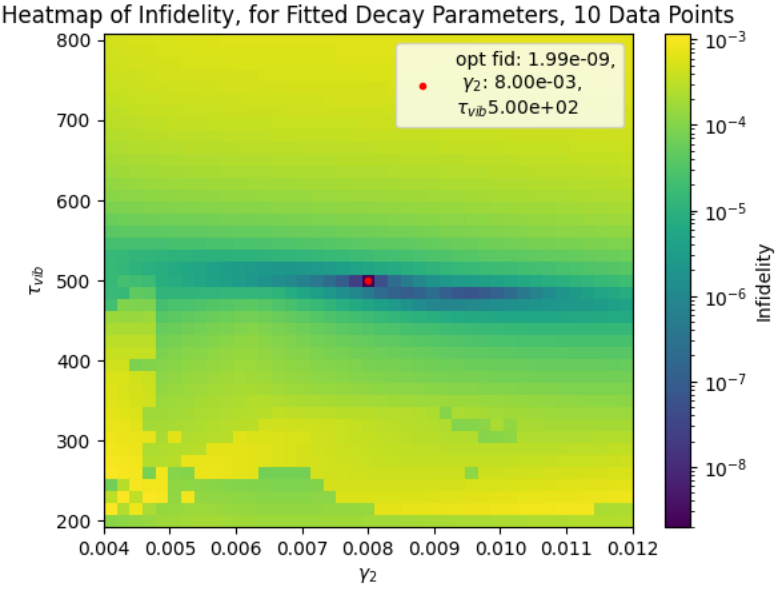

Figure 8.3: Heat map of infidelity, the optimal infidelity is 1.99×10^{-9}, with the correct decay parameters identified. Here, the optimal infidelity is larger than in the case with 7 data points; however, this is because we have summed more infidelities for each data point together.

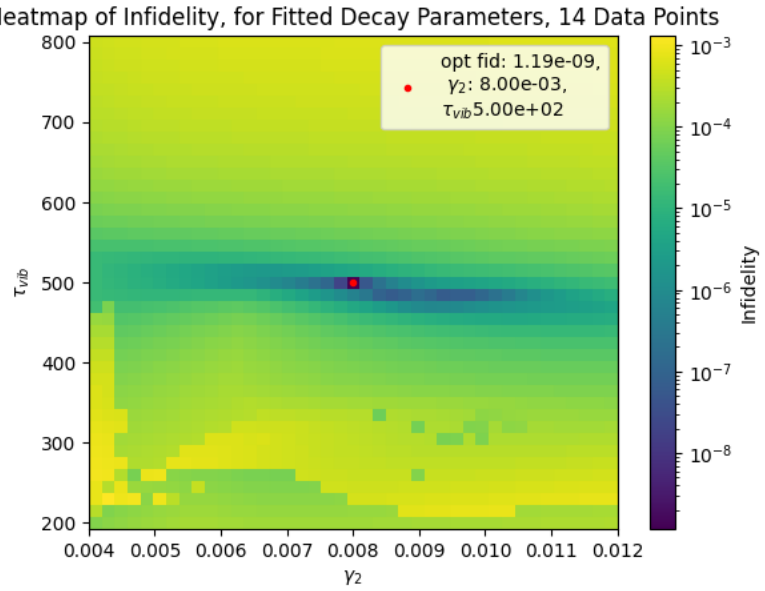

Figure 8.4: Heat map of infidelity, the optimal infidelity is $1.19 \cdot 10^{-9}$, with the correct decay parameters identified.

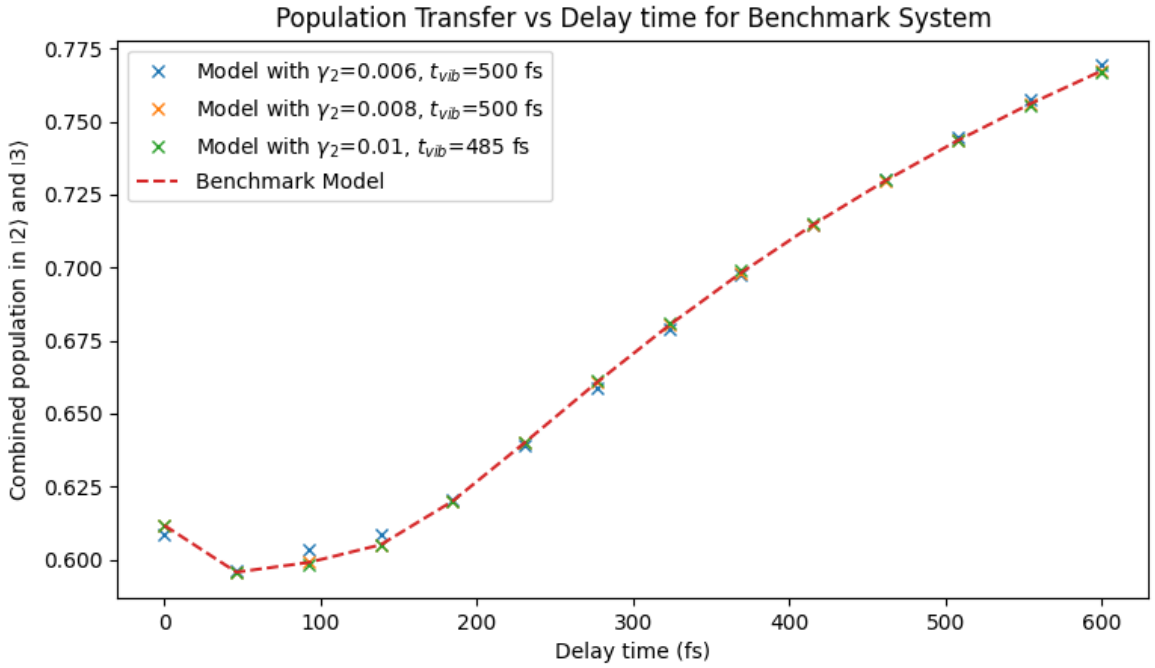

Figure 8.5: Delay trace of the benchmark model with differing choices of γ_2 and τ_{vib}. The other fitting parameters are displayed in Table 8.1 and were obtained using the parameter identification related to Figure 8.4.

8.3 Benchmark with Added Noise

We will now repeat the same parameter identification experiments, but will now include some random uncorrelated noise to each data point of the delay trace, as shown in Figure 8.6. The experimental noise was modelled by adding to each data point a random number drawn from a Gaussian distribution with a standard deviation of 0.003, 0.01 or 0.03, respectively.

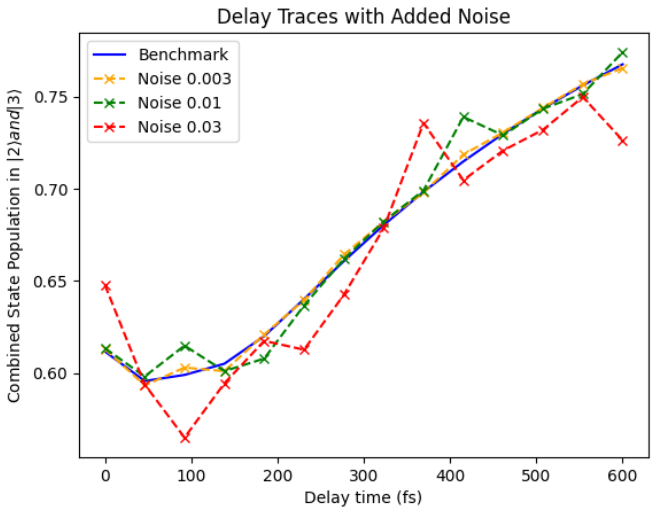

Figure 8.6: Here the original delay trace of Section 8.2 is displayed together with three traces, where Gaussian noise is added with standard deviations of 0.003, 0.01 and 0.03 respectively.

In Figure 8.7, the benchmark decay parameters are still almost correctly identified; however the optimal infidelity has significantly increased compared to Figure 8.4, where no experimental noise was present.

In Figure 8.8, interestingly enough, the model parameters are identified with good accuracy; however, the combination of a higher optimal infidelity and a larger region of decay parameters which produce close-to-optimal infidelity, compared to the previous figure, suggests that this may have happened by chance.

In Figure 8.9, the correct decay parameters are no longer recovered, the optimal infidelity is now associated with a set of decay parameters in the bottom-left of the heat map, creating a region of low infidelity which is disconnected from the benchmark parameters, which suggests that with this level of experimental noise, parameter identification, even with large error bars, is no longer possible.

