Technische Universität München

Fakultät für Mathematik

Lehrstuhl für Analysis

Sparse Time-Frequency Control of Bilinear Quantum Systems

Felix Wilhelm Henneke

Vollständiger Abdruck der von der Fakultät für Mathematik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:		Univ.-Prof. Dr. Michael Wolf
Prüfer der Dissertation:	1.	Univ.-Prof. Gero Friesecke, Ph. D.
	2.	Univ.-Prof. Dr. Karl Kunisch Karl-Franzens-Universität, Graz, Österreich
	3.	Prof. Eric Cancès, Ph. D. École de Ponts ParisTech, Paris, Frankreich (schriftliche Beurteilung)

Die Dissertation wurde am 26.2.2015 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 17.9.2015 angenommen.

Abstract

In this thesis, optimal control of bilinear quantum systems is studied. A framework is developed which results in optimal controls with simple time-frequency structure. This is achieved by the use of cost functionals that promote sparsity in frequency direction and smoothness in time. Existence of minimizers and necessary optimality conditions of the resulting nonsmooth and nonconvex optimization problem are analyzed. Efficient numerical solution methods, in particular for the treatment of the quantum system, are studied. The framework is applied to a problem from molecular control.

Zusammenfassung

Die vorliegenden Arbeit behandelt die optimale Steuerung bilinearer Quantensysteme. Es wird ein Framework entwickelt, das auf optimale Kontrollen mit einfacher Zeit-Frequenz-Struktur führt. Diese wird durch die Verwendung von Kostenfunktionalen erreicht, die „Sparsity“ in Frequenzrichtung und Glattheit in der Zeit begünstigen. Die Existenz von Minimierern und notwendige Optimalitätsbedingungen des resultierenden nichtkonvexen und nicht-glatten Optimierungsproblems wird analysiert. Es werden effiziente numerische Lösungsverfahren, insbesondere für die Behandlung des Quantensystems, untersucht. Das Framework wird auf ein Kontrollproblem für Moleküle angewendet.

Acknowledgement

I would like to thank my supervisor Gero Friesecke for giving me the opportunity to work on this interesting project. I greatly benefited from the stimulating discussions with him on the mathematics and physics of quantum control.

I also thank my second supervisor Karl Kunisch for providing deep insights in the area of optimal control and his continuous support in developing my own ideas.

I gratefully acknowledge the funding provided by the German Research Foundation (DFG) through the International Research Training Group (IGDK) 1754 „Optimization and Numerical Analysis for Partial Differential Equations with Nonsmooth Structures“.

I thank Konstantin Pieper and Philip Trautmann for introducing me to the concept of vector measures in optimal control and for our countless discussions, without which this thesis would not have been possible. I also thank Manfred Liebmann for his patience in introducing me to the ideas of high performance computing.

I thank my colleagues Yuen Au Yeung, Michael Fauser, Frank Hofmaier, Dominik Jüstel, Christian Mendl and all the other members of the analysis group, and all the members of the IGDK in Munich and Graz for helpful discussion and making the last three years such an enjoyable time.

Last but not least, I thank my family — my sisters Caro and Tina, my mother Irmgard and my father Karl — for their ongoing support throughout my studies.

Contents

1 Introduction

This thesis is devoted to the optimal control of quantum mechanical systems. The main contributions are the development of a new framework for sparse quantum control and the numerical solution of such problems with efficient numerical solution methods. The framework is designed to generate controls with a simple time-frequency structure using sparsity enhancing functionals.

The goal of quantum control is to manipulate in some desired way a system which is described by quantum mechanics. Examples range from steering chemical reactions to the experimental realization of quantum computers. The main application in this work is the control of a molecule via the electric field generated by a laser source. Typical control goals include conformation changes of the molecule and selective dissociation of bonds. Control fields are often computed using optimal control theory. But the fields obtained using standard optimal control methods exhibit a complicated oscillating structure. This makes it difficult to interpret the field and to identify underlying control mechanisms. It also prohibits the direct implementation in experiments.

In this work a new framework for the optimal control of quantum systems is proposed that results in controls with a very simple time-frequency structure. This is achieved by controlling via the time-frequency representations of the electric field in combination with functionals that promote sparsity in the frequencies and smoothness in time. This results in optimal control problems of the form

$$\text{Minimize}_{u,\psi} \quad \frac{1}{2}\langle \psi(T), \mathcal{O}\psi(T)\rangle + \alpha\|u\|_{\mathcal{M}}, \tag{1.1}$$

$$\text{s.t. } i\partial_t \psi = H(u)\psi, \quad \psi(0) = \psi_0. \tag{1.2}$$

Here, the state equation (1.2) describes the evolution of the quantum state ψ subject to the time-dependent Hamiltonian H for a given time-frequency control u in the measure space $\mathcal{M}$. Subject to this equation, one seeks to minimize a cost function (1.1) consisting of the quantum mechanical expectation value of the observable $\mathcal{O}$ at time T and a cost term given by the total variation norm of u. The main goal of the thesis is to explain why this is a useful approach, provide a rigorous mathematical theory for optimal control

problem, and explain how it can be solved.

There are always two main aspects in the study of an optimal control problem. The first part is the theoretical analysis of the problem, which consists of the modeling and the solution theory of the state equation and of the optimal control problem. The second part is the numerical treatment of the state equation and the optimal control problem. The contributions of this thesis in those two areas are as follows.

- The theoretical part of this thesis sets up a mathematical framework for sparse time-frequency control of bilinear quantum systems. It includes the solution theory of the governing equations with weak assumptions on initial data, the observation and the coupling operators. This lays the foundation for a rigorous analysis of the proposed nonsmooth and nonconvex optimal control problem. A proof of existence of minimizers is given and a rigorous formulation and derivation of first-order optimality conditions of the optimal control problem are provided.

- The main contribution on the numerical side is the analysis of a time stepping method, a generalized Suzuki–Trotter type method, in the context of optimal control. Using this efficient and flexible method, the behavior of optimal controls for different quantum systems and different realizations of our general control framework is studied.

The field of optimal quantum control was initiated by the pioneering work of Pierce, Daleh and Rabitz [PDR88], and has led to many interesting applications [BKT01; Hoh+07; RWP09; Ass+13]. There are several publications that deal in various ways with the oscillating structure of control fields [Ren+06; CB06; Lap+09; WB08; Hin+13; KHK10; SSB10; Rue+11]. There, mainly the frequency structure of control fields is considered. Our framework provides a systematic optimal control approach to handling the frequency structure of controls, and also supplies the crucial extension to time-frequency controls. This generalization is essential for simplifying the structure of control fields. Mathematically, the new framework is based the idea of sparsity enhancing L^1 type minimization [FR08; Sta09; HSW12]. More specifically it builds on the work by Kunisch, Pieper, Vexler [KPV14] on optimization with function-valued measures for parabolic equations. It extends the results therein in several ways. First, we deal with nonlinear control problems, as arise in bilinear quantum control, and not just with linear ones. Second, we propose the use of control operators, accounting for the nontrivial time-frequency structure of controls dictated by quantum physics. Third, we work with measures with values in space of smooth functions. The combination of control operators

with smooth function spaces provides additional flexibility in the optimal control with measures.

The generalized Suzuki–Trotter type time-stepping analyzed in this thesis was proposed by Liebmann [Lie00] and is based on work of Suzuki [Suz90]. In our case it results in an explicit polynomial approximation of the exponential function that describes the time evolution of the quantum system. The structure of the method makes it easy to parallelize, which, in view of the current developments in computational hardware, is essential to be competitive. For the optimization we will follow a first-discretize-then-optimize approach, where we take special care to give memory efficient implementations. The memory efficiency will make the application of the method to large scale problems possible.

The optimal control of quantum systems has an interdisciplinary component. Typical applications are generated by physicists and chemists, who often show a remarkable intuition for useful optimal control formulations of their problem. On the other hand, mathematicians show their strength in analyzing given control problems. This thesis tries to combine a control approach motivated by quantum mechanics with recently developed mathematical tools: time-frequency controls are used in combination with function-valued measures. A lot of the value of this thesis lies in the proposition of this framework. It was generated in an iteration of proposing a framework, analyzing the computed optimal controls, and using the results to propose an adapted framework. I by no means claim that our framework is a fixed point of this iteration. But I hope that it can help to close the gap between the mathematical optimal control community and the chemists and physicists who apply control theory in practice.

The remainder of the thesis is structured as follows. The next three chapters constitute the control theoretic part of the thesis. Thereafter, two chapters on the numerical implementation are given. We close with a chapter on applications. In the following, the content of the individual chapters is summarized.

Dynamics of Quantum Systems with Time-Dependent Fields In Chapter 2 we will provide the analysis of the state equation, which forms the constraint in our optimal control problem. It starts by giving examples of central evolution equations and their interaction with control fields. Then the abstract quantum setting is defined and existence and compactness properties for solutions are derived. We also study the behavior of solutions with respect to differentiation in the direction of the field. In this thesis the concept of mild solutions is used since it provides a natural framework for existence under weak assumptions on initial data and observation and coupling operators.

Sparse Time-Frequency Control of Quantum Systems In Chapter 3 a general framework for sparse time-frequency control of quantum systems is proposed. It starts with a summary of the standard approach to optimal quantum control. This motivates our new framework, which uses function-valued measures for the control in the time-frequency plane. We give a rigorous definition of the framework and prove the existence of optimal solutions. Our approach is illustrated with several examples. We close the chapter by putting the proposed framework in the context of the existing literature.

Necessary Optimality Conditions In Chapter 4 we derive first order optimality conditions for our abstract optimization problem. Its starts with the setup of adjoint equations to provide expressions for derivatives of expectation values in the direction of the control. Then we derive first order optimality conditions for solutions of the optimal control problem in the measure space. We will also have a close look at these conditions for particular time-frequency controls and derive results on additional regularity of optimal controls and their control fields.

A Generalized Suzuki–Trotter type Method in Optimal Control In Chapter 5 a generalized Suzuki–Trotter type time-stepping is analyzed in the context of optimal control problems. First, the method is introduced and structural properties of it are studied. Also, we will provide expressions for derivatives of the time stepping in the direction of the control field. Then, those results will be used to find representations of derivatives of discrete expectation values where the state is discretized using the generalized Suzuki–Trotter method. We will see that the adjoint state equation can be discretized using the adjoint time-stepping.

The Discrete Optimization Problem In Chapter 6 we formulate the discrete optimization problem and explain it can be solved. We first discuss the discretization of the measure space and the control operator to obtain a discrete problem. We then regularize the problem to obtain a smooth optimization problem. The chapter concludes with a discussion of the optimization methods used for the numerical experiments.

Applications In Chapter 7 the framework is applied to two quantum systems. The first is a discrete system with three levels, the second a system of two coupled one-dimensional Schrödinger equations. For the simpler first example, we give an in-depth analysis of the structure of the optimal control and study the optimality conditions. In the more challenging second example, we compare optimal controls for different realizations of the

new control framework to controls obtained using the standard approach.

2 Dynamics of Quantum Systems with Time-Dependent Fields

In this chapter the analysis of the quantum system is given that serves as a constraint in the optimization problem. After a short discussion of model quantum systems, we formulate the abstract quantum setting used throughout the thesis. We present existence and differentiability results for solutions of the time-dependent quantum systems. We also derive a compactness result important for the optimal control theory in the next chapter.

2.1 Central Evolution Equations with Controls

In this section we present two interesting models in the bilinear control of quantum systems. We also sketch how time dependent fields interact with quantum systems.

2.1.1 Models

The dynamics of the quantum systems we study is described by the equation

$$i\partial_t\psi(t) = \big(H_0 + \sum_{l=1}^{L} E_l(t)H_l\big)\psi(t). \tag{2.1}$$

where the state $\psi(t)$ belongs to some Hilbert space $\mathcal{H}$, H_0 and the H_l are possibly unbounded self-adjoint operators in $\mathcal{H}$ and the E_l are real-valued amplitudes of electric or magnetic fields.

Here we already use important simplifications of the field–matter interaction [DAl08, Section 2.2]. We do not model the backaction of the quantum system on the electric field using quantum field theory but just use a classical field strength E_l. Actually modeling the quantum field would result in a much more difficult problem. We also assume that the field is just a scalar function of time and that it enters the equation in a linear fashion. These assumptions are not essential and control theoretical work has been done without them [IK07; Tur07]. Since the focus of this thesis lies on the control approach,

we restrict our attention to the case of equation 2.1. We expect that most of the control concepts are also applicable in a more general setting.

We now discuss two examples fitting in this setting.

Example 1. We consider the spin of a spin-1/2 particle in a magnetic field. The spin of the particle is described by a unit vector in the space $\mathcal{H} = \mathbb{C}^2$. The evolution in a time-dependent magnetic field $B\colon \mathbb{R} \to \mathbb{R}^3$ is given by

$$i\partial_t \psi(t) = \gamma \frac{1}{2}(\sigma_x B_x(t) + \sigma_y B_y(t) + \sigma_z B_z(t))\psi(t) \tag{2.2}$$

with the Pauli matrices

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

and the gyromagnetic ratio γ of the particle. One often applies a time-independent field in direction z and controls the spin with the fields in the x and y directions. Equation (2.2) can then be written in the form of equation 2.1 with $L = 2$ by setting $E_1 = B_x$, $E_2 = B_y$,

$$H_0 = \gamma \frac{1}{2}\sigma_z B_z, \quad H_1 = \gamma \frac{1}{2}\sigma_x, \quad H_2 = \gamma \frac{1}{2}\sigma_y.$$

For theoretical purposes it is sometimes useful to write this model in the form

$$i\partial_t \psi(t) = \left(H_0 + \gamma \frac{1}{2} \begin{pmatrix} 0 & E^*(t) \\ E(t) & 0 \end{pmatrix} \right) \psi(t)$$

with the complex valued function $E(t) = E_1(t) + iE_2(t)$.

This model is widely studied in the literature, as a basic example in NMR and more recently as a model of a single qubit in quantum information theory. The simplicity of the model make it a reasonable starting point to explore control phenomena analytically. The basic case of a time-harmonic control field is analyzed in a classical paper by Rabi [Rab37]. It is considered in the context of optimal control in [Bos+02][WG07, A.3][DAl08, p. 6.3]. A recent careful experimental realization of this system is described in [Sch+14].

Example 2. In this example we study a multi level Born–Oppenheimer approximation. It describes the evolution of nuclei on potential energy surfaces generated by the electrons. The model can be used to study laser guided chemical reactions. For a number

$M \in \mathbb{N}$ of surfaces the equation reads

$$i\partial_t \psi(t) = (\hat{T} + \hat{V} + E(t)\hat{\mu})\psi(t) \tag{2.3}$$

Here, the state $\psi(t) = (\psi_1(t), \dots, \psi_M(t))^T$ lies in the Hilbert space $L^2(\mathbb{R}^n; \mathbb{C}^M)$, $\hat{T} = \mathrm{diag}(T, \dots, T)$ is an operator-valued matrix with the kinetic energy operator T on the diagonal, $\hat{V} = (V_{ij})$ is an operator-valued matrix of potentials $V_{ij} = \overline{V_{ji}}$, $\hat{\mu}$ is an operator-valued matrix of transition dipole operators of the electric field, and E is the real-valued electric field strength. In the case $M = 1$, we obtain the standard Born–Oppenheimer approximation for the evolution of the atomic nuclei. The model can be derived similarly to the standard Born-Oppenheimer approximation by projection of the full molecular Hamiltonian on spaces generated by the eigenstates of the electronic part of the full Hamiltonian for a vanishing control field. A careful mathematical account in the absence of control fields is given in [Teu03]. Equation (2.3) fits in the setting of equation (2.1) with $L = 1$ and the unbounded operator $H_0 = \hat{T} + \hat{V}$ and the possibly unbounded operator $H_1 = \hat{\mu}$. For simplicity we will only study the case $Vij = 0$ for $i \neq j$, and without intersections of the V_{ii}.

This model is studied in a control context since the beginning of quantum control [PDR88; Kos+89; GNR92] and is still widely used [KHK10; Hof+12]. There exist extensions of the model. One approach is to project the full Hamiltonian on field dependent electronic spaces [Bal+05]. It is also possible to consider exact factorizations of the wave function instead of a projection ansatz [AMG10; Abe+13]. There is also some work on the optimal control in a semiclassical regime [KN08].

For practical and theoretical purposes, it is often interesting to approximate infinite dimensional systems by finite dimensional ones. The dimension of the finite dimensional approximation varies from small numbers corresponding to handpicked physically relevant degrees of freedom to large numbers coming from discretization of underlying partial differential equations. In any case, we obtain matrices H_0 and H_l on the state space $\mathbb{C}^N$ for some $N \in \mathbb{N}$.

2.1.2 Transition Frequencies

Numerical experiments for the optimal control of quantum systems of the form (2.1) show that oscillating fields are necessary for most control purposes. This is clear from a physical point of view since one needs photons with an energy proportional to the difference in the eigenvalues to induce transitions between eigenstates of a quantum system. This energy of the photon is proportional to its frequency.

The effect can also be understood mathematically. In the simplest case, for the spin system presented in Example 1, the coordinate transformation $\psi(t) \mapsto \tilde{\psi}(t) = \exp(iH_0t)\psi(t)$, that leaves the eigenspaces invariant, results in the equation

$$i\partial_t\tilde{\psi}(t) = \begin{pmatrix} 0 & E(t)^* e^{i\gamma B_z t} \\ E(t)e^{-i\gamma B_z t} & 0 \end{pmatrix} \tilde{\psi}(t).$$

For time-harmonic fields given by

$$E_1(t) = A\cos(\omega t), \quad E_2(t) = A\sin(\omega t) \tag{2.4}$$

this simplifies to

$$i\partial_t\tilde{\psi}(t) = \begin{pmatrix} 0 & Ae^{i(\gamma B_z - \omega)t} \\ Ae^{-i(\gamma B_z - \omega)t} & 0 \end{pmatrix} \tilde{\psi}(t).$$

One can explicitly solve the evolution of this equation. For an initial state $(1,0)^T$ we obtain

$$|\tilde{\psi}_2(t)|^2 = \frac{\Delta^2}{\Omega^2}\sin^2\left(\frac{\Omega}{2}t\right) \tag{2.5}$$

for the population of the second level, where $\Delta^2 = (\omega - \gamma B_z)^2$ and $\Omega = (\gamma A)^2 + \Delta^2$. This formula is already given by Rabi, see [Rab37, (12)]. In Figure 2.1 we see plotted $|\psi_2(t)|^2$ for the initial state $(1,0)^T$ for different ω. In accordance with (2.5), we see that the field with a frequency matching the difference in eigenvalues induces full transitions to the second eigenspace spanned by $(0,1)^T$. Using a different frequency only induces partial transitions. Similar results for transitions between eigenstates also hold in a long-time–small-amplitude limit for more general finite dimensional quantum systems [DAl08, p. 7.1] and also for particular infinite dimensional systems [Cha12]. For the quantum system from Example 2, the situation is more complicated since we do not just want to introduce transition between eigenstates. In particular we also want to induce transitions of the state between the different surfaces, that is in the electronic structure, while staying at the same point in space. These states do not correspond to eigenstates of the uncontrolled Hamiltonian. Numerical experiments show that for states well localized in space those transitions can be induced by frequencies roughly proportional to the difference in potentials at this location. In Figure 2.2 we plotted two fields and the corresponding populations on the second surface for a problem with two surfaces. Here, the situation becomes more complicated. Since the states move on the surfaces, the difference in the potentials varies in time. To induce a transition, we therefore also need to have a field localized in time. Transitions between surfaces are

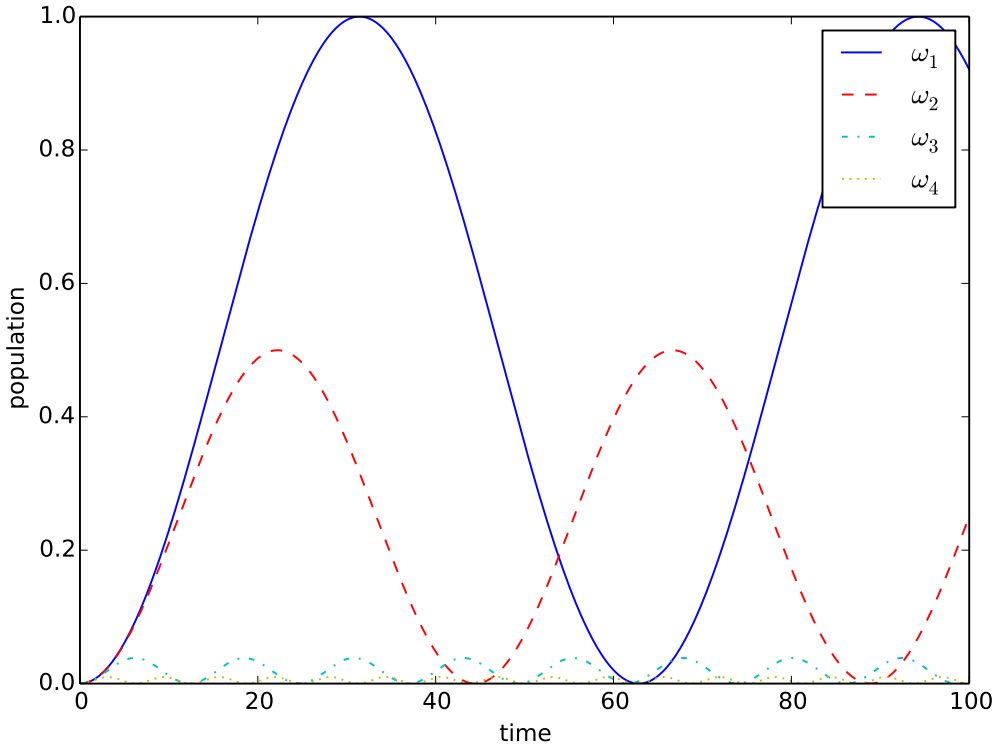
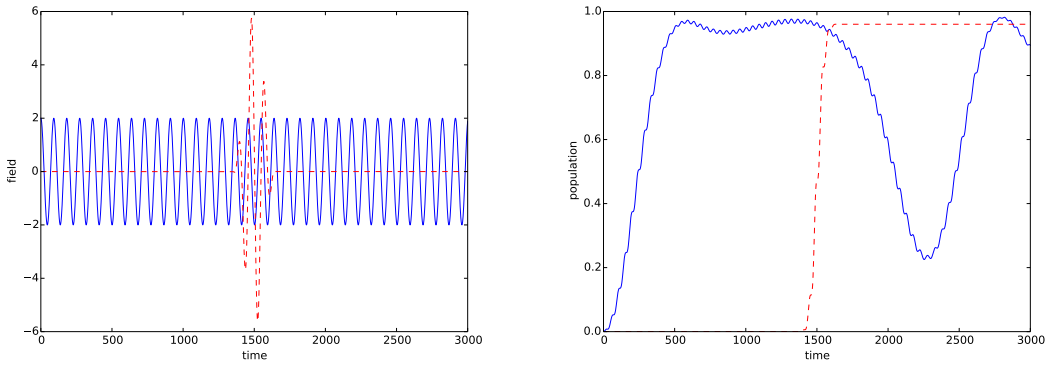


Figure 2.1: Time evolution of populations of the second eigenspace for the spin system. The control fields are given by (2.4) for the frequencies $\omega_1 = \gamma B_z$ (solid), $\omega_2 = 1.05\gamma B_z$ (dashed), $\omega_3 = 1.25\gamma B_z$ (dash-dotted), $\omega_4 = 1.5\gamma B_z$ (dotted).



(a) (b)

Figure 2.2: Time evolution of (a) two control fields (solid and dashed) and (b) their corresponding populations for the second surface (solid and dashed, respectively) for a two-level Born–Oppenheimer system.

studied in the theory of nonadiabatic transitions [Nak12].

2.2 Abstract Bilinear Quantum Systems

In this section we present the functional analytic setting for the quantum systems we want to consider and discuss the solution theory for the equation. We present results on the existence and differentiability of mild solutions to (2.6). Then we prove a compactness result for the solution.

2.2.1 Functional Analytic Setting

The functional analytic setting presented here will be the foundation for the rigorous results in the following chapters. It is based on the concept of mild solutions.

We consider quantum systems with a bilinear control interaction of the form (2.1). The uncontrolled quantum system has the *drift Hamiltonian* H_0. The coupling is introduced through the fields E_l, $l = 1, \cdots, L$ for some $L \in N$, and the *coupling Hamiltonians* H_l for $l = 1, \ldots, L$. Let $\mathcal{H}$ be a Hilbert space. We will throughout this thesis use the following two assumptions for the Hamiltonians.

(A1) The drift Hamiltonian H_0 is a possibly unbounded self-adjoint linear operator in $\mathcal{H}$.

(A2) The coupling Hamiltonians H_l are bounded self-adjoint linear operators for $l = 1, \ldots, L$.

The first assumption guarantees that the drift Hamiltonian generates a unitary group G via the relation $G(t) = \exp(-itH_0)$ for $t \in \mathbb{R}$ [BEH08, Prop. 5.9.1]. The assumption on the coupling Hamiltonians will make a solution theory for control fields with low regularity accessible through mild solutions.

Both assumptions (A1) and (A2) are satisfied for Example 1 and in discretized finite dimensional setting for Example 2. In the infinite dimensional case for Example 2, assumption (A1) results in assumptions on the potentials. In the diagonal case, one can essentially use results for the uncoupled case [BEH08, p. 14.1], combined with the result that direct sums of self-adjoint operators are again self-adjoint. The boundedness assumption in (A2) is a restriction in some models. It is not possible to consider the unbounded dipole operator on $\mathbb{R}^n$. In this case a restriction to a bounded domain can be used [WBV10].

For the solution theory, we will typically assume the following regularity on initial data and the control fields. The *initial state* ψ_0 satisfies $\psi_0 \in \mathcal{H}$. This means we only

assume the natural regularity given by the Hilbert space $\mathcal{H}$. The *control fields* E_l are given by the components of the multicomponent control field $E \in L^1(0,T;\mathbb{R}^L)$ for some $T > 0$. The L^1 regularity is the natural assumption in the theory of mild solutions. For optimal control we will often have to work in the space $L^p(0,T;\mathbb{R}^L)$ for some $1 < p \leq \infty$. The reason lies in the needed compactness results described in Section 2.2.3. We will work in the setting of mild solutions. The mild formulation of equation (2.1) is

$$\psi(t) = G(t)\psi_0 + \int_0^t G(t-s)E(s) \cdot \tilde{H}\psi(s)\,\mathrm{d}s, \tag{2.6}$$

where we used the short hand notation $v \cdot \tilde{H} := \sum_l v_l(-iH_l)$ for $v \in \mathbb{R}^L$. The vector valued integral here and in the following should be understood in the sense of Bochner integrals [AB06, p. 11.8]. In Section 2.2.2 we will see that this equation indeed has a unique solution. The solution ψ lies in the space $C([0,T];\mathcal{H})$.