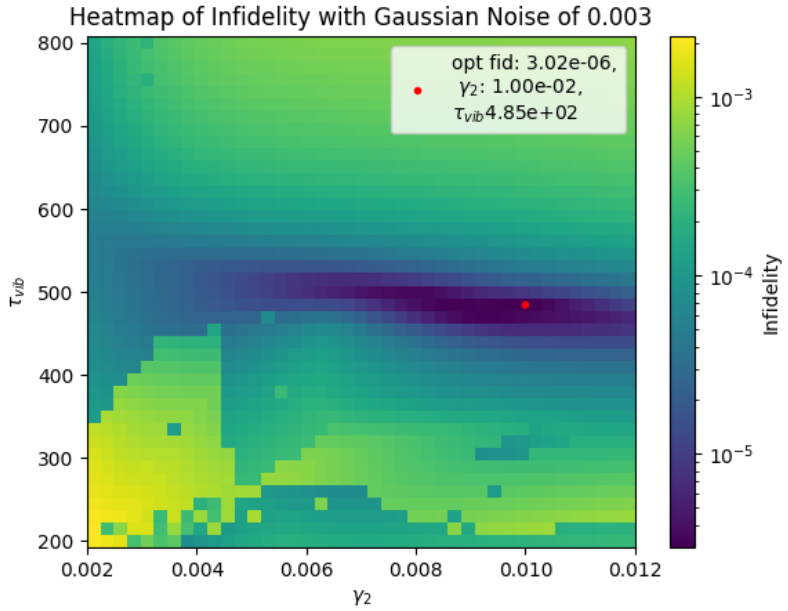

Figure 8.7: Heat map of infidelity, obtained using the parameter identification algorithm for a delay trace with Gaussian noise with a standard deviation of 0.003. The optimal infidelity is 3.0×10^{-6}, which is three orders of magnitude larger than in Figure 8.4.

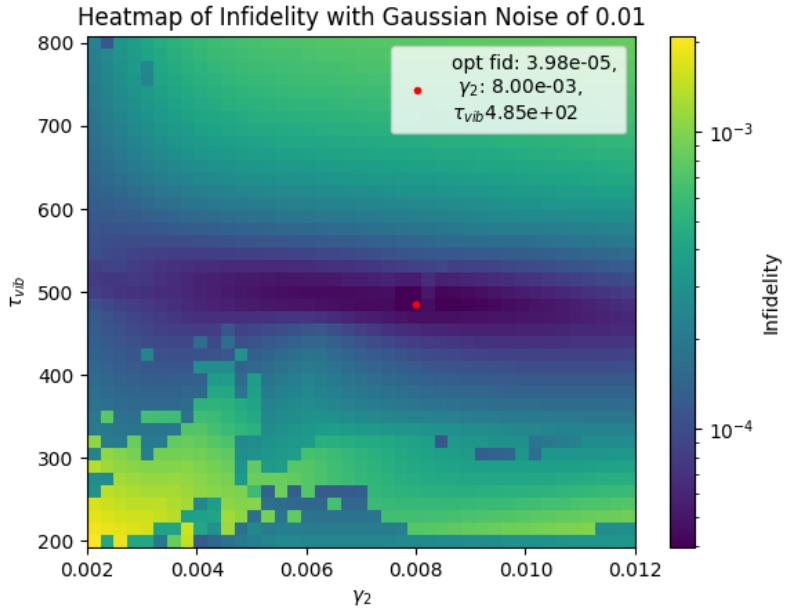

Figure 8.8: Heat map of infidelity, obtained using the parameter identification algorithm for a delay trace with Gaussian noise with a standard deviation of 0.01. The optimal infidelity is 4.0×10^{-5}, which is one order of magnitude larger than in Figure 8.7.

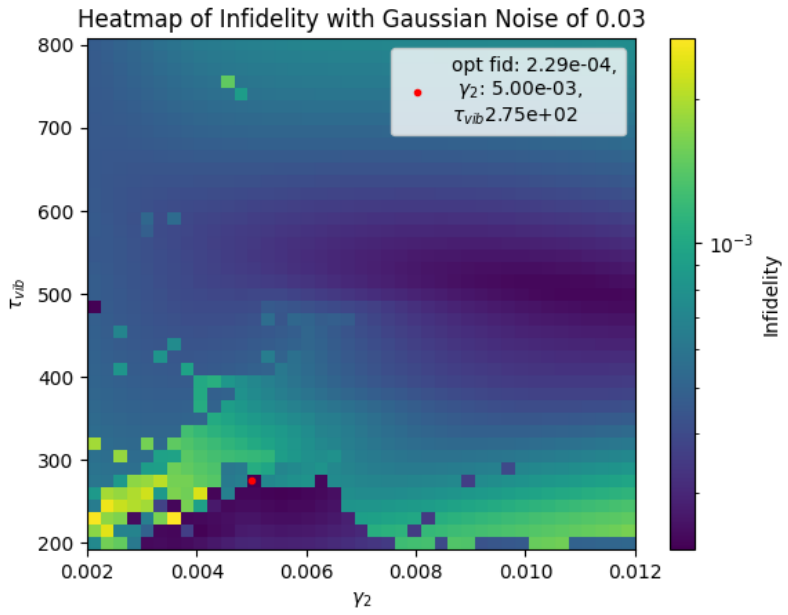

Figure 8.9: Heat map of infidelity, obtained using the parameter identification algorithm for a delay trace with Gaussian noise with a standard deviation of 0.01. The optimal infidelity is 2.3×10^{-4}, which significantly larger than in Figure 8.8.

8.4 Benchmark with Scale Factor

One complication when comparing the experimental data and the Python simulations is the issue that the experimental data provide only a proxy for the excited state population: namely, the measurement photon counts per second coming from excited state decay. We can assume that the emission is linearly correlated with the excited state population; however, the conversion factor between the two is difficult to determine at times.

There are two ways for dealing with this: either a scale factor is selected manually based on best estimates, or the scale factor $S \in \mathbb{R}^+$ becomes part of the optimisation algorithm In our approach, we handle this by first translating the emissions counts using a fixed ratio to excited state populations $P_{|2\rangle,|3\rangle}(\Delta t)$ using a reasonable estimate and then we introduce a scale factor $S \in \mathbb{R}^+$, which is generally restricted to $S \in [\frac{1}{2}, 2]$, which allows us the scale the initial excited state populations as follows:

$$P'_{|2\rangle}(\Delta t) = SP_{|2\rangle}(\Delta t), \tag{8.11}$$
$$P'_{|3\rangle}(\Delta t) = SP_{|3\rangle}(\Delta t). \tag{8.12}$$

This means that the ground state population changes like this:

$$\begin{aligned} P'_{|1\rangle}(\Delta t) &= 1 - P'_{|2\rangle}(\Delta t) - P'_{|3\rangle}(\Delta t), && (8.13)\\ &= 1 - S(P_{|2\rangle}(\Delta t) + P_{|3\rangle}(\Delta t)), && (8.14)\\ &= 1 - S(1 - P_{|1\rangle}(\Delta t)). && (8.15) \end{aligned}$$

This is incorporated into the cost function as follows:

$$\mathcal{C} = \sum_{\Delta t \in A} \left(1 - S(1 - P_{|1\rangle}(\Delta t)) - |\, \langle 1|\psi(T;\Delta t)\rangle \,|\right)^2, \tag{8.16}$$

with the GOAT algorithm being provided the following derivative:

$$\frac{\partial \mathcal{C}}{\partial S} = 2 \sum_{\Delta t \in A} (P_{|1\rangle}(\Delta t) - 1) \left[1 - S(1 - P_{|1\rangle}(\Delta t)) - |\, \langle 1|\psi(T;\Delta t)\rangle \,|\right]. \tag{8.17}$$

In Figure 8.10, the results of the parameter identification algorithm, with the scale factor S as an optimisation parameter, are displayed. On the left side of the heat map, a checkerboard pattern seems to emerge; such drastic changes of infidelity for neighbouring decay parameter pairs suggest that the GOAT algorithm has issues reaching a global infidelity minima due to the introduction of the scale factor parameter.

The parameter identification algorithm takes significantly more resources when the scale factor is introduced, even if the termination conditions of the underlying GOAT algorithm are left unchanged. That suggests that the termination conditions don't happen to be slightly too restrictive

after including the scaling factor; instead, the convergence has become more inefficient because of the scaling factor.

This suggests that the introduction of the scaling factor as an optimisation variable is likely not an efficient solution; a better solution would likely be to run the parameter identification algorithm for a few different fixed scaling factors.

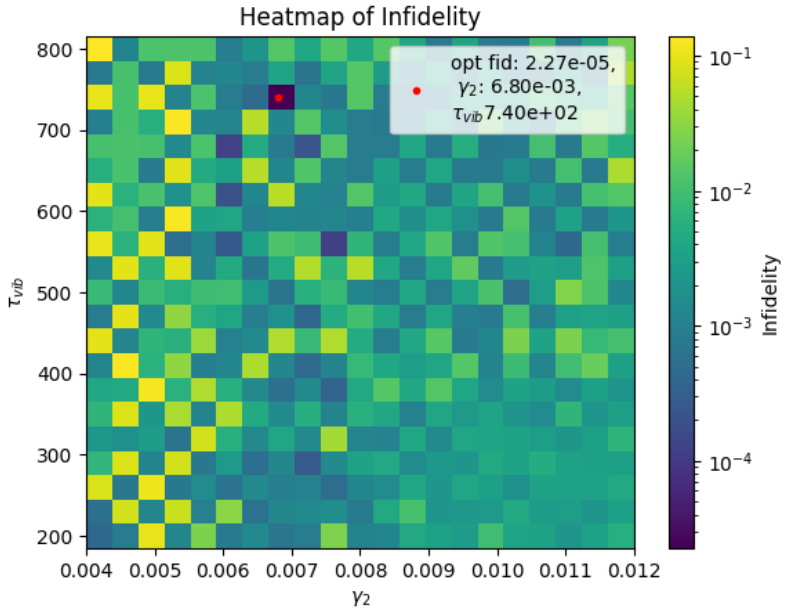

Figure 8.10: Heat map of infidelity, obtained using the parameter identification algorithm on the benchmark delay trace, but including the scaling factor as an optimisation variable. The optimal infidelity, which is achieved, is four orders of magnitude compared to the benchmark from Figure 8.4. The decay parameters identified as optimal also do not correspond to the benchmark decay parameters.

8.5 Fit of Experimental Data using Two-Level System

We will now start with applying the parameter identification algorithm to real-world experimental data. We will start with a delay trace that looks rather simple; it resembles an exponential decay. We thus restrict ourselves to a two-level system in the parameter identification algorithm, we will use the mathematical model provided in Section 7.2.

The only form in which this experimental data is available to us is the graphs from the paper itself. That means that we have to manually extract the data points from these graphs. One complication is that the experimental data points are plotted as continuous, but jagged black lines. All the peaks in the graph are, of course, data points, but if a point is not a local minima or maxima, it might be hard to distinguish it from the line segment connecting its neighbours. We took every part of the graph where the derivative of the line changes to be a data point, which resulted in the data points from Figure 8.11. We used the website `automeris.io` to aid with the data point retrieval from the graphs. It determines the location of each data point by measuring its pixel distance away from the axis. We manually selected the pixels which would count as our data points.

Our procedure for selecting which data points of the experimental delay trace to include in the fitting process was kept simple. In this case, every fourth data point was kept, and the remainder were discarded. This was a rather rudimentary method, and it did happen to create some bias in the results, unfortunately. We can not include every data point of the delay trace due to computational resource constraints of the parameter identification algorithm.

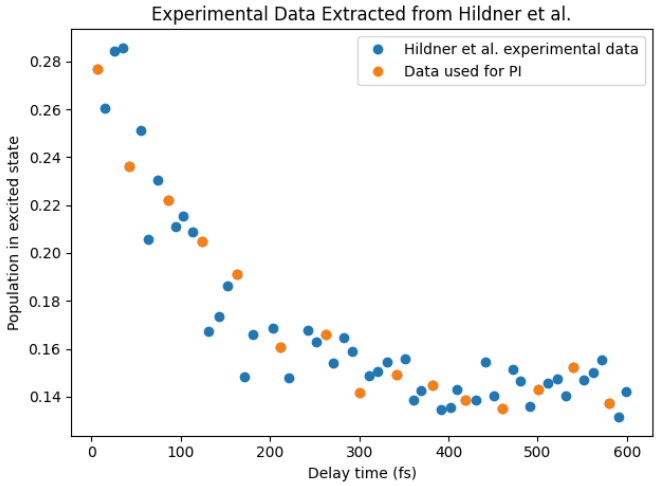

Figure 8.11: In blue and orange, all the data points extracted from the delay trace in Figure 2a of [43]. The orange points are specifically the data points which are used in the parameter identification algorithm.

In Figure 8.12, the optimised infidelity resulting from the parameter identification algorithm is shown. The optimised infidelity is minimal for $\gamma_2 = 0.00345$ fs^{-1}. It is clear that including the scaling factor turns the optimisation less reliable.

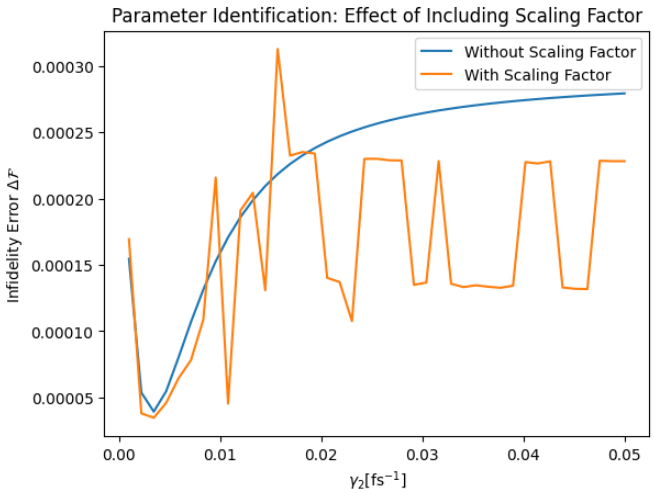

Figure 8.12: The optimised infidelity resulting from the parameter identification algorithm using a given electronic dephasing lifetime γ_2 is depicted in this figure. For the blue curve, the scale factor is not included as an optimisation variable, and we see a smooth change in optimised infidelity. For the orange curve, the scale factor is included, and the curve is rather spiky, suggesting that the GOAT algorithm again had issues consistently finding the global optima of infidelity.

In Figure 8.13, multiple sample delay traces are depicted with control parameters given in Table 8.2. The green and red delay traces follow the data points used in the parameter identification algorithm very well. However, due to bias in the sampling of the data points, the parameter identification algorithm suggests that smaller values of γ_2 produce better results, but it is clear that these delay traces fail to capture the trend of the entire data set.

In the original paper, a Rabi frequency of $\omega_{R,max} = 0.01$ fs^{-1} and a γ_2 between 0.02 fs^{-1} and 0.04 fs^{-1} is suggested as a suitable fit. The purple and brown delay traces in Figure 8.13 have similar parameters and seem to partly capture the trend of the entire data set.

γ_2	Δ	$\omega_{R,max}$	$\Delta\mathcal{F}$	S
0.0035	$1.2 \cdot 10^{-5}$	0.0082	$3.9 \cdot 10^{-5}$	1.0
0.0035	-0.0027	0.0077	$3.5 \cdot 10^{-5}$	0.89
0.011	$1.1 \cdot 10^{-5}$	0.011	$1.7 \cdot 10^{-4}$	1.0
0.033	$-2.5 \cdot 10^{-5}$	0.017	$2.7 \cdot 10^{-4}$	1.0

Table 8.2: Each row shows the optimal fits identified by the parameter identification algorithm, when γ_2 is fixed. All quantities have units of fs^{-1} except for the scale factor S.

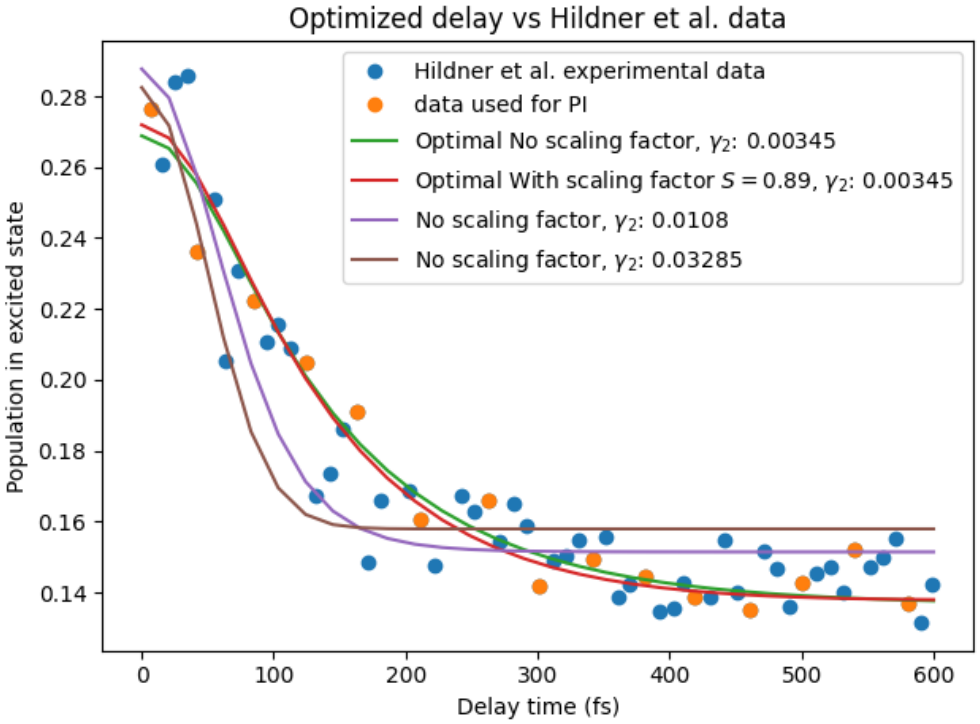

Figure 8.13: Comparison between experimental data and simulated delay traces using different control parameters, which are given in Table 8.2.

8.6 Fit of Experimental Data using Three-Level System

We will now continue our work on parameter identification by looking at a delay trace with more complicated behaviour: Figure 8.14. Therefore, we will attempt to model it using the three-level system, which we discussed at the start of this chapter.

In the figure, the fit does not capture the initial stagnant/downward trend. In this plot, you see some unexpected behaviour, where the data is first declining or flat and only around ~ 200 fs starts to increase again. You can see that the blue fit does not correspond with this behaviour. The goal for this section is to find control parameters which more clearly follow the shape of this delay trace.