Other authors use stronger assumptions on the control field, initial state and the coupling Hamiltonians [IK07; WBV10; Hin+13]. The typical variational setting can be obtained by working in spaces derived from the drift Hamiltonian H_0. Assuming positivity of H_0, one can define the space $\mathcal{V} = \mathcal{D}(H_0^{\frac{1}{2}})$, the domain of definition of the square root of H_0. Then $\psi_0 \in \mathcal{V}$, $H_l \in \mathcal{B}(\mathcal{H}) \cap \mathcal{B}(\mathcal{V})$, $E \in L^2(0,T;\mathbb{R}^L)$, leads to the additional regularity $\psi \in C([0,T];\mathcal{V}) \cap H^1(0,T;\mathcal{V}^*)$. Together with compact embeddings $\mathcal{V} \hookrightarrow \mathcal{H}$, compactness properties of the equations are easier to show. The theoretical treatment of optimal control for the Schrödinger equation then is very similar to the control of parabolic equations, compare [WBV10] and [Trö10]. We choose to present the optimal control of quantum systems without those additional regularity assumptions and will stay in the setting of mild solutions. There is also work with weaker assumptions on the coupling. In [FZC14] a coupling through singular potentials is considered, albeit with stronger assumptions on the field.

All spaces — also those containing complex valued objects — are equipped with a real Banach or Hilbert space structure. That means linear always means $\mathbb{R}$-linear and the scalar product is real-valued and $\mathbb{R}$-bilinear. A complex Hilbert space with scalar product $\langle\cdot,\cdot\rangle^{\mathbb{C}}$ becomes a real Hilbert space by using the scalar product $\langle\cdot,\cdot\rangle := \langle\cdot,\cdot\rangle_{\mathcal{H}}^{\mathbb{R}} := \mathrm{Re}\langle\cdot,\cdot\rangle_{\mathcal{H}}^{\mathbb{C}}$. This change of Hilbert space structure does not change the topology. This change corresponds to considering $\mathbb{C}$ as a two-dimensional real vector space. This mathematical trick is often used in theory [HTC83, pp. 2613][IK07]. The real structure of the spaces will not be exploited in this chapter, but it will become important when the differentiability of the cost functional is considered. Using real spaces, we will avoid the additional regularity inherent in the theory of complex differentiation. In particular, the

derivatives we will obtain will not be linear maps with respect to the complex structure, but they will be with respect to the real structure. Also see the remark and reference in [WBV10, p. 4177].

We will use the following notations. Normed spaces X are equipped with a norm $\|\cdot\|_X$, Hilbert spaces X are equipped with a scalar product $\langle\cdot,\cdot\rangle_X$ and the norm induced by the scalar product. The space of bounded linear operators between the normed spaces X and Y is denoted by $\mathcal{B}(X,Y)$ and equipped with the operator norm. We write $\mathcal{B}(X)$ for $\mathcal{B}(X,X)$. The scalar product in R^L is denoted by $v\cdot w$ for $v,w\in\mathbb{R}^L$. In addition to the notation $v\cdot\tilde{H}=\sum_{l=1}^L v_l(-iH_l)$, we will write $v\cdot\tilde{H}^*:=\sum_{l=1}^L v_l(-iH_l)^*$ and $\langle\chi_1,\tilde{H}\chi_2\rangle_{\mathcal{H}}:=(\langle\chi_1,-iH_l\chi_2\rangle_{\mathcal{H}})_{l=1}^L\in\mathbb{R}^L$ for $v\in\mathbb{R}^L$ and $\chi_1,\chi_2\in\mathcal{H}$. We will write $\|\tilde{H}\|:=(\sum_l\|H_l\|^2_{\mathcal{B}(\mathcal{H})})^{1/2}$. We will use the standard notation for Bochner and Sobolev spaces.

2.2.2 Existence and Differentiability of Solutions

We will now discuss the existence of solutions to (2.6) and their differentiability in the direction of the control. The results are small variations of what can already be found in the work of Ball, Marsden and Slemrod [BMS82]. We will present the proofs of the unitarity of the time evolution and of the differentiability of the state. The first proof is instructive and prototypical of proofs in the setting of mild solutions. The second proof is elementary and a little more detailed than the proof in [BMS82].

We look at the existence of the inhomogeneous equation

$$i\partial_t\psi(t)=(H_0+E(t)\cdot i\tilde{H})\psi(t)+f(t).$$

The inhomogeneity is not needed for the existence of the state itself, the evolution of a closed quantum system is described by a homogeneous equation. The derivative in the direction of the control, however, satisfies an inhomogeneous equation. This makes it useful to consider existence also in this case.

Proposition 1. *Let $\psi_0\in\mathcal{H}$, $E\in L^1(0,T;\mathbb{R}^L)$ and $f\in L^1(0,T;\mathcal{H})$. Then there exists a unique function $\psi\in C([0,T];\mathcal{H})$ satisfying*

$$\psi(t)=G(t)\psi_0+\int_0^t G(t-s)E(s)\cdot\tilde{H}\psi(s)\,\mathrm{d}s+\int_0^t G(t-s)f(s)\,\mathrm{d}s. \tag{2.7}$$

for all $t\in[0,T]$.

The proof relies on a fixed point argument for short times and the extension to large times by using naive stability estimates [BMS82, Prop. 2.1, Thm. 2.5]. The restriction

to a finite time horizon is not essential here. But since we only need such results in our optimal control framework, we restrict ourself to this case.

For given $E \in L^1(0,T;\mathbb{R}^L)$ and $f = 0$ there is a unique function

$$\mathfrak{G}\colon \{\, (t,s) \in [0,T] \mid 0 \le t \le s \le T \,\} \to \mathcal{B}(\mathcal{H})$$

such that the solution ψ of (2.6) satisfies

$$\mathfrak{G}(t,s)\psi(s) = \psi(t), \quad s \le t,$$

for all initial states ψ_0 [LY95, Chapter 2, §5.3]. We call $\mathfrak{G}$ the *evolution operator* for the field E. Using $\mathfrak{G}$ we can write solutions to (2.7) as

$$\begin{aligned} \psi(t) &= \mathfrak{G}(t,s)\psi(s) + \int_s^t \mathfrak{G}(t,s) f(s)\,\mathrm{d}s, \quad s \le t \\ \psi(0) &= \psi_0. \end{aligned} \tag{2.8}$$

We will switch between the representations of solutions given by (2.7) and (2.8) as needed.

For optimal control purposes, we will also have to consider the solution φ of the equation adjoint to (2.7),

$$\varphi(t) = G(T-t)^*\varphi_T + \int_t^T G(s-t)^* E(s) \cdot \tilde{H}^* \varphi(s)\,\mathrm{d}s + \int_t^T G(s-t)^* f(s)\,\mathrm{d}s, \tag{2.9}$$

for $\varphi_T \in \mathcal{H}$. In the strong form this reads

$$i\partial_t \varphi(t) = (H_0 + E(t) \cdot i\tilde{H})\varphi + f, \quad \varphi(T) = \varphi_T.$$

This directly translates into taking the adjoint of the evolution operator. We have

$$\begin{aligned} \varphi(s) &= \mathfrak{G}(t,s)^*\varphi(t) + \int_s^t \mathfrak{G}(t,s)^* f(s)\,\mathrm{d}s, \quad s \le t, \\ \varphi(T) &= \varphi_T, \end{aligned} \tag{2.10}$$

again see [LY95, Chapter 2, §5.3]. Using the variable substitution $t \mapsto T - t$ we see that φ can be written in the form (2.7),

$$\varphi(T-t) = G(t)^*\varphi_T + \int_0^t G(t-s)^* E(T-s) \cdot \tilde{H}^* \varphi(T-s)\,\mathrm{d}s + \int_0^t G(t-s)^* f(T-s)\,\mathrm{d}s,$$

for the operators $-H_0$ and $-H_l$ which still satisfy (A1) and (A2). This makes results for the ψ also applicable to φ.

The next lemma is concerned with the unitarity of the evolution operator $\mathfrak{G}$. In contrast to the results of the preceding propositions, it makes explicit use of the self-adjointness of the Hamiltonian and the coupling operators. The typical proof of unitarity goes as follows. We have

$$\frac{\mathrm{d}}{\mathrm{d}t}\|\psi(t)\|_{\mathcal{H}}^2 = 2\langle\psi(t), (-iH_0 + E(t)\cdot\tilde{H})\psi(t)\rangle_{\mathcal{H}},$$

and since $-iH_0$ and $E(t)\cdot\tilde{H}$ are skew-adjoint, this implies that the norm is conserved in time. Note that we do not have to take the real part here since we work with the real scalar product structure only. This proof of unitarity does not work in the setting of mild solutions because the state is not differentiable in a suitable sense. One can overcome this difficulty by approximation arguments. We will give a proof that stays in the setting of mild solutions.

Lemma 2. *The evolution operator $\mathfrak{G}$ satisfies*

$$\mathfrak{G}(t,s)^*\mathfrak{G}(t,s) = I = \mathfrak{G}(t,s)\mathfrak{G}(t,s)^* \tag{2.11}$$

for all $0 \le s \le t \le T$.

Proof. Let $\chi \in \mathcal{H}$ and $0 \le s \le t \le T$, and set $\phi(t) = \mathfrak{G}(t,s)\chi$. We will show

$$\chi = \mathfrak{G}(t,s)^*\phi(t), \tag{2.12}$$

which implies the first equality in (2.11). Let $r \in [s,t]$. By (2.9) and (2.10) for $f = 0$, we know that $\varphi(r) = \mathfrak{G}(t,r)^*\phi(t)$ solves

$$\varphi(r) = G(t-r)^*\phi(t) + \int_r^t G(q-r)^*E(q)\cdot\tilde{H}^*\varphi(q)\,\mathrm{d}q.$$

We will show that $\phi = \varphi$. For the difference we obtain, using the group property of G

and the self-adjointness of the H_l,

$$\begin{aligned}
\phi(r) - \varphi(r) = {} & G(r)\chi + \int_s^r G(r-q)E(q)\cdot\tilde{H}\phi(q)\,\mathrm{d}q \\
& - G(t-r)^*\phi(t) - \int_r^t G(q-r)^*E(q)\cdot\tilde{H}^*\varphi(q)\,\mathrm{d}q \\
= {} & G(r)\chi + \int_s^r G(r-q)E(q)\cdot\tilde{H}\phi(q)\,\mathrm{d}q \\
& - G(t-r)^*\left(G(t)\chi + \int_s^t G(t-q)E(q)\cdot\tilde{H}\phi(q)\,\mathrm{d}q\right) \\
& - \int_r^t G(q-r)^*E(q)\cdot\tilde{H}^*\varphi(q)\,\mathrm{d}q \\
= {} & -\int_r^t G(q-r)^*E(q)\cdot\tilde{H}\Big(\phi(q)-\varphi(q)\Big)\,\mathrm{d}q.
\end{aligned}$$

Therefore we obtain

$$\|\phi(r)-\varphi(r)\|_{\mathcal{H}} \le \int_r^t \sum_{l=1}^{L} |E(q)_l|\|H_l\|\|\phi(q)-\varphi(q)\|_{\mathcal{H}}\,\mathrm{d}q.$$

Applying Gronwall's inequality, e.g. the version [BMS82, Lem. 2.6], to $f(r) = \|\phi(t-r)-\varphi(t-r)\|$ yields $\phi(r)=\varphi(r)$ for all $r\in[s,t]$. Thus we have

$$\chi = \varphi(s) = \mathfrak{G}(t,s)^*\varphi(t) = \mathfrak{G}(t,s)^*\phi(t)$$

which is (2.12).

Applying this result to the adjoint equation gives

$$I = \mathfrak{G}(t,s)^{**}\mathfrak{G}(t,s)^* = \mathfrak{G}(t,s)\mathfrak{G}(t,s)^*,$$

which proves the second equality in (2.11). □

As a direct consequence we obtain the following stability estimates for solutions of (2.7).

Corollary 3. *Let ψ be the solution of (2.7) for given $\psi_0 \in H$, $E \in L^1(0,T;\mathbb{R}^L)$ and $f \in L^1(0,T;\mathcal{H})$. Then ψ satisfies*

$$\|\psi\|_{C([0,T];\mathcal{H})} \le \|\psi_0\|_{\mathcal{H}} + \|f\|_{L^1(0,T;\mathcal{H})}. \tag{2.13}$$

If $f = 0$ then for all $t \in [0,T]$, we have

$$\|\psi(t)\|_{\mathcal{H}} = \|\psi_0\|_{\mathcal{H}}. \tag{2.14}$$

Corresponding results hold for the solution φ of (2.9).

Proof. By Lemma 2 we have that for all $t \in [0,T]$,

$$\|\mathfrak{G}(t,0)\psi_0\|_{\mathcal{H}}^2 = \langle \psi_0, \mathfrak{G}(t,0)^* \mathfrak{G}(t,0)\psi_0 \rangle_{\mathcal{H}} = \langle \psi_0, \psi_0 \rangle_{\mathcal{H}} = \|\psi_0\|_{\mathcal{H}}^2.$$

In the case $f = 0$ we therefore have (2.14). In general we obtain the estimate

$$\begin{aligned} \|\psi(t)\|_{\mathcal{H}} &\leq \|\mathfrak{G}(t,0)\psi_0\|_{\mathcal{H}} + \int_0^t \|\mathfrak{G}(t,s)f(s)\|_{\mathcal{H}} \,\mathrm{d}s \\ &\leq \|\psi_0\|_{\mathcal{H}} + \|f\|_{L^1(0,T;\mathcal{H})}. \end{aligned}$$

Taking the maximum over all $t \in [0,T]$ yields (2.13). The results for solutions of the adjoint equation can be proven in the same way. □

We now study the differentiability of the solution of (2.7). The next proposition covers the differentiability in the direction of the control field, as well as differentiability in the direction of the initial state and the inhomogeneity. We will not follow the proof idea in [BMS82], but will explicitly check that a candidate for the derivative satisfies the definition of differentiability.