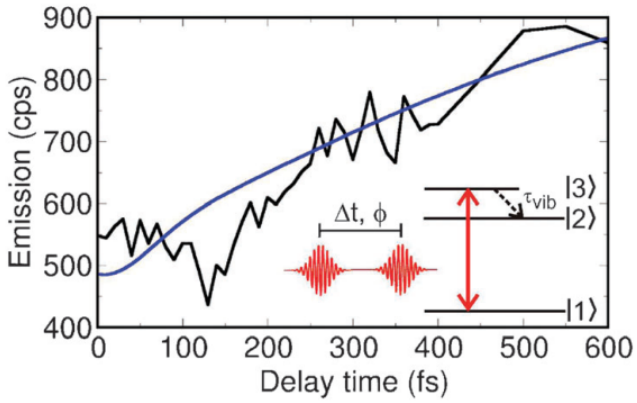

Figure 8.14: In black is the experimental data. In blue, a fit was performed using a two-level system model. Taken from Figure 3 in [42].

We have to be careful when selecting which data points to include in the cost function, which is optimised in the parameter identification algorithm. There is significant experimental noise, the data points are unevenly spaced, and we can't include every data point due to computational cost. Contrary to the previous section, we will manually select the data points which will represent the entire delay trace in the cost function, since this is the least complicated method to deal with all of the aforementioned issues. We take the conversion factor between emission in counts per second and excited state population to be 1100.

In Figure 8.16, the result of the parameter identification are displayed. The optimal pair of decay parameters is in the top right corner. This would suggest that it might be wise to expand the grid search which we perform; however, as shown in Figure 8.17, these combinations of decay and control parameters result in very similar plots. Additionally, increasing the electronic dephasing and vibrational decay lifetimes beyond those values will create minimal changes because their lifetimes are already far larger than the dynamics timescales which we are interested in, namely replicating that dip in excited state population in the first 200 fs.

In Table 8.3, the optimised control parameters for two decay parameter pairs are given. Although the decay parameters differ significantly, the differences in the control parameters are minimal,

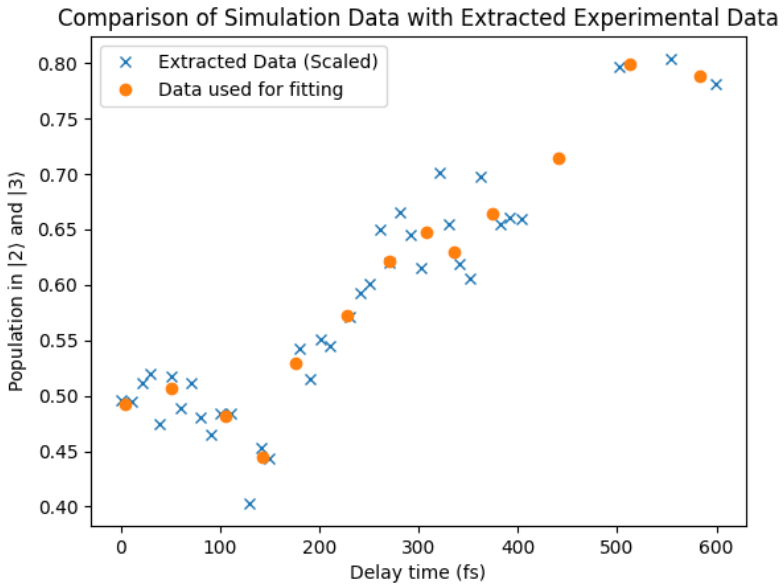

Figure 8.15: Here, the data extracted from the graph in Figure 8.14 is displayed, and the data points which were selected by hand to represent the experimental data for the parameter identification algorithm.

with the same thing holding true for the optimised infidelity. This suggests that the parameter identification algorithm produces a large plateau of similar-looking fits, as illustrated in Figure 8.17. The two delay traces are very similar and seem to largely fit within the bounds of experimental error.

γ_2	τ_{vib} [fs]	Δ_1	Δ_2	ω_{12}	ω_{13}	$\Delta\mathcal{F}$
0.008	500	0	0	0	0.05	N/A
$2.95 \cdot 10^{-4}$	800	-0.012	-0.014	0.011	0.029	3.1×10^{-4}
0.001	530	-0.011,	-0.014	0.0090	0.030	3.2×10^{-4}

Table 8.3: The first row displays the parameters used in the source of the experimental data itself. The second row is the optimal set of parameters identified in Figure 8.16. For the third row, a pair of decay parameters in the low-infidelity region, but relatively far away from the optimum, was selected; the control parameters were optimised using the parameter identification algorithm. All quantities have units of fs^{-1} except for τ_{vib}.

In Figure 8.17, the parameter identification algorithm is applied to the experimental data provided in Figure 8.15. The last fit has $T_2 = \frac{1}{\gamma_2} \approx 3400$ fs, which suggests that in this 3-level model, it might not be required for the model to have any electronic dephasing. However, we have to keep in mind that we assumed whilst constructing the model that the energy difference between $|2\rangle$ and $|3\rangle$ would be small. Table 8.3 shows that this is not the case. However, the Rabi frequency ω_{12} is about three times as small as ω_{13}, meaning that there is some sense of hierarchy.

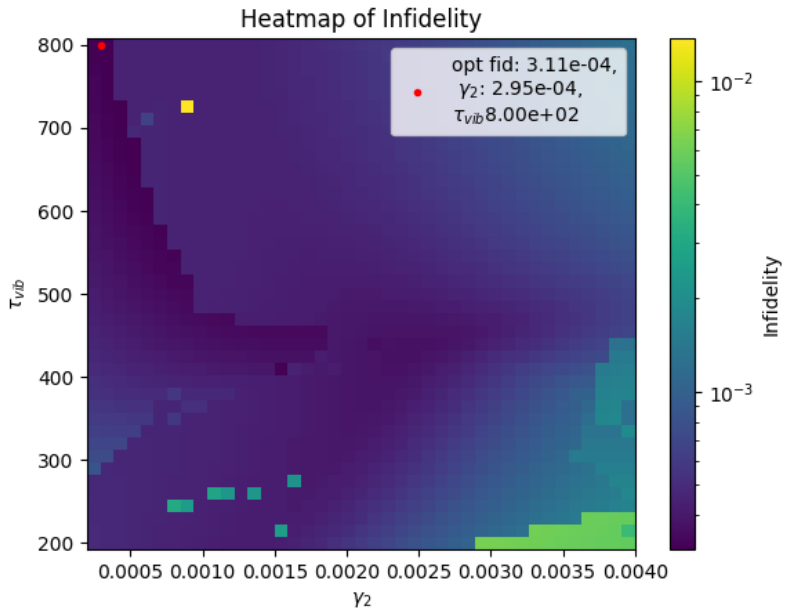

Figure 8.16: Here the heatmap resulting from the parameter identification algorithm are displayed. It can be seen that for large regions in the decay parameter space, a similar optimised infidelity is achieved.

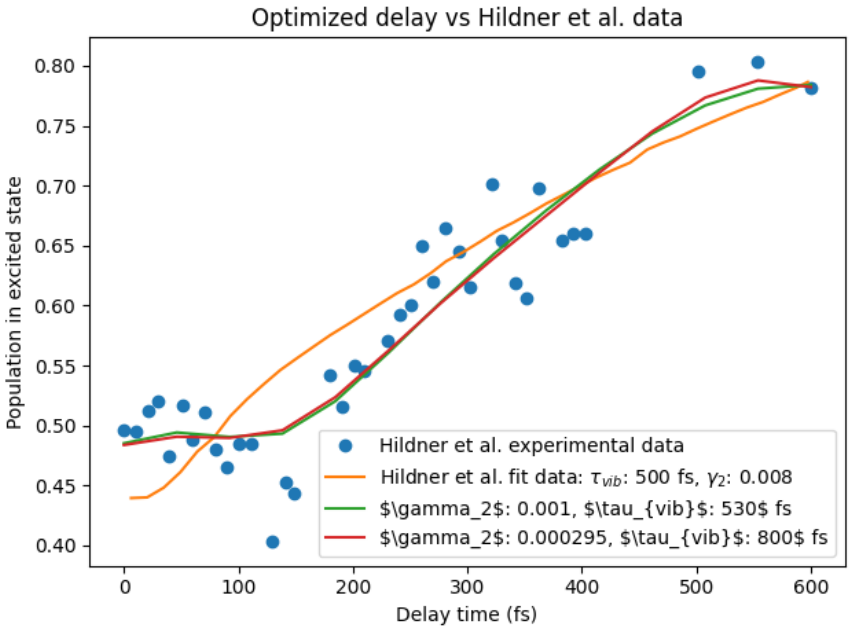

Figure 8.17: The delay traces, with the experimental data and various fits. In orange is the original fit, and in green and red are two different fits which were performed on the data. Although their decay parameters vary significantly, they produce the same delay trace, suggesting this curve is optimal in explaining the data due to the experimental noise. This suggests that we can not determine the decay parameters with any satisfactory precision.

8.7 Conclusion

In this chapter, we have examined three benchmarks and two experimental data sets, each bringing its own insights. In Section 8.2, we established that the number of data points required is such that the parameter identification algorithm performs adequately. In the following Section 8.3, we looked at the effects of experimental noise, determining how much random noise can be added to the delay trace before the parameter identification algorithm fails to function. Both of these benchmarks highlighted the benefits of maximising the quantum Fisher information of the control pulse. For example, in Section 8.6, we saw that very different decay parameters can still produce the same control pulse, meaning the parameter identification algorithm can not suggest a single-best set of decay parameters.

In Section 8.4, we introduced the scaling factor; however, it highlighted the issues which can be caused by the need to use non-local optimisation algorithms: sometimes, no longer having a guarantee that the algorithm will converge consistently to global optima.

In Section 8.5, we applied the parameter identification algorithm to experimental data, but it was largely successful. However, issues were caused by bias in the selection of fourteen data points, which would be used to fit the simulated delay traces. Manually selecting those data points might alleviate some issues, but it is also not a rigorous solution to the problem. Applying a more mathematically rigorous method would likely be beneficial.

In Section 8.6, the parameter identification algorithm was applied to a three-level system, and the algorithm seemed to consistently find the global optimum for each decay parameter pair, because unlike heat maps like Figure 8.10, the optimised infidelity seems to smoothly vary across the decay parameter space for the most part.

Future avenues of research could involve developing methods to design experiments which are optimised for quantum Fisher information by creating a GOAT-based optimal identification algorithm. It could also be interesting to more explicitly connect the heat maps discussed in this chapter with the quantum Fisher information metric. Additionally, it could also be interesting to perform a quantitative comparison of performance between GOAT-based parameter identification and optimal identification algorithm.

Moreover, the parameter identification algorithm allows for a more quantitative comparison of the ability of quantum models to explain experimental data. Models with too many degrees of freedom are prone to overfitting, like in Figure 8.16. The most appropriate model should have a well-defined infidelity minimum in the decay parameter space with a sharp increase as you move away from the optimum.

Appendix A

BFGS Optimisation Algorithm

The BFGS optimisation algorithm is a quasi-Newton method which uses its own approximations of the cost function Hessian to accelerate the gradient descent. It is a gradient-based method, meaning that you have to supply it with the gradient of the cost function yourself. GRAPE's core idea is exactly the computation of the gradient of the infidelity. Which means it often pairs well with the BFGS algorithm. For a more thorough introduction, consider the textbook [47]. The algorithm was named after Broyden, Fletcher, Goldfarb and Shanno, who first developed the algorithm in 1970 [48, 49, 50, 51].

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function which we want to minimise without applying any constraints. Before using the algorithm, we pick an initial guess $x_0 \in \mathbb{R}^n$. We will then continue iteratively to get a better approximation. The Taylor series of f around x_k is:

$$f(x_k + \Delta x) = f(x_k) + \nabla f(x_k)^T \Delta x + \frac{1}{2}\Delta x^T B_k \Delta x + \mathcal{O}\left(|\Delta x|^3\right), \tag{A.1}$$

where B_k is the Hessian of f at x_k. We wish to the find Δx such that $\nabla f(x_k + \Delta x) = 0$, differentiating the equation above, we get:

$$0 = \nabla f(x_k) + B_k \Delta x. \tag{A.2}$$

Which suggests that we should pick $\Delta x_k = -B_k^{-1} \nabla f(x_k)$ as our descent direction, and then select the appropriate step size α_k, which minimises $f(x_k + \alpha_k \Delta x_k)$. The updated B_{k+1} satisfies:

$$B_{k+1}(x_{k+1} - x_k) = \nabla f(x_{k+1}) - \nabla f(x_k), \tag{A.3}$$

which follows from the Taylor expansion for ∇f centred at x_{k+1}:

$$\nabla f(x_{k+1}) = \nabla f(x_k) + B_{k+1}(x_{k+1} - x_k) + \mathcal{O}\left(|x_{k+1} - x_k|^2\right). \tag{A.4}$$

We require that this approximate Hessian is positive definite. If f is not strongly convex, this is an additional condition which has to be applied during the line search. Set $s_k := x_{k+1} - x_k$ and $y_k := \nabla f(x_{k+1}) - \nabla f(x_k)$. However, there is no unique solution to the above equation. Therefore, we want to select the solution which minimises the matrix norm $||B_{k+1} - B_k||$, whilst B_{k+1} still satisfies equation (A.3) and is symmetric. The following construction satisfies those conditions.

$$B_{k+1} = B_k + \alpha u u^T + \beta v v^T, \tag{A.5}$$

Choosing $u = y_k$ and $v = B_k s$ and imposing equation (A.3), we get

$$\alpha = \frac{1}{y_k^T s_k}, \quad \beta = -\frac{1}{s_k^T B_k s_k}. \tag{A.6}$$

Plugging these parameters back in, we get update equation for B_{k+1}:

$$B_{k+1} = B_k + \frac{y_k y_k^T}{y_k^T s_k} - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k}. \tag{A.7}$$

A combination of termination conditions can be selected, based on: the number of iterations, the estimated error $|f(x_k) - f(x_*)|$ or the norm of the gradient $|\nabla f(x_k)|$.

Appendix B

Ancillary Figures: Small Final Times

Here are the figures related to the discussion in Section 5.7. The optimisations are performed using the GRAPE algorithm and the four-level Hamiltonian given in the aforementioned section.

The method of repeated GRAPE optimisation, as in Figure 5.13, is employed to ensure both the infidelity and the control amplitude cost are optimised.

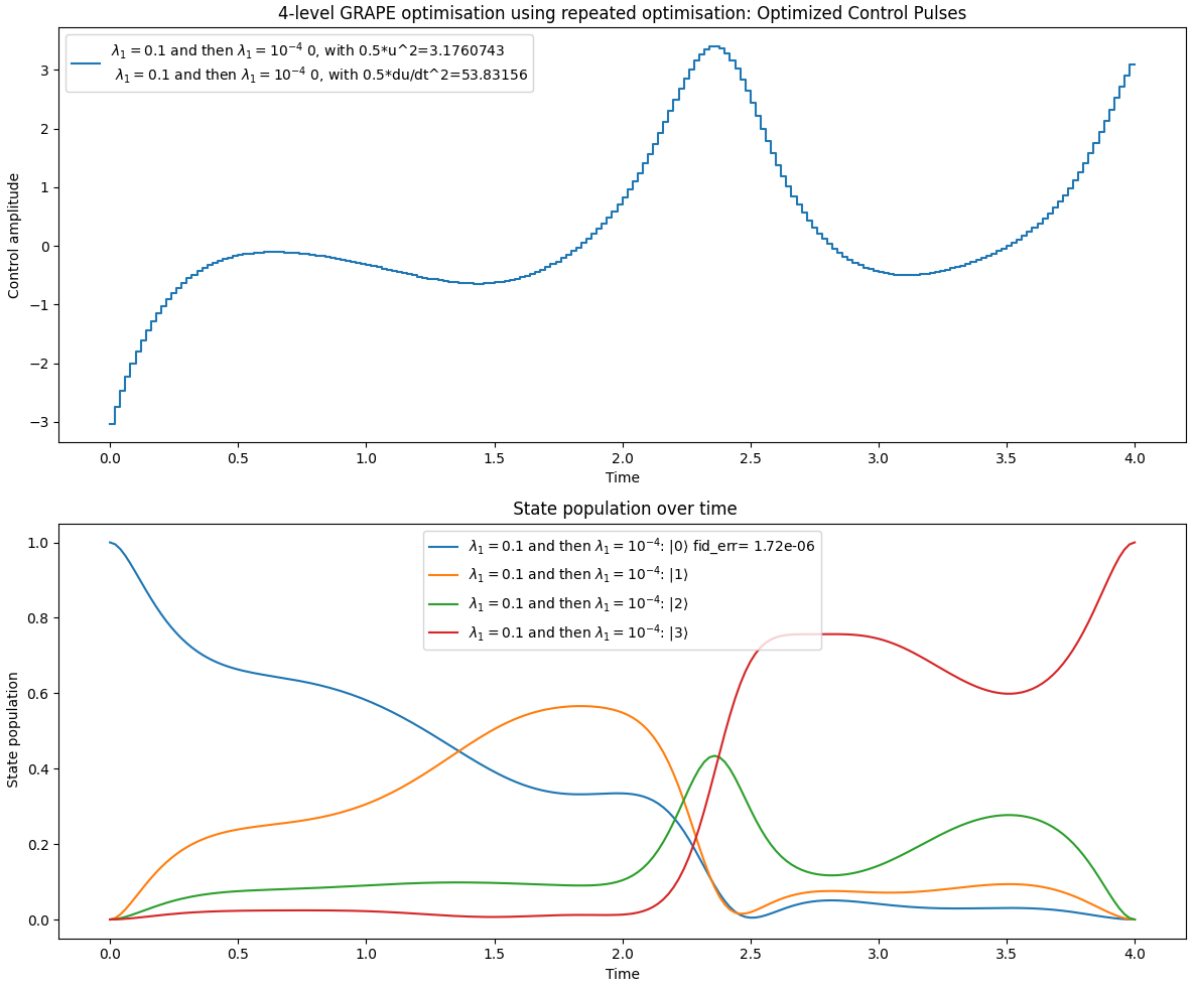

Figure B.1: Here the final time T is 4. We see that the maximal control amplitude is ~ 3. The final infidelity error is $\sim 1.7 \cdot 10^{-6}$, meaning that near-perfect state transfer takes place.