Proposition 4. *The map $F\colon L^1(0,T;\mathbb{R}^L) \times \mathcal{H} \times L^1(0,T;\mathcal{H}) \to C([0,T];\mathcal{H})$ given by $F(E,\psi_0,f) = \psi$ where ψ solves* (2.7) *is continuously differentiable. The derivative is given by $F'(E,\psi_0,f)(\delta E, \delta\psi_0, \delta f) = \psi'$ where ψ' solves the equation*

$$\psi'(t) = \mathfrak{G}(t,0)\delta\psi_0 + \int_0^t \mathfrak{G}(t,s)\delta E(s) \cdot \tilde{H}\psi(s) \,\mathrm{d}s + \int_0^t \mathfrak{G}(t,s)\delta f(s) \,\mathrm{d}s. \tag{2.15}$$

Proof. To prove differentiability we will construct the canonical candidate for the derivative of F and then explicitly check that it actually is the derivative in the proposed sense. Let $E, \delta E \in L^1(0,T;\mathbb{R}^L)$, $\psi_0, \delta\psi_0 \in \mathcal{H}$ and $f, \delta f \in L^1(0,T;\mathcal{H})$. Formally differentiating (2.7) gives

$$\begin{aligned} \psi'(t) = G(t)\delta\psi_0 &+ \int_0^t G(t-s)E(s) \cdot \tilde{H}\psi'(s) \,\mathrm{d}s \\ &+ \int_0^t G(t-s)\delta E(s) \cdot \tilde{H}\psi(s) \,\mathrm{d}s + \int_0^t G(t-s)\delta f(s) \,\mathrm{d}s. \end{aligned} \tag{2.16}$$

Proposition 1 tells us that (2.15) indeed has a unique solution in $\psi' \in C([0,T];\mathcal{H})$. Using the evolution operator $\mathfrak{G}$, equation (2.16) can be equivalently rewritten as (2.15). From (2.15) we see that the map $(\delta E, \delta\psi_0, \delta f) \mapsto \psi'$ is linear, and bounded with

$$\|\psi'\|_{C([0,T];\mathcal{H})} \leq \|\delta\psi_0\|_{\mathcal{H}} + \|\tilde{H}\|\|\psi_0\|_{\mathcal{H}}\|\delta E\|_{L^1(0,T;\mathbb{R}^L)} + \|\delta f\|_{L^1(0,T;\mathcal{H})}. \tag{2.17}$$

We will now show that $F'(E,\psi_0,f)(\delta E,\delta\psi_0,\delta f) = \psi'$. We introduce the notation $\psi_\delta = F(E+\delta E, \psi_0+\delta\psi_0, f+\delta f)$ and define

$$r(t) = \frac{\psi_\delta(t) - \psi(t) - \psi'(t)}{\|\delta E\|_{L^1(0,T;\mathbb{R}^L)} + \|\delta\psi_0\|_{\mathcal{H}} + \|\delta f\|_{L^1(0,T;\mathcal{H})}}$$

for $t \in [0,T]$. We need to show $\sup_t \|r(t)\|_{\mathcal{H}} \to 0$ for $(\delta E, \delta\psi_0, \delta f) \to 0$. Using the definitions of ψ_δ, ψ and ψ', (2.7) and (2.16), we obtain

$$\begin{aligned} r(t) = &\int_0^t G(t-s)(E(s)+\delta E(s))\cdot \tilde{H} r(s)\,\mathrm{d}s \\ &+ \int_0^t G(t-s)\frac{\delta E(s)}{\|\delta E\|_{L^1(0,T;\mathbb{R}^L)} + \|\delta\psi_0\|_{\mathcal{H}} + \|\delta f\|_{L^1(0,T;\mathcal{H})}}\cdot \tilde{H}\psi'(s)\,\mathrm{d}s. \end{aligned}$$

Therefore,

$$\|r(t)\|_{\mathcal{H}} \leq \|\tilde{H}\| \int_0^t \|E(s)+\delta E(s)\|_{\mathbb{R}^L}\|r(s)\|_{\mathcal{H}}\,\mathrm{d}s + \|\tilde{H}\|\|\psi'\|_{C([0,T];\mathcal{H})}$$

By Gronwall's inequality this yields

$$\|r(t)\|_{\mathcal{H}} \leq \|\tilde{H}\|\|\psi'\|_{C([0,T];\mathcal{H})}\exp(\|\tilde{H}\|\|E+\delta E\|_{L^1(0,T;\mathbb{R}^L)}).$$

Using estimate (2.17) this implies $\sup_t\|r(t)\|_{\mathcal{H}} \to 0$ for $(\delta E, \delta\psi, \delta f) \to 0$. Therefore F is Fréchet differentiable.

It remains to show the continuity of F'. By (2.16) we have

$$F'(E,\psi_0,f)(\delta E,\delta\psi_0,\delta f) = F(E,\delta\psi_0,\delta f + \delta E\cdot\tilde{H}F(E,\psi_0,f)). \tag{2.18}$$

Let $(E_n,\psi_{0,n},f_n)$ be a sequence in $L^1(0,T;\mathbb{R}^L)\times\mathcal{H}\times L^1(0,T;\mathcal{H})$ with $(E_n,\psi_{0,n},f_n) \to (E,\psi_0,f)$. We denote by (ψ_n) the corresponding sequence of solutions of (2.7) and by (ψ_n') the sequence of solutions to (2.15),

$$\psi_n' = F'(E_n,\psi_{0,n},f_n)(\delta E,\delta\psi_0,\delta f).$$

Then (2.18) and the continuity of F imply

$$\|\psi_n' - \psi'\|_{C([0,T];\mathcal{H})} \to 0.$$

This convergence is also uniform in $(\delta E, \delta\psi_0, \delta f)$ since by (2.18) and (2.16) we have

$$\begin{aligned}\psi_n'(t) - \psi'(t) = \int_0^t G(t-s)(E_n(s)\cdot\tilde{H}\psi_n'(s) - E(s)\cdot\tilde{H}\psi'(s))\,\mathrm{d}s \\ + \int_0^t G(t-s)\delta E\cdot\tilde{H}(\psi_n(s)-\psi(s))\,\mathrm{d}s\end{aligned}$$

which gives

$$\begin{aligned}\|\psi_n' - \psi'\|_{C([0,T];\mathcal{H})} \leq &\|E_n\|_{L^1(0,T;\mathbb{R}^L)}\|\tilde{H}\|\|\psi_n' - \psi'\|_{C([0,T];\mathcal{H})} \\ &+ \|E_n - E\|_{L^1(0,T;\mathbb{R}^L)}\|\tilde{H}\|\|\psi'\|_{C([0,T];\mathcal{H})} \\ &+ \|\delta E\|_{L^1(0,T;\mathbb{R}^L)}\|\tilde{H}\|\|\psi_n - \psi\|_{C([0,T];\mathcal{H})}\end{aligned}$$

uniformly in $(\delta E, \delta\psi_0, \delta f)$. Therefore F' is continuous. □

We obtain the following corollary for the differentiability of solution to the adjoint equation (2.9).

Corollary 5. *The map $F^*\colon L^1(0,T;\mathbb{R}^L)\times\mathcal{H}\times L^1(0,T;\mathcal{H}) \to C([0,T];\mathcal{H})$ given by $F^*(E,\varphi_T,f) = \varphi$ where φ solves (2.9) is continuously differentiable. The derivative is given by $F^{*\prime}(E,\varphi_T,f)(\delta E,\delta\varphi_T,\delta f) = \varphi'$ where φ' solves the equation*

$$\varphi'(t) = \mathfrak{G}(t,0)^*\delta\psi_0 + \int_0^t \mathfrak{G}(t,s)^*\delta E(s)\cdot\tilde{H}^*\psi(s)\,\mathrm{d}s + \int_0^t \mathfrak{G}(t,s)^*\delta f(s)\,\mathrm{d}s.$$

2.2.3 A Compactness Result

We will show a compactness result in this section. The typical proof of such results in the context of optimal control problems relies on compact embeddings. Since we use a weaker solution concept and do not obtain sufficient regularity for such embeddings, we need to show compactness explicitly. A stronger version of the results in this section can be found in the preprint [FHK15].

In this section we will need a slightly higher regularity for the control field E compared to the preceding section. We will assume $E \in L^p(0,T;\mathbb{R}^L)$ for some p with $1 < p \leq \infty$. We include the case $p = \infty$. We use the notation $E_n \xrightarrow{(*)} E$ in $L^p(0,T;\mathbb{R}^L)$ for weak convergence if $1 < p < \infty$ and for weak-$*$ convergence if $p = \infty$.

Proposition 6. *Let $1 < p \leq \infty$ and let $(E_n)_n$ be a sequence in $L^p(0,T;\mathbb{R}^L)$ such that $E_n \overset{(*)}{\rightharpoonup} E$. Then there exists a subsequence such that the solutions ψ_n of (2.6) corresponding to E_n satisfy $\psi_n(t) \to \psi(t)$ in $\mathcal{H}$ for all t, where ψ is the solution of (2.6) for E.*

Proof. The proof is split up into three steps. First we will show existence of a pointwise weak limit for all $t \in [0,T]$ for some subsequence of $(\psi_n)_n$. Then we prove that this pointwise limit is the mild solution corresponding to the limit field. Then the pointwise strong convergence of the original sequence $(\psi_n)_n$ is established.

Step 1 By Lemma 3 we know that $(\psi_n(t))_n$ is bounded for all $t \in [0,T]$. Using a diagonal sequence argument, there exists a subsequence, again denoted by $(\psi_n)_n$, such that $\psi_n(t)$ converges weakly for all $t \in S$ for some countable dense subset $S \subset [0,T]$. To show weak convergence for all $t \in [0,T]$, we use the fact that the ψ_n solve (2.6). Let $t' \in [0,T] \setminus S$. We show that $(\langle \chi, \psi_n(t')\rangle)_n$ is a Cauchy sequence for each $\chi \in \mathcal{H}$. Let $\chi \in \mathcal{H}$ and $\epsilon > 0$. For $t \in S$ with $t < t'$ we have

$$\langle \chi, \psi_n(t') - \psi_m(t')\rangle = \langle \chi, \psi_n(t') - \psi_n(t)\rangle + \langle \chi, \psi_n(t) - \psi_m(t)\rangle + \langle \chi, \psi_m(t) - \psi_m(t')\rangle \quad (2.19)$$

for all $n, m \in \mathbb{N}$. We first look at the first and last term on the right hand side of (2.19). We have

$$\psi_n(t') - \psi_n(t) = (G(t'-t) - I)\psi_n(t) + \int_t^{t'} G(t'-s)E_n(s) \cdot \tilde{H}\psi_n(s)\,\mathrm{d}s$$

for all $n \in \mathbb{N}$ and therefore also

$$\begin{aligned} |\langle \chi, \psi_n(t') - \psi_n(t)\rangle| &\leq \|(G(t'-t)^* - I)\chi\|\|\psi_n(t)\| + \|\tilde{H}\| \int_t^{t'} \|E_n(s)\|_{\mathbb{R}^L}\|\psi_n(s)\|\mathrm{d}s \\ &\leq \|(G(t'-t)^* - I)\chi\| + (t'-t)^{\frac{1}{q}}\|\tilde{H}\|\|E_n(s)\|_{L^p(0,T;\mathbb{R}^L)} \end{aligned}$$

for q with $\frac{1}{p} + \frac{1}{q} = 1$. Since G is continuous in the strong topology and $(\|E_n\|_{L^p})_n$ is bounded as a weak(-$*$) convergent sequence, we can choose t such that

$$|\langle \chi, \psi_n(t') - \psi_n(t)\rangle| \leq \epsilon.$$

for all $n \in \mathbb{N}$. The term in the middle on the right-hand side in (2.19) becomes smaller than ϵ for $n, m \geq N$ for some $N \in \mathbb{N}$ since $t \in S$. Therefore we have

$$|\langle \chi, \psi_n(t') - \psi_m(t) \rangle| < 3\epsilon$$

for all $n, m \geq N$, which proves that $(\langle \chi, \psi_n(t') \rangle)_n$ is a Cauchy sequence. Therefore the sequence converges for each $\chi \in \mathcal{H}$. Weak convergence of $(\psi_n(t'))_n$ now follows from the following general argument. Let $(\chi_\iota)_\iota$ be a orthonormal basis of the Hilbert space $\mathcal{H}$ and set $a_\iota^n := \langle \chi_\iota, \psi_n(t') \rangle$ and $a_\iota := \lim_n a_\iota^n$. Since $(\psi_n(t'))_n$ is bounded, every subsequence contains another weakly convergent subsequence. The limit is always given by $\sum_\iota a_\iota \chi_\iota$. Therefore the original subsequence already converged weakly. Thus for all $t \in [0, T]$, the sequence $(\psi_n(t))_n$ has a subsequence such that

$$\psi_n(t) \rightharpoonup \psi(t) \quad \text{in } \mathcal{H} \tag{2.20}$$

for some $\psi \in L^\infty(0, T; \mathcal{H})$.