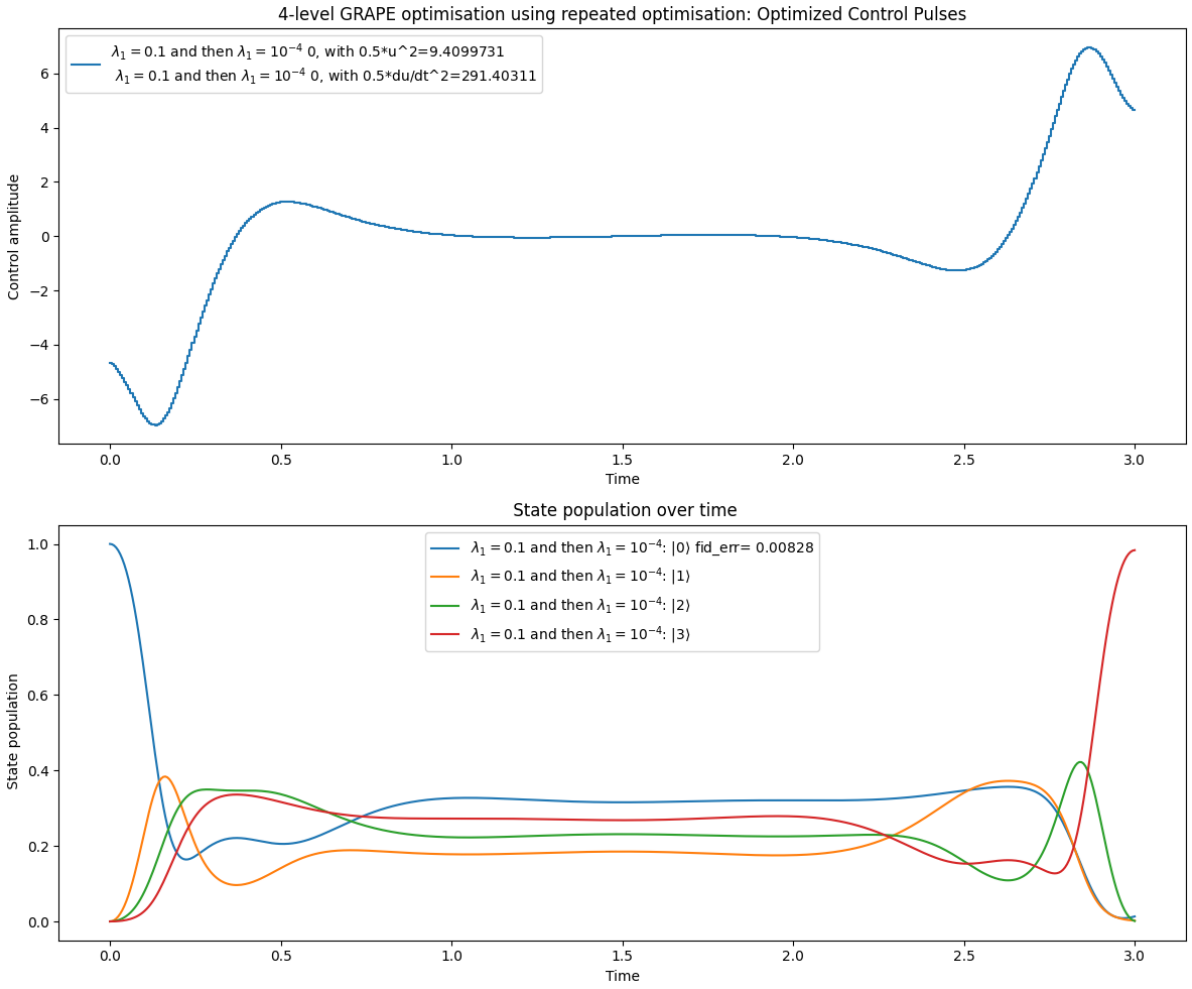

Figure B.2: Here the final time T is 3. We see that the maximal control amplitude is ~ 6, which is double compared to the previous figure. The final infidelity error is $\sim 8.3 \cdot 10^{-3}$, which is also worse than the previous figure.

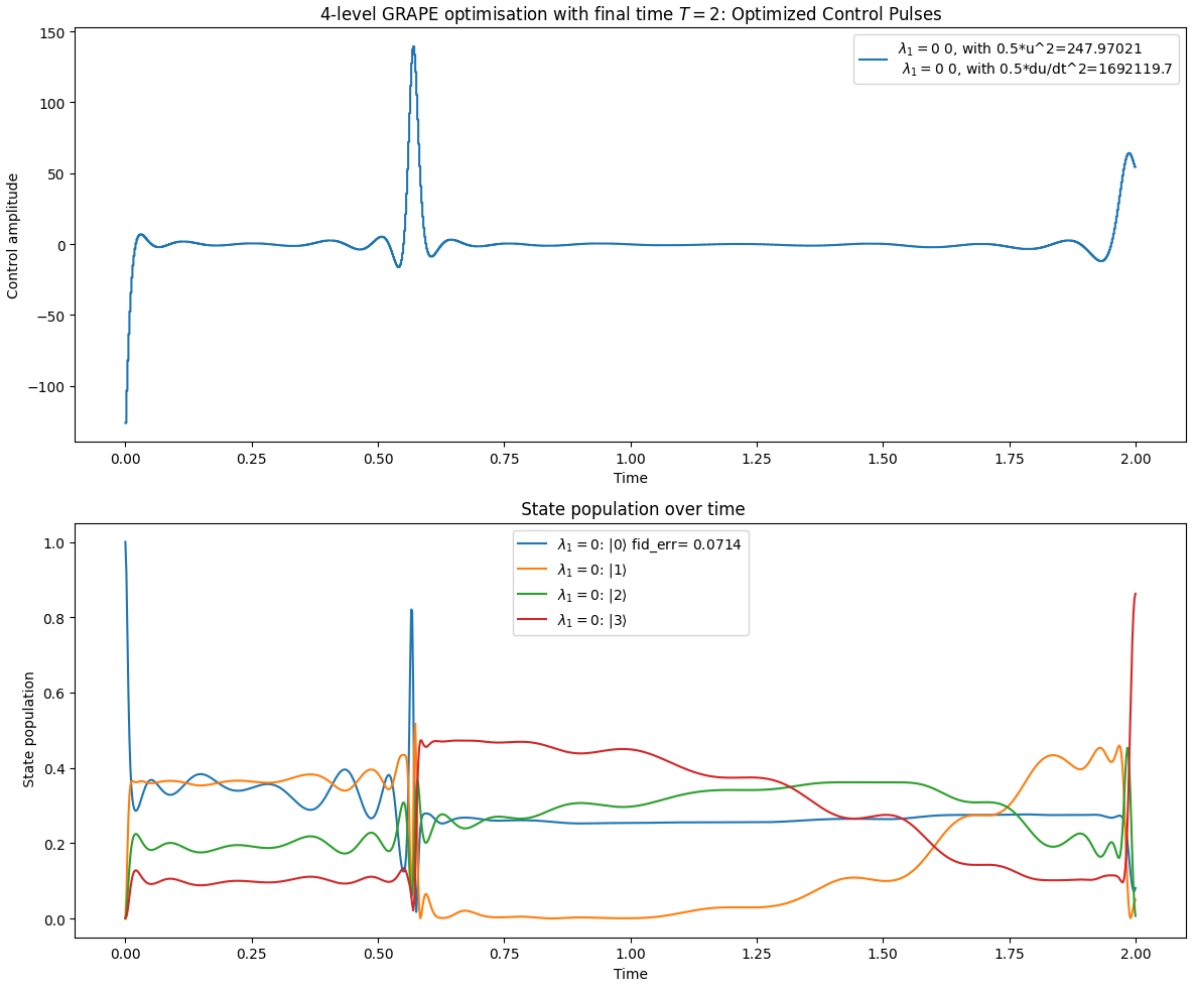

Figure B.3: Here the final time T is 2. We see that the maximal control amplitude has become ~ 150, which is a massive increase compared to the previous figures. The final infidelity error is ~ 0.07, meaning that the state transfer has become noticeably imperfect. For smaller final times, the control pulses will become more erratic and larger, which makes it very hard for the GRAPE algorithm to find a zero-infidelity control pulse.

Bibliography

[1] Stefffen J. Glaser, Ugo Boscain, Tommaso Calarco, Christiane P. Koch, Walter Köckenberger, Ronnie Kosloff, Ilya Kuprov, Burkard Luy, Sophie Schirmer, Thomas Schulte-Herbrüggen, D. Sugny, and Frank K. Wilhelm. "Training Schrödinger's cat: quantum optimal control". In: *The European Physical Journal D* 69.12 (Dec. 2015). arXiv:1508.00442 [quant-ph], p. 279. ISSN: 1434-6060, 1434-6079. DOI: 10.1140/epjd/e2015-60464-1. URL: http://arxiv.org/abs/1508.00442 (visited on 10/26/2025).