Step 2 We will now show that ψ is the mild solution for the field E. Let $\chi \in \mathcal{H}$. We know

$$\langle \chi, \psi_n(t) \rangle_{\mathcal{H}} = \langle \chi, G(t)\psi_0 \rangle_{\mathcal{H}} + \sum_{l=1}^{L} \langle \chi, \int_0^t G(t-s) E_n(s)_l (-iH_l) \psi_n(s) \, \mathrm{d}s \rangle_{\mathcal{H}} \tag{2.21}$$

for all $n \in \mathbb{N}$ and $t \in [0, T]$ and will pass to the limit $n \to \infty$ to obtain

$$\langle \chi, \psi(t) \rangle_{\mathcal{H}} = \langle \chi, G(t)\psi_0 \rangle_{\mathcal{H}} + \sum_{l=1}^{L} \langle \chi, \int_0^t G(t-s) E(s)_l (-iH_l) \psi(s) \, \mathrm{d}s \rangle_{\mathcal{H}}. \tag{2.22}$$

For the term on the left-hand side of (2.21) we use (2.20). For each summand of the second term on the right we define $\tilde{\chi}_l(s) := (-iH_l)^* G(t-s)^* \chi$. Then

$$\langle \chi, \int_0^t G(t-s) E_n(s)_l (-iH_l) \psi_n(s) \, \mathrm{d}s \rangle_{\mathcal{H}} = \int_0^t E_n(s)_l \langle \tilde{\chi}_l(s), \psi_n(s) \rangle_{\mathcal{H}} \, \mathrm{d}s.$$

Since $(E_n)_l \xrightarrow{(*)} E_l$ in $L^p(0, T)$, it remains to show strong convergence $\langle \tilde{\chi}_l, \psi_n \rangle_{\mathcal{H}} \to \langle \tilde{\chi}_l, \psi \rangle_{\mathcal{H}}$ in $L^q(0, t)$. We have

$$\| \langle \tilde{\chi}_l, \psi_n \rangle - \langle \tilde{\chi}_l, \psi \rangle \|_{L^q(0,t)}^q = \int_0^t |\langle \tilde{\chi}_l(s), \psi_n(s) - \psi(s) \rangle_{\mathcal{H}}|^q \, \mathrm{d}s.$$

By (2.20) the integrand converges to 0 for each $s \in [0,t]$. Since $H_l \in \mathcal{B}(\mathcal{H})$ we obtain $\tilde{\chi}_l \in L^\infty(0,T;\mathcal{H})$. Therefore Lebesgue's dominated convergence theorem implies

$$\|\langle \tilde{\chi}_l, \psi_n\rangle - \langle \tilde{\chi}_l, \psi\rangle\|_{L^q(0,t)} \to 0.$$

Thus (2.22) holds. Since χ was arbitrary, (2.22) implies

$$\psi(t) = G(t)\psi_0 + \sum_{l=1}^{L} \int_0^t G(t-s)E(s)_l(-iH_l)\psi(s)\,\mathrm{d}t$$

for all $t \in [0,T]$, i.e. ψ is a mild solution of (2.1) for the field $E \in L^p(0,T;\mathbb{R}^L)$.

Step 3 Since ψ is uniquely defined as mild solution for E, we obtain pointwise weak convergence of the original sequence $(\psi_n)_n$. Since $\psi_n(t) \rightharpoonup \psi(t)$ in $\mathcal{H}$ and $1 = \|\psi_n(t)\| \to \|\psi(t)\| = 1$, we obtain strong convergence $\psi_n(t) \to \psi(t)$ in $\mathcal{H}$ for all $t \in [0,T]$. □

Remark. In a more general setting where $\mathfrak{G}$ is not unitary one still obtains $\psi_n(t) \rightharpoonup \psi(t)$ for all $t \in [0,T]$ in Proposition 6.

Using the boundedness of ψ, we immediately obtain the following corollary by the dominated convergence theorem for Bochner integrals [AB06, p. 11.45]. We will not use it in this thesis but it is useful in the analysis of other types of cost functionals like time averaged expectation values.

Corollary 7. *In the setting of Proposition 6, we have strong convergence $\psi_n \to \psi$ in $L^2(0,T;\mathcal{H})$.*

3 Sparse Time-Frequency Control of Quantum Systems

In this chapter a new optimal control framework for control with simple time-frequency structure is formulated. We start the chapter with an introduction to the optimal control of quantum systems. We motivate the standard optimal control approach and then discuss how control fields are typically interpreted. There we will see the shortcomings of the standard approach. Then we will present the new control framework. It is based on two key ideas. The first is to control not the field itself but a time-frequency representation of it. The second is to use cost functionals that promote sparsity in frequency direction and smoothness in time. We will give a rigorous definition of the proposed optimal control formulation and prove existence of optimal solutions. The general framework is then illustrated with several concrete examples. We close the chapter with a short overview of the literature on sparsity in quantum control and put our work into context of this previous work. The results of this section are contained in [FHK15].

3.1 Optimal Control of Quantum Systems

In this section we give an overview of a typical control approach in the context of quantum systems. First we discuss how the quantum control problem can be modeled as an optimal control problem. Then we discuss how to interpret typical control fields using representations in time, frequency and time-frequency.

3.1.1 Modeling Quantum Control Problems as Optimal Control Problems

A typical goal in molecular quantum control is to steer a system from an initial state ψ_0 into a subspace X, which consists of states with a desired property. Interesting choices for X include subspaces of common electronic structure or localization in space. The corresponding control problem can be formulated as follows: Find a control field E in a set of admissible control fields $\mathcal{E}$ such that the corresponding solution of (2.1)

satisfies $\psi(T) \in X$ for some $T > 0$. A nice introduction to this problem from a mathematical perspective is given in [DAl08]. A recent overview from the point of view of theoretical chemistry is given in [BCR10]. The bilinear structure of this control problem allows for a rich mathematical theory. Unfortunately, in general this problem does not have a solution by a fundamental negative result on controllability in infinite dimensions [BMS82]. Even in cases where controllability results tell us that suitable control fields exist [HTC83; TR03; Bos+12], the shape of the fields can often only be derived analytically for very simple systems [Bos+02]. On large time horizons, one can use asymptotic results as mentioned in Section 2.1.2. For shorter time horizons, the problem is more complicated. One therefore often reformulates the control problem as an optimal control problem. Overview articles on optimal quantum control from the point of view of theoretical physics and chemistry are given in [WG07; BZB08; Hof+12]. Other control approaches include open and closed loop feedback control [IK09; QG10].

The optimal control problem can be derived as follows. As a first step, we fix the final time $T > 0$ and consider the minimization of $\operatorname{dist}(\psi(T), X)$, where ψ and E solve (2.1). This can equivalently be written as

$$\text{Minimize}_{\psi,E} \quad \frac{1}{2}\langle \psi(T), \mathcal{O}\psi(T)\rangle \quad \text{s. t. (2.1)}, \tag{3.1}$$

where $\mathcal{O} = I - P_X$ and P_X is the orthogonal projection on X. This problem is not well-posed since, by the same argument as above, solutions might not exist. For example the infimum might be zero but the value zero might not be attained. There are several ways to guarantee the existence of solutions to problem (3.1). Among them are constraints on the admissible controls $\mathcal{E}$ like thresholds for the amplitude or restrictions to finitely many degrees of freedom. Most often one instead regularizes the problem with a Tikhonov term. This means that one adds a term to the cost functional that penalizes the growth of the control in a suitable way. Then the problem reads

$$\text{Minimize}_{\psi,E} \quad \frac{1}{2}\langle \psi(T), \mathcal{O}\psi(T)\rangle + \frac{\alpha}{2}\|E\|_{\mathcal{E}}^2 \quad \text{s. t. (2.1)}. \tag{3.2}$$

Here $\mathcal{E}$ often — but not necessarily [Bal+05] — is a Hilbert space contained in $L^1(0, T; \mathbb{R}^L)$, and $\alpha > 0$ is a regularization or cost parameter. The norm in $\mathcal{E}$ might penalize unfavorable behavior of the control field E. Problem (3.2) has solutions under appropriate assumptions [IK07; WBV10].

The operator $\mathcal{O}$ can also represent a more general observable. Then $\langle \psi(T), \mathcal{O}\psi(T)\rangle$ is the quantum mechanical expectation value of the observable for the state $\psi(T)$. The

observable $\mathcal{O}$ and a time horizon T are typically given by the concrete application. But there is some freedom in the choice of the space $\mathcal{E}$ and the regularization parameter α. Restrictions on the control field through the choice of $\mathcal{E}$ might be of physical nature or address issues of the experimental implementation. For comparisons with our new framework, we will consider the two cases $\mathcal{E} = L^2$ and $\mathcal{E} = H_0^1$, in later parts of this thesis referred to as *Hilbert space case*. The regularization with an L^2 cost term goes back to the first papers on optimal control for quantum molecular systems [PDR88]. It is used for its simplicity and seeming naturalness. The term $\|E\|_{L^2}^2$ is proportional to the energy of the field. Its boundedness is a physical necessity. However, optimal pulses computed for this choice of the cost term seem to suffer from a highly irregular shape and from oscillations which prohibit direct implementations in experiments. The H_0^1 regularization does not suffer from these oscillations since it penalizes them. Additionally, controls in H_0^1 satisfy homogeneous Dirichlet boundary conditions that are suitable for a laser pulse that gets switched on and off. A modified H_0^1 norm is related to the physical work done in the system [Hin+13]. Cost terms including a H_0^1 norm are successfully applied in applications [Hoh+07]. However, for the problem at hand, it introduces an undesired behavior of the control in the sense that high frequencies are penalized even though they might be useful to induce particular transitions, see Section 2.1.2. In the next section, we will take a closer look at how control fields are interpreted with regard to their time and frequency structure.

3.1.2 Representation and Interpretation of Control Fields

In quantum control the structure of control fields is important for the interpretation of control mechanism and the experimental realization of the fields. The structure of the control fields is often analyzed by using three different representations of the field: time representations, frequency representations and time-frequency representations. A more detailed interpretation of time and frequency structure of fields is given [Fec+07; MWC10] and in the context of quantum control in [Rue+11]. A general introduction to time-frequency analysis is given in [Grö01].

In Figure 3.1, two different fields are plotted in those three representations. For the time frequency representation we use a short-time Fourier transform. For the frequency and time-frequency representation, we only plotted the absolute values on a logarithmic scale. The first field has a relatively easy structure. From the time representation, we see that it consists of two *pulses* at different times with different frequencies. The duration of the pulses and the time gap between them can be derived from this representation. The time structure can often be translated into a sequence of transitions in the quantum

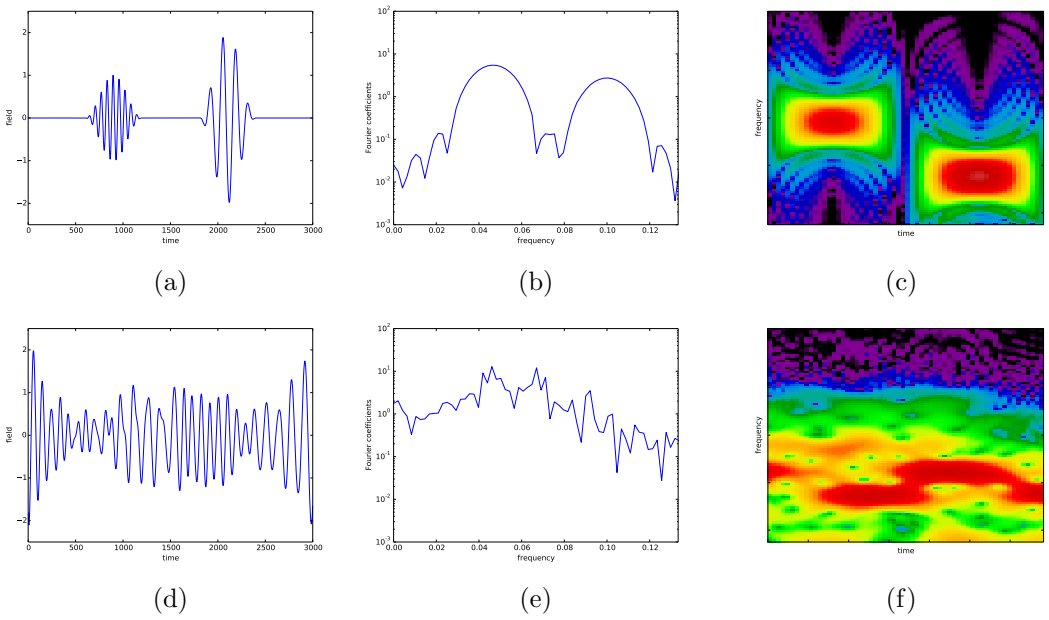


Figure 3.1: Two different fields (top and bottom) given in a time (left), frequency (middle) and time-frequency (right) representation.

system, see Section 2.1.2. The frequencies, which are important to understand which transition happens, are not so easily quantified from the time representation. The frequencies can be clearly seen in the frequency representation of the pulse. As expected there are two important frequencies corresponding to the two pulses. The additional structure in the neighborhood of those two frequencies comes from the time structure in the pulses. This time structure cannot so easily be derived from the frequency representation. Time-frequency representations combine the advantages of time and frequency representations at the cost of uncertainty in time and frequency. Here we can clearly see that the field consists of two pulses. For this control field, one would expect the field to introduce two transitions in the quantum system, first a transition with the higher frequency and then, after some quantifiable time lag, a transition with the lower frequency.

For the second control field, which actually originates from an optimal control computation, the analysis gets more complicated. In the time representation, one can see two main frequencies which occur in an alternating fashion. But there are non-trivial effects where the frequency contributions overlap. In the frequency representation, two frequencies stand out, but again the time structure is not easily reconstructed. For this complicated pulse, the time-frequency representation is most useful for the interpreta-

tion. We can see four main contributions, two for each frequency, with an alternating time structure. But one also sees a lot of additional structure in the time-frequency representation. For example, there are also substantial low frequency components contained in the field. It is not clear how they contribute to the field and how they should be interpreted. Some of this structure is of course also due to artifacts inherent in time-frequency transformation. In the next section, we propose a control framework that generates controls with a much simpler time-frequency structure.