[2] David J. Tannor and Stuart A. Rice. "Control of selectivity of chemical reaction via control of wave packet evolution". In: *The Journal of Chemical Physics* 83.10 (Nov. 1985), pp. 5013–5018. ISSN: 0021-9606. DOI: 10.1063/1.449767. URL: https://doi.org/10.1063/1.449767 (visited on 10/26/2025).

[3] Q. Ficheux, S. Jezouin, Z. Leghtas, and B. Huard. "Dynamics of a qubit while simultaneously monitoring its relaxation and dephasing". en. In: *Nature Communications* 9.1 (May 2018). Publisher: Nature Publishing Group, p. 1926. ISSN: 2041-1723. DOI: 10.1038/s41467-018-04372-9. URL: https://www.nature.com/articles/s41467-018-04372-9 (visited on 10/26/2025).

[4] Chungwei Lin, Yanting Ma, and Dries Sels. "Optimal Control for Quantum Metrology via Pontryagin's principle". In: *Physical Review A* 103.5 (May 2021). arXiv:2105.06931 [quant-ph], p. 052607. ISSN: 2469-9926, 2469-9934. DOI: 10.1103/PhysRevA.103.052607. URL: http://arxiv.org/abs/2105.06931 (visited on 10/26/2025).

[5] J. J. Sakurai and Jim Napolitano. *Modern Quantum Mechanics*. en. ISBN: 9781108587280 Publisher: Cambridge University Press. Sept. 2020. DOI: 10.1017/9781108587280. URL: https://www.cambridge.org/highereducation/books/modern-quantum-mechanics/DF43277E8AEDF83CC12EA62887C277DC (visited on 01/30/2025).

[6] Leslie E. Ballentine. *Quantum Mechanics: A Modern Development*. en. Google-Books-ID: sHJRFHz1rYsC. World Scientific, 1998. ISBN: 978-981-02-4105-6.

[7] Arthur S. Wightman, ed. *Part I: Particles and Fields. Part II: Foundations of Quantum Mechanics*. en. Berlin, Heidelberg: Springer, 1997. ISBN: 978-3-642-08179-8 978-3-662-09203-3. DOI: 10.1007/978-3-662-09203-3. URL: https://link.springer.com/10.1007/978-3-662-09203-3 (visited on 01/30/2025).

[8] Stephan Fackler. *Mathematical Foundations of Quantum Mechanics*. en. Lecture notes, University of Ulm. 2015.

[9] Carlo Alabiso and Ittay Weiss. *A Primer on Hilbert Space Theory: Linear Spaces, Topological Spaces, Metric Spaces, Normed Spaces, and Topological Groups*. en. UNITEXT for Physics. Cham: Springer International Publishing, 2021. ISBN: 978-3-030-67416-8 978-3-030-67417-5. DOI: 10.1007/978-3-030-67417-5. URL: http://link.springer.com/10.1007/978-3-030-67417-5 (visited on 01/30/2025).

[10] P. Facchi, D. A. Lidar, and S. Pascazio. "Unification of Dynamical Decoupling and the Quantum Zeno Effect". en. In: *Physical Review A* 69.3 (Mar. 2004). arXiv:quant-ph/0303132, p. 032314. ISSN: 1050-2947, 1094-1622. DOI: 10.1103/PhysRevA.69.032314. URL: http://arxiv.org/abs/quant-ph/0303132 (visited on 01/20/2025).

[11] *Parseval equality - Encyclopedia of Mathematics*. URL: https://encyclopediaofmath.org/index.php?title=Parseval_equality (visited on 02/03/2025).

[12] Jussi Lindgren and Jukka Liukkonen. "Quantum Mechanics can be understood through stochastic optimization on spacetimes". en. In: *Scientific Reports* 9.1 (Dec. 2019). Publisher: Nature Publishing Group, p. 19984. ISSN: 2045-2322. DOI: 10.1038/s41598-019-56357-3. URL: https://www.nature.com/articles/s41598-019-56357-3 (visited on 01/30/2025).

[13] David J. Griffiths and Darrell F. Schroeter. *Introduction to quantum mechanics*. English. Third edition. Section: xiii, 495 pages : illustrations ; 26 cm. Cambridge, United Kingdom: Cambridge University Press, 2018. ISBN: 978-1-107-18963-8 1-107-18963-2 978-1-108-79110-6 1-108-79110-7.

[14] N. Bohr. "I. On the constitution of atoms and molecules". In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 26.151 (July 1913). Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/14786441308634955, pp. 1–25. ISSN: 1941-5982. DOI: 10.1080/14786441308634955. URL: https://doi.org/10.1080/14786441308634955 (visited on 10/20/2025).

[15] *Bohr Model Of Hydrogen Atom - Unifyphysics*. en-US. Section: Home. June 2024. URL: https://unifyphysics.com/bohr-model-of-hydrogen-atom/ (visited on 08/19/2025).

[16] *6: Adiabatic Approximation*. en. Sept. 2018. URL: https://chem.libretexts.org/Bookshelves/Physical_and_Theoretical_Chemistry_Textbook_Maps/Time_Dependent_Quantum_Mechanics_and_Spectroscopy_(Tokmakoff)/06%3A_Adiabatic_Approximation (visited on 08/03/2025).

[17] Daniel Manzano. "A short introduction to the Lindblad master equation". en. In: *AIP Advances* 10.2 (Feb. 2020), p. 025106. ISSN: 2158-3226. DOI: 10.1063/1.5115323. URL: https://pubs.aip.org/adv/article/10/2/025106/1021638/A-short-introduction-to-the-Lindblad-master (visited on 01/20/2025).

[18] Heinz-Peter Breuer and Francesco Petruccione. *The Theory of Open Quantum Systems*. Oxford University Press, Jan. 2007. ISBN: 978-0-19-921390-0. DOI: 10.1093/acprof:oso/9780199213900.001.0001. URL: https://doi.org/10.1093/acprof:oso/9780199213900.001.0001 (visited on 08/05/2025).

[19] A. G. Redfield. "The Theory of Relaxation Processes*". In: *Advances in Magnetic and Optical Resonance*. Ed. by John S. Waugh. Vol. 1. Advances in Magnetic Resonance. Academic Press, Jan. 1965, pp. 1–32. DOI: 10.1016/B978-1-4832-3114-3.50007-6. URL: https://www.sciencedirect.com/science/article/pii/B9781483231143500076 (visited on 09/17/2025).

[20] R. de la Madrid. "The role of the rigged Hilbert space in Quantum Mechanics". en. In: *European Journal of Physics* 26.2 (Apr. 2005). arXiv:quant-ph/0502053, pp. 287–312. ISSN: 0143-0807, 1361-6404. DOI: 10.1088/0143-0807/26/2/008. URL: http://arxiv.org/abs/quant-ph/0502053 (visited on 01/20/2025).

[21] *Nlab: Gelfand Triple*. URL: https://ncatlab.org/nlab/show/Gelfand+triple.

[22] Daniel W. Stroock. *Essentials of Integration Theory for Analysis*. en. Vol. 262. Graduate Texts in Mathematics. Cham: Springer International Publishing, 2020. ISBN: 978-3-030-58477-1 978-3-030-58478-8. DOI: 10.1007/978-3-030-58478-8. URL: http://link.springer.com/10.1007/978-3-030-58478-8 (visited on 06/27/2025).

[23] J.J. Duistermaat and J.A.C. Kolk. *Distributions: Theory and Applications*. en. Boston: Birkhäuser, 2010. ISBN: 978-0-8176-4672-1 978-0-8176-4675-2. DOI: 10.1007/978-0-8176-4675-2. URL: https://link.springer.com/10.1007/978-0-8176-4675-2 (visited on 01/30/2025).

[24] Elias M. Stein and Guido Weiss. *Introduction to Fourier Analysis on Euclidean Spaces*. Princeton: Princeton University Press, 1972. ISBN: 978-1-4008-8389-9. DOI: doi:10.1515/9781400883899. URL: https://doi.org/10.1515/9781400883899 (visited on 06/27/2025).

[25] Walter Rudin. *Real and complex analysis*. eng. 3rd ed., International ed. McGraw-Hill series in higher mathematics. OCLC: 1101294288. New York: McGraw-Hill, 1987. ISBN: 978-0-07-100276-9.

[26] R Beukema, W M De Muynck, and J De Graaf. "A GELFAND TRIPLE APPROACH TO WIGNER AND HUSIMI REPRESENTATIONS". en. In: ().

[27] Gjerrit Meinsma and Arjan Van Der Schaft. *A Course on Optimal Control*. en. Springer Undergraduate Texts in Mathematics and Technology. Cham: Springer Nature Switzerland, 2023. ISBN: 978-3-031-36654-3 978-3-031-36655-0. DOI: 10.1007/978-3-031-36655-0. URL: https://link.springer.com/10.1007/978-3-031-36655-0 (visited on 01/20/2025).