3.2 A Framework for Sparse Time-Frequency Control

In this section we will present a new control framework for the generation of controls with simple time-frequency structure. It is the main contribution of this thesis. After a motivation we will give a rigorous definition of our new framework and prove the existence of solutions. Then several concrete examples for the general framework are given and the framework is put into context of existing literature.

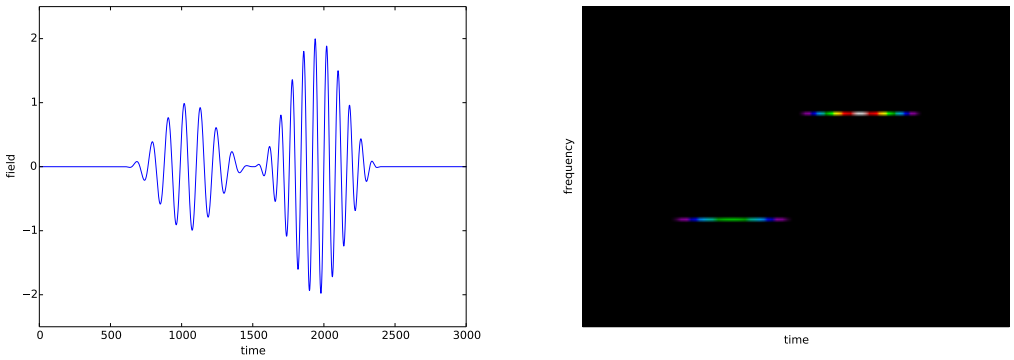
3.2.1 Optimal Quantum Control with Function-Valued Measures

The new framework is based on optimal control with function-valued measures in combination with special control operators accounting for the nontrivial physics of quantum control. The key ideas behind the framework are

- the control of a time-frequency representation and not the field itself and
- a penalization of the time-frequency representation to obtain sparse frequency structure and smooth time structure.

Here the use of time-frequency representations is different from what is done, in the context of parabolic problems, in the control theory literature, and the use of a sparsity-promoting penalization is different from what is done in the physics literature. In the following we give a motivation for our approach.

An important drawback of previous optimal control approaches is the complicated structure of controls in a time-frequency representation. This structure can be simplified by using a suitable penalization. But what would an ideal time-frequency representation look like? One possible idea is given in Figure 3.2b. This corresponds to a perfect frequency resolution with just two frequencies being present. Each of those frequencies then is modulated by a smooth envelope. If we stay in a setting of one-to-one time-frequency representations, this ideal case is not possible. One always has to account for uncertainty in frequency and time: A good frequency resolution implies a bad time resolution



(a) (b)

Figure 3.2: A field (left) and a corresponding idealized time-frequency representation (right).

and vice versa. The first key idea behind our approach is to avoid this uncertainty by considering a time-frequency object as the control and by mapping it to a control field with a control operator that is not one-to-one. In a sense, we over-parametrize the control field. For example the control u could be a function defined on the time-frequency plane $\mathbb{R}^+ \times [0, T]$ and the control operator B mapping u to a field could have the form

$$(Bu)(t) = \mathrm{Re} \int_{\mathbb{R}^+} u(\omega, t) e^{i\omega t} \, \mathrm{d}\omega.$$

The next question is how to force the control u to have a simple structure as in Figure 3.2b. Here we use a recently developed tool from optimal control, namely directional or joint sparsity [FR08; HSW12; KW13; KPV14]. Directional sparsity means that one has a function of several variables and has sparsity just in one direction. This is exactly what we want here: We want our time-frequency control to be sparse in the frequency direction but not sparse in the time direction. This behavior is typically obtained by penalizing in nested L^p spaces like

$$L^1(\mathbb{R}^+; L^2(0, T)).$$

The work by Kunisch, Pieper and Vexler, which uses measure spaces instead of L^1 spaces, seems closest to what we need in our application. With measures one can realize the idealized time-frequency structure from Figure 3.2b by using sums of Dirac measures in frequency multiplied by envelopes in time. Compared to previous uses of directional sparsity, we will need to adapt the setting to also obtain the desired smooth behavior in

the time direction. A natural choice for the control space then is

$$\mathcal{M}(\Omega; H_0^1(0,T)),$$

a space of function-valued measures defined on the frequency domain Ω with valued in the function space $H_0^1(0,T)$ of weakly differentiable functions in time. Having homogeneous Dirichlet boundary conditions is appropriate in the setting of a control field that is switched on and off.

The setting described in this section fits in a more general control framework. In the next section, we will give a rigorous definition of this framework.

3.2.2 An Abstract Control Framework

In this section we will present an abstract control framework for the sparse control of quantum systems. It is a suitable generalization of the motivating example given in the preceding section. We will formulate the optimal control problem and state explicit assumptions on the involved mathematical objects. Examples will be provided in Section 3.2.3.

The optimal control problem we solve will be of the form

$$\begin{aligned} \text{Minimize}_{\psi,u} \quad & \frac{1}{2}\langle \psi(T), \mathcal{O}\psi(T)\rangle_{\mathcal{H}} + \alpha\|u\|_{\mathcal{M}(\Omega;\mathcal{U})} \\ \text{s.t. } & i\partial_t\psi = (H_0 + Bu \cdot H_1)\psi, \quad \psi(0) = \psi_0. \end{aligned} \tag{3.3}$$

Here, we minimize the *cost functional*

$$J(\psi, u) = \frac{1}{2}\langle \psi(T), \mathcal{O}\psi(T)\rangle_{\mathcal{H}} + \alpha\|u\|_{\mathcal{M}(\Omega;\mathcal{U})}$$

consisting of the scaled *expectation value* of the *observable* $\mathcal{O}$ at the *final time* $T > 0$ and the scaled measure norm of the *control* u in the space $\mathcal{M}(\Omega;\mathcal{U})$ of bounded regular vector measures defined on the *sparsity domain* Ω with values in the space $\mathcal{U}$. The second summand is weighted by a *cost parameter* $\alpha > 0$. The *control operator* B maps the control u to a *control field* Bu, which enters in the equality constraint. The equality constrained is to be understood in the mild sense as discussed in Chapter 2. We will now present the precise assumptions for the general control framework.

For the measure space, we assume the following.

(A3) The sparsity domain Ω is a locally compact subset of $\mathbb{R}^n$ for some $n \in \mathbb{N}$.

(A4) The space $\mathcal{U}$ is a separable Hilbert space.

Assumption (A3) on the sparsity domain allows for sufficient flexibility to consider discrete sets like $\mathbb{N}$ or a continuum like $\mathbb{R}$. In fact we allow for those sets Ω that can be written as a set theoretic difference of two closed sets. The restriction to subsets of $\mathbb{R}^N$ is probably not necessary and the theory should also hold for general locally compact spaces. Assumption (A4) gives us the flexibility to consider different function spaces, but also finite dimensional spaces or just the real numbers. The choice of this space has a big influence on the structure of optimal controls, as we will see in Chapters 4 and 7. We will formulate the control framework in the space $\mathcal{M}(\Omega;\mathcal{U})$ of bounded regular vector measures. An introduction to vector measures can be found in [Lan93] and an exhaustive treatment is given in [DU77]. For most parts of this thesis it is not strictly necessary to be familiar with vector measures. This is because we can often use the decomposition of a vector measure $u \in \mathcal{M}(\Omega;\mathcal{U})$ into a positive measure $|u| \in \mathcal{M}(\Omega)$ and a direction $u' \in L^1(\Omega, |u|;\mathcal{U})$. Here, $|u|$ is the total variation measure of u, see [Lan93, p. VII 3.1], and u' is the Radon–Nikodym derivative of u with respect to $|u|$, see [Lan93, pp. VII 4.1,4.2]. Thus, instead of working with the vector measure, one works with an ordinary measure and a vector-valued function. The total variation norm of u in $\mathcal{M}(\Omega;\mathcal{U})$ is given by

$$\|u\|_{\mathcal{M}(\Omega;\mathcal{U})} = \int_\Omega \mathrm{d}|u|(\omega).$$

Under assumptions (A3) and (A4) we have the duality

$$C_0(\Omega;\mathcal{U})^* = \mathcal{M}(\Omega;\mathcal{U}). \tag{3.4}$$

A proof of this result, known as Singer's theorem [Zin57], is given in [Mez09]. It can also be derived from the original result for compact Ω, see the comment in [Cam76, p. 2]. The duality pairing is given by

$$\langle \varphi, u \rangle_{\mathcal{C}_0,\mathcal{M}} = \int_\Omega \langle \varphi(\omega), u'(\omega) \rangle_{\mathcal{U}} \, \mathrm{d}|u|(\omega) \tag{3.5}$$

for $u \in \mathcal{M}(\Omega;\mathcal{U})$ and $\varphi \in C_0(\Omega;\mathcal{U})$. We identify the dual $\mathcal{U}^*$ with $\mathcal{U}$ by the Fréchet–Riesz representation theorem. This is a natural identification in this abstract framework. Choosing non-Hilbert spaces for $\mathcal{U}$ is also interesting, but for simplicity we restrict ourself to the simpler Hilbert space case. The duality (3.4) and the separability in assumption (A4) make $\mathcal{M}(\Omega;\mathcal{U})$ the dual of a separable space [Bou65, Prop. 1]. This makes the sequential Banach–Alaoglu theorem applicable [AB06, pp. 6.25, 6.34, 3.17]. This will become important for showing the existence of solutions to problem (3.3). As in Chapter 2 we will equip $\mathcal{U}$ with a real Hilbert space structure.

For the control operator, we will assume the following.

(A5) The control operator B satisfies $B \in \mathcal{B}(\mathcal{M}(\Omega;\mathcal{U}), L^p(0,T;\mathbb{R}^L))$ for some $1 < p \leq \infty$ and has a predual operator.

Here, the natural number L has to fit the description of the equation. The condition $p > 1$ is used to make the compactness result in Proposition 6 applicable. We say that a bounded linear operator B has a predual operator A if $A^* = B$ in the sense of operators between Banach spaces. In our case this means that

$$A\colon L^q(0,T;\mathbb{R}^L) \to C_0(\Omega;\mathcal{U}),$$

where $\frac{1}{p} + \frac{1}{q} = 1$ and

$$\langle Af, u\rangle_{C_0,\mathcal{M}} = \langle f, Bu\rangle_{L^q,L^p}$$

for all $f \in L^q(0,T;\mathbb{R}^L)$ and $u \in \mathcal{M}(\Omega;\mathcal{U})$. Here we already used the duality (3.4). For non-reflexive spaces we do not necessarily have $B^* = A$. Instead B^* is an extension of A. Having a predual operator often corresponds to additional regularity of the operator B^*. More specifically, it implies weak-$*$–weak(-$*$) continuity of B and in particular $u_n \overset{*}{\rightharpoonup} u$ in $\mathcal{M}(\Omega;\mathcal{U})$ implies $Bu_n \overset{(*)}{\rightharpoonup} Bu$ in $L^p(0,T;\mathbb{R}^L)$. The existence of a predual operator will be used in the existence proof and to provide optimality conditions. In slight abuse of notation, we will use B^* as a notation for the predual operator. Since B^* appears in the optimality system its structure is of importance. In the examples we will see the influence of the choice of B and $\mathcal{U}$ on B^*.

The observable $\mathcal{O}$ has the following regularity.

(A6) The observable $\mathcal{O}$ is a self-adjoint bounded linear operator on $\mathcal{H}$.

Using bounded observables fits naturally in our equation setting. Other authors use stricter assumptions on the observable like $\mathcal{O} \in B(\mathcal{H}) \cap B(\mathcal{V})$ for the space of higher regularity $\mathcal{V}$ discussed as in Section 2.2. This assumption on higher regularity is need for their more restrictive solution theory since one needs to study solutions of the adjoint equation with terminal condition given by $\mathcal{O}\chi$ for some $\chi \in \mathcal{H}$. We do not need this higher regularity here. Using observables with even less regularity, like unbounded operators, is of physical interest and is, for example, done in [Hin+13]. Minimizing a functional containing the expectation value $\langle \psi(T), \mathcal{O}\psi(T)\rangle_{\mathcal{H}}$ results in final states $\psi(T)$ close to the eigenspace corresponding to the smallest eigenvalue of $\mathcal{O}$. Often $\mathcal{O}$ is an orthogonal projection on a subspace we want to leave. Then $\frac{1}{2}\langle \psi(T), \mathcal{O}\psi(T)\rangle_{\mathcal{H}} \in [0, 0.5]$ and we can use a probabilistic interpretation of the scaled expectation value. The value

0 correspond to a 100% and the value 0.5 corresponds to 0% chance of achieving the control objective to leave the subspace. Instead of expectation values at the final time one could also use a cost distributed in time [KHK10]. To allow for the control into a subspace of a prescribed eigenvalue, and not just the lowest one, one can use squared expectation values [Hin+13].

Note, that although equation (2.1) is linear in the state and linear in the control, the resulting bilinear control problem is not convex. In addition, the measure norm is convex but not differentiable. Therefore (3.3) defines a nonconvex and nonsmooth optimization problem.

Instead of the constrained optimization problem (3.3), we can equivalently study the unconstrained problem

$$\text{Minimize} \quad j(u), \quad u \in \mathcal{M}(\Omega;\mathcal{U}) \tag{3.6}$$

of the *reduced cost functional*

$$j(u) = J(\psi(u), u) = \frac{1}{2}\langle \psi(u)(T), \mathcal{O}\psi(u)(T)\rangle_{\mathcal{H}} + \alpha\|u\|_{\mathcal{M}(\Omega;\mathcal{U})},$$

where $\psi(u)$ denotes the solution of (2.1) for a given control u. The functional j is called reduced cost functional since the explicit dependence of the cost functional on ψ is dropped. Unconstrained problems are sometimes easier to study from a theoretical point of view. We will use the reduced approach to show existence of solutions and to prove optimality conditions. From the numerical point of view, it also reduces the degrees of freedoms of the problem and it is possible to directly apply techniques from unconstrained optimization theory. We also use the reduced approach in the numerics.

We will now prove the existence of solutions to problem (3.3). After the preparations in Chapter 2 and defining a suitable rigorous framework in this section, we can follow a standard proof pattern: the direct method in the calculus of variation.

Theorem 8. *There exists a solution* $(\bar{\psi}, \bar{u}) \in C([0,T];\mathcal{H}) \times \mathcal{M}(\Omega;\mathcal{U})$ *of* (3.3).

Proof. We will use the reduced formulation (3.6). By Lemma 3 and (A6) the reduced functional j is bounded from below. Hence, there exists a minimizing sequence $(u_n)_n$ satisfying

$$\lim_n j(u_n) = \inf_u j(u). \tag{3.7}$$

The sequence of minimizing controls $(u_n)_n$ is bounded in $\mathcal{M}(\Omega;\mathcal{U})$ as a consequence of $\alpha > 0$. Because of the duality (3.4), $\mathcal{M}(\Omega;\mathcal{U})$ is the dual of a separable Banach space. Therefore, by the sequential Banach–Alaoglu theorem, the sequence $(u_n)_n$ has a weak-$*$ convergent subsequence still denoted by $(u_n)_n$ with limit $\bar{u} \in \mathcal{M}(\Omega;\mathcal{U})$. The weak-$*$–

weak(-$*$) continuity of B implies $Bu_n \overset{(*)}{\rightharpoonup} B\bar{u}$ in $L^p(0,T;\mathbb{R}^L)$. By Proposition 6, the corresponding sequence of states $(\psi_n)_n$ satisfies $\psi_n(T) \to \bar{\psi}(T)$. Thus, the first summand of j converges, i.e. $\langle \psi_n(T), \mathcal{O}\psi_n(T)\rangle \to \langle \bar{\psi}(T), \mathcal{O}\bar{\psi}\rangle$. The second summand of j is weak-$*$ lower semi-continuous as it is a norm in a dual space [AB06, p. 6.26]). Thus, we obtain $\lim_n j(u_n) \geq j(\bar{u})$. Together with (3.7) this implies the claim. □

3.2.3 Examples

In this section we will present several examples for the abstract framework presented in the preceding section. We discuss different choices for the sparsity domain Ω, the space $\mathcal{U}$ and the control operator B. In the examples we will always use $L = 1$. Generalizations of the examples to $L > 1$ are often immediate and depend on the equation studied.

The first example was the motivation for the general functional analytic framework. In this example u will represent a time-frequency representation and B will generate the corresponding field. The control u consists of different frequencies with smooth envelopes. This can be modeled as follows.

Example 3. Let $\Omega \subset \mathbb{R}^+$ be closed, $\mathcal{U} = H_0^1(0,T;\mathbb{C})$, $p = \infty$ and $L = 1$. We define B by

$$(Bu)(t) = \operatorname{Re} \int_\Omega u'(\omega,t) e^{i\omega t} \,\mathrm{d}|u|(\omega), \tag{3.8}$$

Here u' is the Radon–Nikodym derivative of u with respect to $|u|$ as in (3.5).

Let us check the validity of assumptions (A3-A5). A closed subset $\Omega \subset \mathbb{R}^+$ is locally compact, which shows that (A3) holds. The space $H_0^1(0,T;\mathbb{C})$ is a separable Hilbert space, thus (A4) holds. Here, it is equipped with the scalar product

$$\langle v, w\rangle_{\mathcal{U}} = \operatorname{Re} \int_0^T \overline{\partial_t v(t)} \partial_t w(t) \,\mathrm{d}t,$$

where bar denotes complex conjugation. To show that B is a bounded linear operator, we use a duality argument. Define the operator $A\colon L^1(0,T) \to C_0(\Omega;\mathcal{U})$ for $f \in L^1(0,T)$ and $\omega \in \Omega$ by the weak solution of the Poisson equation

$$-\Delta_t (Af)(\omega) = f e^{-i\omega\cdot}, \tag{3.9}$$

with homogeneous Dirichlet boundary conditions, that is $(Af)(\omega) \in \mathcal{U}$ such that

$$\langle (Af)(\omega), \varphi\rangle_{\mathcal{U}} = \operatorname{Re} \int_0^T \overline{f(t) e^{-i\omega t}} \varphi(t) \,\mathrm{d}t$$

for all $\varphi \in \mathcal{U}$. Indeed, unique solutions exist with

$$\|(Af)(\omega)\|_{\mathcal{U}} \leq \|fe^{-i\omega\cdot}\|_{H^{-1}(0,T;\mathbb{C})}$$

and the solutions depend continuously on the right hand side. This gives $Af \in C(\Omega;\mathcal{U})$. If Ω is bounded, this implies $Af \in C_0(\Omega;\mathcal{U})$ since Ω is closed and therefore compact. If Ω is unbounded, we have $fe^{-i\omega\cdot} \rightharpoonup 0$ in $L^1(0,T;\mathbb{C})$ for $\omega \to \infty$. Together with the compact embedding of $L^1(0,T;\mathbb{C})$ into $H^{-1}(0,T;\mathbb{C})$ in one dimension, we obtain $(Af)(\omega) \to 0$ for $\omega \to \infty$. Thus, A is a well-defined bounded linear operator. Hence, its dual operator $A^*\colon \mathcal{M}(\Omega;\mathcal{U}) \to L^\infty(0,T)$ is bounded and linear. Since, using Fubini's theorem, we have

$$\begin{aligned}
\langle Af, u\rangle_{C_0(\Omega;\mathcal{U}),\mathcal{M}(\Omega;\mathcal{U})} &= \int_\Omega \langle (Af)(\omega), u'(\omega)\rangle_{\mathcal{U}} \,\mathrm{d}|u|(\omega) \\
&= \int_\Omega \operatorname{Re} \int_0^T f(t) e^{i\omega t} u'(\omega, t) \,\mathrm{d}t \,\mathrm{d}|u|(\omega) \\
&= \int_0^T f(t) \operatorname{Re} \int_\Omega u'(\omega, t) e^{i\omega t} \,\mathrm{d}|u|(\omega) \,\mathrm{d}t \\
&= \langle f, Bu\rangle_{L^1(0,T),L^\infty(0,T)}
\end{aligned}$$

for all $f \in L^1(0,T)$ and $u \in \mathcal{M}(\Omega;\mathcal{U})$, we obtain $A^* = B$. Therefore $B\colon \mathcal{M}(\Omega,\mathcal{U}) \to L^\infty(0,T)$ is a bounded linear operator and has a predual operator.

Typical choices for frequency domain Ω include the whole positive real line $\mathbb{R}^+$, a closed interval or even a discrete set of admissible frequencies. Taking the real part of the integral in the definition of the control operator (3.8) could be avoided by using $\Omega \subset \mathbb{R}$ with $-\Omega = \Omega$ and requiring u to satisfy $u(-X) = \overline{u(X)}$.

For controls which are sums of Dirac measures, $u = \sum_k c_k \delta_{\omega_k}$, where $\omega_k \in \Omega$ and $c_k \in H_0^1(0,T;\mathbb{C})$, we obtain

$$(Bu)(t) = \sum_k \operatorname{Re} c_k(t) e^{i\omega_k t}.$$

We allow $\mathcal{U}$ to contain complex-valued functions. This allows for a shift in the different frequencies and their phases without leaving the linear setting. If c_k is decomposed as $c_k(t) = |c_k(t)|e^{i\theta_k(t)}$, then

$$(Bu)(t) = \sum_k |c_k(t)| \cos(\omega_k t + \theta_k(t)).$$

If the function θ_k is constant, the phase of the frequency ω_k is shifted by θ_k. If the

function θ_k is linear, the frequency ω_k itself is shifted by θ_k' in addition to the shift in its phase. A general θ_k can have a complicated influence on the control field.

In this example the identification of $\mathcal{U}^*$ with $\mathcal{U}$ leads to an identification of $(H_0^1)^*$ with H_0^1 itself and not with H^{-1} as it is usually the case. I hope that this causes no confusion. One can of course also use the standard identification, but then one has to be careful about the definition of the predual operator and also with the optimality conditions as presented in the next chapter.

Equation (3.9) gives that for each ω the function $(B^*f)(\omega)$ satisfies a one-dimensional complex-valued differential equation. The equations are not coupled for different ω. It is important that Ω is closed. Otherwise the operator B^* would not satisfy the necessary homogeneous Dirichlet boundary conditions on Ω. This means that B would not have a predual operator but just a dual operator and would in particular not be weak-$*$–weak(-$*$) continuous. The operator B^* can also be written as

$$(B^*f)(\omega, t) = \int_0^T G(t,s) f(s) e^{-i\omega s} \,\mathrm{d}s,$$

where G is the Green's kernel for the one dimensional Laplace equation.

The control operator B is related to the Zak transform [Grö01, p. 8.], in particular, see the inversion formula (8.9) in this reference.

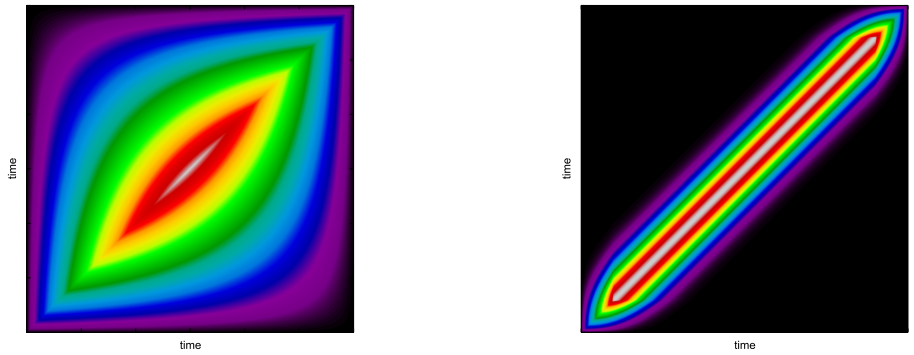
In the preceding example, the smoothness of the envelope functions comes from the choice $\mathcal{U} = H_0^1(0,T;\mathbb{C})$. The envelopes have to be functions with a certain smoothness. A different approach would be to put the smoothness of the envelopes into the control operator B, as is done in the following example.

Example 4. Let $\Omega \subset \mathbb{R}^+$ be closed, $\mathcal{U} = L^2(0,T;\mathbb{C})$, $p = \infty$ and $L = 1$. We define B for a smooth, non-negative and symmetric kernel $k\colon [0,T]^2 \to \mathbb{R}$ by

$$(Bu)(t) = \mathrm{Re} \int_\Omega \int_0^T k(t,s) u'(\omega, s) \,\mathrm{d}s \; e^{i\omega t} \,\mathrm{d}|u|(\omega). \tag{3.10}$$

The kernel k is often of the form $k(t,s) = \eta(t-s)$ for some smoothing kernel η. A typical example for η is a Gaussian kernel, $\eta(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{t^2}{2\sigma^2})$. The kernel k can also induce suitable boundary conditions like homogeneous Dirichlet boundary conditions. We will assume that k is such that the operator K defined by

$$(Kf)(t) = \int_0^T k(t,s) f(s) \,\mathrm{d}s \tag{3.11}$$



(a) (b)

Figure 3.3: Coefficients of different kernels. In (a) the Green's function of the one-dimensional Poisson equation, in (b) a modified Gaussian kernel.

is a compact self-adjoint operator on $L^2(0,T;\mathbb{C})$ and can be extended to a compact operator $K\colon L^1(0,T;\mathbb{C}) \to L^\infty(0,T;\mathbb{C})$. The control operator can also be written as

$$(Bu)(t) = \int_\Omega (Ku'(\omega))(t) e^{i\omega t} \,\mathrm{d}|u|(\omega).$$

In contrast to the control operator in Example 3, here we first smooth the envelopes and then modulate the frequencies with these smoothed functions. It follows that B maps to $L^\infty(0,T)$. For a Gaussian kernel we even obtain C^∞ regularity of the fields.

In this setting the predual operator $B^*\colon L^1(0,T) \to C_0(\Omega;\mathcal{U})$ has the form

$$(B^* f)(\omega,t) = \int_0^T k(t,s) f(s) e^{-i\omega s} \,\mathrm{d}s = K(f e^{-i\omega\cdot})(t). \tag{3.12}$$

For a Gaussian smoothing kernel, B^* has the form of a short-time Fourier transform [Grö01, p. 3.].

We can compare the structure of the predual operator B^* for Examples 3 and 4. Example 3 leads to a global time-frequency representation B^*. That means that it has a global window kernel equal to the Green's function G depicted in Figure 3.3a On the other hand, Example 4 with a modified Gaussian kernel leads to a nice time-frequency representation with a Gaussian window, see Figure 3.3b.

One can ask the question whether one can choose a better space $\mathcal{U}$ such that the natural control operator B as defined in (3.8) results in a predual B^* as in (3.12) with a useful kernel. This is possible, as we will see in the next example. In fact we will

see that the choice of $\mathcal{U}$ presented in the next example gives rise to an optimal control problem equivalent to the one from Example 4.