[28] Q. Ansel, E. Dionis, F. Arrouas, B. Peaudecerf, S. Guérin, D. Guéry-Odelin, and D. Sugny. "Introduction to Theoretical and Experimental aspects of Quantum Optimal Control". en. In: *Journal of Physics B: Atomic, Molecular and Optical Physics* 57.13 (July 2024). arXiv:2403.00532 [quant-ph], p. 133001. ISSN: 0953-4075, 1361-6455. DOI: 10.1088/1361-6455/ad46a5. URL: http://arxiv.org/abs/2403.00532 (visited on 01/20/2025).

[29] Tak-San Ho and Herschel Rabitz. "Why do effective quantum controls appear easy to find?" en. In: *Journal of Photochemistry and Photobiology A: Chemistry* 180.3 (June 2006), pp. 226–240. ISSN: 10106030. DOI: 10.1016/j.jphotochem.2006.03.038. URL: https://linkinghub.elsevier.com/retrieve/pii/S1010603006001948 (visited on 01/20/2025).

[30] O. V. Morzhin and A. N. Pechen. "Krotov method for optimal control of closed quantum systems". en. In: *Russian Mathematical Surveys* 74.5 (Oct. 2019), pp. 851–908. ISSN: 0036-0279, 1468-4829. DOI: 10.1070/RM9835. URL: https://www.mathnet.ru/eng/rm9835 (visited on 01/20/2025).

[31] H. Rabitz, T.-S. Ho, M. Hsieh, R. Kosut, and M. Demiralp. "Topology of optimally controlled quantum mechanical transition probability landscapes". en. In: *Physical Review A* 74.1 (July 2006), p. 012721. ISSN: 1050-2947, 1094-1622. DOI: 10.1103/PhysRevA.74.012721. URL: https://link.aps.org/doi/10.1103/PhysRevA.74.012721 (visited on 01/20/2025).

[32] Jason Dominy and Herschel Rabitz. "Dynamic homotopy and landscape dynamical set topology in quantum control". en. In: *Journal of Mathematical Physics* 53.8 (Aug. 2012), p. 082201. ISSN: 0022-2488, 1089-7658. DOI: 10.1063/1.4742375. URL: https://pubs.aip.org/jmp/article/53/8/082201/232400/Dynamic-homotopy-and-landscape-dynamical-set (visited on 01/20/2025).

[33] Yidian Fan, Re-Bing Wu, Tak-San Ho, Gaurav V. Bhole, and Herschel Rabitz. *The Top Manifold Connectedness of Quantum Control Landscapes*. en. arXiv:2409.15139 [quant-ph]. Sept. 2024. DOI: 10.48550/arXiv.2409.15139. URL: http://arxiv.org/abs/2409.15139 (visited on 01/20/2025).

[34] Katharine Moore Tibbetts and Herschel Rabitz. "Constrained control landscape for population transfer in a two-level system". en. In: *Physical Chemistry Chemical Physics* 17.5 (2015), pp. 3164–3178. ISSN: 1463-9076, 1463-9084. DOI: 10.1039/C4CP04792C. URL: https://xlink.rsc.org/?DOI=C4CP04792C (visited on 01/20/2025).

[35] Navin Khaneja, Timo Reiss, Cindie Kehlet, Thomas Schulte-Herbrüggen, and Steffen J. Glaser. "Optimal control of coupled spin dynamics: design of NMR pulse sequences by gradient ascent algorithms". en. In: *Journal of Magnetic Resonance* 172.2 (Feb. 2005), pp. 296–305. ISSN: 10907807. DOI: 10.1016/j.jmr.2004.11.004. URL: https://linkinghub.elsevier.com/retrieve/pii/S1090780704003696 (visited on 01/21/2025).

[36] R. M. Wilcox. "Exponential Operators and Parameter Differentiation in Quantum Physics". In: *Journal of Mathematical Physics* 8.4 (Apr. 1967), pp. 962–982. ISSN: 0022-2488. DOI: 10.1063/1.1705306. URL: https://doi.org/10.1063/1.1705306 (visited on 08/08/2025).

[37] Neill Lambert, Eric Giguère, Paul Menczel, Boxi Li, Patrick Hopf, Gerardo Suárez, Marc Gali, Jake Lishman, Rushiraj Gadhvi, Rochisha Agarwal, Asier Galicia, Nathan Shammah, Paul Nation, J. R. Johansson, Shahnawaz Ahmed, Simon Cross, Alexander Pitchford, and Franco Nori. *QuTiP 5: The Quantum Toolbox in Python*. en. arXiv:2412.04705 [quant-ph]. Dec. 2024. DOI: 10.48550/arXiv.2412.04705. URL: http://arxiv.org/abs/2412.04705 (visited on 01/21/2025).

[38] *Qutip-QTRL Github*. URL: https://github.com/qutip/qutip-qtrl.

[39] *Qutip-QOC Github*. URL: https://github.com/qutip/qutip-qoc.

[40] *Quantum Optimal Control — qutip_qtrl 0.2.0.dev documentation*. URL: https://qutip-qtrl.readthedocs.io/en/latest/guide/guide-control.html (visited on 08/09/2025).

[41] Shai Machnes, Elie Assémat, David Tannor, and Frank K. Wilhelm. "Tunable, Flexible, and Efficient Optimization of Control Pulses for Practical Qubits". en. In: *Physical Review Letters* 120.15 (Apr. 2018), p. 150401. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.120.150401. URL: https://link.aps.org/doi/10.1103/PhysRevLett.120.150401 (visited on 01/20/2025).

[42] Richard Hildner, Daan Brinks, Fernando D. Stefani, and Niek F. van Hulst. "Electronic coherences and vibrational wave-packets in single molecules studied with femtosecond phase-controlled spectroscopy". In: *Physical Chemistry Chemical Physics* 13.5 (2011), p. 1888. DOI: 10.1039/c0cp02231d.

[43] Richard Hildner, Daan Brinks, and Niek F. van Hulst. "Femtosecond coherence and quantum control of single molecules at room temperature". In: *Nature Physics* 7.2 (Dec. 2010), pp. 172–177. DOI: 10.1038/nphys1858.

[44] Claude Le Bris, Mazyar Mirrahimi, Herschel Rabitz, and Gabriel Turinici. "Hamiltonian identification for quantum systems: well-posedness and numerical approaches". en. In: *ESAIM: Control, Optimisation and Calculus of Variations* 13.2 (Apr. 2007), pp. 378–395. ISSN: 1292-8119, 1262-3377. DOI: 10.1051/cocv:2007013. URL: http://www.esaim-cocv.org/10.1051/cocv:2007013 (visited on 10/17/2025).

[45] J. M. Geremia and H. Rabitz. "Optimal Hamiltonian identification: The synthesis of quantum optimal control and quantum inversion". English (US). In: *Journal of Chemical Physics* 118.12 (Mar. 2003). Publisher: American Institute of Physics, pp. 5369–5382. ISSN: 0021-9606. DOI: 10.1063/1.1538242. URL: https://collaborate.princeton.edu/en/publications/optimal-hamiltonian-identification-the-synthesis-of-quantum-optim (visited on 10/17/2025).

[46] Shibei Xue, Rebing Wu, Shan Ma, Dewei Li, and Min Jiang. "Gradient algorithm for Hamiltonian identification of open quantum systems". en. In: *Physical Review A* 103.2 (Feb. 2021), p. 022604. ISSN: 2469-9926, 2469-9934. DOI: 10.1103/PhysRevA.103.022604. URL: https://link.aps.org/doi/10.1103/PhysRevA.103.022604 (visited on 10/17/2025).

[47] Jorge Nocedal and Stephen J. Wright. *Numerical optimization*. en. 2nd ed. Springer series in operations research. OCLC: ocm68629100. New York: Springer, 2006. ISBN: 978-0-387-30303-1.

[48] C. G. BROYDEN. "The Convergence of a Class of Double-rank Minimization Algorithms 1. General Considerations". In: *IMA Journal of Applied Mathematics* 6.1 (Mar. 1970), pp. 76–90. ISSN: 0272-4960. DOI: 10.1093/imamat/6.1.76. URL: https://doi.org/10.1093/imamat/6.1.76 (visited on 01/30/2025).

[49] R. Fletcher. "A new approach to variable metric algorithms". In: *The Computer Journal* 13.3 (Jan. 1970), pp. 317–322. ISSN: 0010-4620. DOI: 10.1093/comjnl/13.3.317. URL: https://doi.org/10.1093/comjnl/13.3.317 (visited on 01/30/2025).

[50] Donald Goldfarb. "A family of variable-metric methods derived by variational means". en. In: *Mathematics of Computation* 24.109 (1970), pp. 23–26. ISSN: 0025-5718, 1088-6842. DOI: 10.1090/S0025-5718-1970-0258249-6. URL: https://www.ams.org/mcom/1970-24-109/S0025-5718-1970-0258249-6/ (visited on 01/30/2025).

[51] D. F. Shanno. "Conditioning of quasi-Newton methods for function minimization". en. In: *Mathematics of Computation* 24.111 (1970), pp. 647–656. ISSN: 0025-5718, 1088-6842. DOI: 10.1090/S0025-5718-1970-0274029-X. URL: https://www.ams.org/mcom/1970-24-111/S0025-5718-1970-0274029-X/ (visited on 01/30/2025).