Example 5. Let Ω and B be as in Example 3 and let k be a kernel as in Example 4. Under suitable assumptions, which are satisfied for Gaussian kernels, the kernel k defines an injective positive compact operator K on $L^2(0,T;\mathbb{C})$ by (3.11). We use its inverse $A = K^{-1}$ to define the scalar product $\langle\cdot,\cdot\rangle_k := \langle A\cdot, A\cdot\rangle_{L^2}$ on the domain of definition $\mathcal{U}_k := \mathcal{D}(A)$ of A. Then the control operator B defined by (3.8) results in a predual operator given by

$$(B^* f)(\omega, t) = (K^2(fe^{-i\omega\cdot}))(t),$$

since

$$\begin{aligned}
\langle f, Bu\rangle_{L^1(0,T),L^\infty(0,T)} &= \int_0^T f(t) \operatorname{Re} \int_\Omega u'(\omega,t) e^{i\omega t} \,\mathrm{d}|u|(\omega)\,\mathrm{d}t \\
&= \int_\Omega \operatorname{Re} \int_0^T \overline{f(t)e^{-i\omega t}} u'(\omega,t)\,\mathrm{d}t\,\mathrm{d}|u|(\omega) \\
&= \int_\Omega \langle K^{-1}K^2(fe^{-i\omega\cdot}), K^{-1}u'(\omega)\rangle_{L^2(0,T;\mathbb{C})}\,\mathrm{d}|u|(\omega) \\
&= \int_\Omega \langle K^2(fe^{-i\omega\cdot}), u'(\omega)\rangle_{\mathcal{U}_k}\,\mathrm{d}|u|(\omega).
\end{aligned}$$

Next we will discuss the equivalence of this example to Example 4. We will show that there is an isomorphism between the control spaces such that the control field and the cost functional value are preserved. We denote by $\tilde{\mathcal{U}}$ and $\tilde{B}$ the space and operator from Example 4. Then K defines a isometric isomorphism from $\tilde{\mathcal{U}}$ to $\mathcal{U}_k$. We define the operator $X\colon C_0(\Omega;\tilde{\mathcal{U}}) \to C_0(\Omega;\mathcal{U}_k)$ by

$$(X\varphi)(\omega) = K\varphi(\omega).$$

Then X is an isometric isomorphism and so is its dual $X^*\colon \mathcal{M}(\Omega;\mathcal{U}_k) \to \mathcal{M}(\Omega;\tilde{\mathcal{U}})$, which implies

$$\|X^*u\|_{\mathcal{M}(\Omega;\tilde{\mathcal{U}})} = \|u\|_{\mathcal{M}(\Omega;\mathcal{U}k)} \tag{3.13}$$

for $u \in \mathcal{M}(\Omega;\mathcal{U}_k)$. Since for $\varphi \in C_0(\Omega;\tilde{\mathcal{U}})$ and $u \in \mathcal{M}(\Omega;\mathcal{U}_k)$, we have

$$\begin{aligned}
\langle X\varphi, u\rangle_{C_0(\Omega;\mathcal{U}_k),\mathcal{M}(\Omega;\mathcal{U}_k)} &= \int_\Omega \langle K(\varphi(\omega)), u'(\omega)\rangle_{\mathcal{U}_k}\,\mathrm{d}|u|(\omega) \\
&= \int_\Omega \langle \varphi(\omega), K^{-1}(u'(\omega))\rangle_{\tilde{\mathcal{U}}}\,\mathrm{d}|u|(\omega),
\end{aligned}$$

and due to $\|K^{-1}u'(\omega)\|_{\tilde{\mathcal{U}}} = \|u'(\omega)\|_{\mathcal{U}_k} = 1$, the dual X^* of X is given by

$$\mathrm{d}X^*u = K^{-1}u' \,\mathrm{d}|u|.$$

Therefore we obtain

$$\begin{aligned}(\tilde{B}X^*u)(t) &= \int_\Omega (K(K^{-1}u'(\omega)))(t)e^{i\omega t} \,\mathrm{d}|u|(\omega) \\ &= \int_\Omega u'(\omega,t)e^{i\omega t} \,\mathrm{d}|u|(\omega) \\ &= (Bu)(t).\end{aligned}$$

Same control fields result in the same state ψ and in the same expectation value, and the norm in the cost functional is preserved due to (3.13). Thus, we see that the values of the cost functionals for controls in $\mathcal{M}(\Omega;\mathcal{U}_k)$ and the corresponding functional and controls in $\mathcal{M}(\Omega;\tilde{\mathcal{U}})$ is the same. Therefore locally optimal controls are mapped to locally optimal controls with the same cost functional value.

Our functional analytic setting also covers interesting cases where $\mathcal{U}$ does not contain time dependent functions. This corresponds to what is already done in the literature, as will be explained in Section 3.2.4. The next example does not work with time-frequency controls but uses just frequency control.

Example 6. Let $\Omega \subset \mathbb{R}^+$ and let $\mathcal{U} = \mathbb{C}$. Then the control space $\mathcal{M}(\Omega;\mathbb{C})$ is the space of complex measures on Ω. We define the control operator as the inverse Fourier transformation for measures,

$$(Bu)(t) = \mathrm{Re}\int_\Omega u'(\omega)e^{i\omega t} \,\mathrm{d}|u|(\omega). \tag{3.14}$$

In the ideal case, optimal controls will be sums of few Dirac measures. The predual control operator will just be the Fourier transformation for functions restricted to $[0,T]$,

$$(B^*f)(\omega) = \int_0^T f(t)e^{-i\omega t} \,\mathrm{d}t.$$

For $\Omega = \mathbb{N}$ we obtain the interesting case of sparsity in the Fourier coefficients.

Working just with the frequency structure of the fields is what is often done in the literature. This will be explained in more detail in the next section.

The next example shows that Ω need not always be a frequency domain. There it will be a time-frequency domain where sparsity is desired in both the time and frequency

direction.

Example 7. Let $\Omega \subset \mathbb{R}^+ \times [0,T]$ be closed and let $\mathcal{U} = \mathbb{C}$. We define the control operator by

$$(Bu)(t) = \operatorname{Re} \int_\Omega u'(\omega, s) g_{\omega,s}(t) \,\mathrm{d}|u|(\omega, s) \tag{3.15}$$

for the functions

$$g_{\omega,s}(t) = k(t,s) e^{i\omega(t-s)}$$

with a Gaussian kernel k. With this control operator, each Dirac measure $\delta_{(\omega,t)}$ corresponds to a Gaussian wave packet located around time t with frequency ω. Using sums of Dirac measures one can generate control fields with a pulse structure. The predual control operator is then given by

$$(B^* f)(\omega, t) = \int_0^T \overline{g_{\omega,t}(s)} f(s) \,\mathrm{d}s.$$

For discrete sparsity domains Ω, we are in a setting of optimizing coefficients in front of given ansatz functions g. It is possible to use different ansatz functions. As in Example 4, the control operator and adjoint control operator are related to short-time Fourier transformations.

There are much more examples that fit to our framework. We just wanted to give an idea of how flexible the framework is. We will now put our framework into context with the existing literature.

3.2.4 Previous Approaches on Sparsity in Quantum Control

There are several papers on dealing with the complicated oscillating structure of controls in optimal quantum control. They can be broadly classified into two categories. One approach is to use modifications of the original L^2 optimal control techniques [Ren+06; CB06; Lap+09; Hoh+07; WB08; Hin+13; KHK10]. Another approach is to use low dimensional, often nonlinear, parametrizations of the control field [Dio+02; TLR04; SSB10; CCM11; Rue+11]. In this section we describe how those approaches are related to our new control framework.

Let us first discuss the approach that uses modified optimal control techniques. They often proceed as follows. First one sets up an iteration procedure to solve an optimal control problem that generates control fields with an undesirable frequency structure. Then they include an additional step in the procedure to make the frequency structure of the control field more desirable. Examples include a smoothed version of the restriction

to predefined admissible frequencies [CB06; Lap+09]. A different modification does not require predefined frequencies but just damps frequencies with small contributions [Ren+06]. In our framework, manipulating only the frequency structure of control fields corresponds to a setup like the one in Example 6. There, constraints on the frequencies can be obtained by modifying the sparsity domain Ω. Using minimization with measure norms, as generalization of an L^1 norm, leads to controls that only have few frequencies with large contributions. Most other frequencies have vanishing or negligibly small contributions. This leads to comparable results as the damping of small frequencies. This idea will be discussed further in Section 4.2.2. Our framework provides a more systematic approach to simplifying the frequency structure of control fields by formulating a suitable optimal control problem. This systematic approach makes fast optimization methods applicable. In numerical experiments we saw that sparsity in frequencies often does not lead to satisfactory controls since sparsity in frequency decreases time structure of the control fields. Our new framework allows to work with time-frequency representations where good localization in frequency does not destroy localization in time. This is a major advance compared to the literature.

Another optimal control approach that deal with oscillating controls uses a setup that penalizes fast oscillations. This is done by using H^1-type norms of the control field as costs in the cost functional [Hoh+07; WB08; Hin+13]. For the applications that we have in mind, this sometimes leads to undesirable controls. Penalizing fast oscillations makes transitions with large energy differences less favorable, which causes problems for quantum systems where those transitions are the desired control mechanisms. We will see this effect in the numerical experiments in Chapter 7.

The second approach, which is based on low dimensional parametrizations, typically starts from an expression of the control field of the form

$$E(t) = \sum_{n=1}^{N} a_n b_n(t) \cos(\omega_n t + \theta_n) \tag{3.16}$$

where $N \in \mathbb{N}$, $a_n \in \mathbb{R}$ are scaling coefficients, $b_n \colon [0,T] \to \mathbb{R}$ are envelope functions, ω_n are the frequencies and θ_n are phase shifts, see [Dio+02; TLR04; SSB10; Rue+11; CCM11]. The problem lies in finding suitable values for the parameters. For a fixed small N, and fixed frequencies ω_n and phases θ_n, one can try to find optimal b_n which are parametrized with a low number of degrees of freedom [SSB10]. One can also additionally optimize the phase shifts [Dio+02; TLR04]. For this highly nonlinear problem, derivative free methods are used, also see [CCM11]. Those methods have the advantage that they can be used directly in an experimental setup. Methods based on derivatives require

more knowledge about the controlled system since derivatives are not directly accessible in experiments. Disadvantages are the inherent nonlinearity, which makes them difficult to solve, and the heavy dependence on the chosen parametrization, which makes them inflexible. With respect to a parametrization of the form (3.16), our framework can be understood as follows. We make N large and allow for frequencies ω_n on a fine grid. As a limit one can think of the frequencies as a continuum. On the other hand, we force a to be sparse in the sense of having only few nonzero entries or, in the limiting case, to be a measure $a \in \mathcal{M}(\Omega)$ with small support. In our framework we can solve for b_n without restricting ourself to a small number of parameters. Instead, we look for b_n or, in the limit case, $b(\omega)$ in some space $\mathcal{U}$ of complex-valued functions. The complex phase of b then results in phase shifts. Normalizing b, we end up with an object $u \in \mathcal{M}(\Omega;\mathcal{U})$ with $u' = b$ and $|u| = a$. The parametrization (3.16) then corresponds to a control operator of the form (3.8). In our framework one can also restrict the frequencies to a prescribed small discrete set and only solve for the envelopes b_n. One possible disadvantage of our approach is that we cannot directly model the parameters of a realistic laser source because these parameters typically influence the field in a nonlinear way. This is a problem inherent in our current approach since it is a linear one. Adding realistic experimental constraints is interesting and relevant, but beyond the scope of this thesis. Another disadvantage is the often large number of degrees of freedom, which make derivative free methods difficult to apply.

We see that most previous approaches on sparsity in quantum control can be realized with our general framework. In addition, our framework provides a powerful generalizations to the case of time-frequency control.

4 Necessary Optimality Conditions

In this chapter we derive optimality conditions for optimal solutions to the problem proposed in the previous chapter. We first look at derivatives of expectation values in the context of mild solutions. Those results will be used to study necessary optimality conditions for our new framework. We will then briefly discuss optimality conditions for the case of discrete sparsity domains and for concrete realizations of our control framework.

4.1 Optimality Conditions and Mild Solutions

The first order optimality conditions we derive are based on the differential of the cost functional. In our case, where the cost functional consists of a smooth and a nonsmooth part, we will first study the smooth part. The smooth part is given by a quantum mechanical expectation value. Therefore we study the derivative of expectations values in the direction of the control field. Since the quantum system enters the cost functional only in the expectation value, this is where we need to deal with the solutions of differential equations. In particular we will see how it is possible to represent the derivatives in the setting of mild solutions. We will use those results to derive the optimality conditions for the Hilbert space case that is typically studied in the literature. In contrast to the literature, we will not follow a Lagrange functional approach but directly differentiate the reduced cost functional.

4.1.1 Differentiation of Expectation Values

In this section we will provide representations for derivatives of expectation values in the direction of the field. The representation can formally be derived using a Lagrange functional approach [PDR88]. Under suitable assumptions this approach can be made rigorous [WBV10]. But the Lagrange approach does not fit conveniently to the mild solution framework. Therefore, we will directly compute the derivative using the differentiability results from Section 2.2.2.

Let $f\colon L^1(0,T;\mathbb{R}^L) \to \mathbb{R}$ be defined as

$$f(E) = \frac{1}{2}\langle \psi(T), \mathcal{O}\psi(T)\rangle, \tag{4.1}$$

where ψ is the solution of (2.6) for the field $E \in L^1(0,T;\mathbb{R}^L)$. We are interested in derivatives of f in the direction E. In the next lemma, we will see that the derivative is given by

$$f'(E) = \langle \varphi, \tilde{H}\psi\rangle_{\mathcal{H}},$$

where φ is the mild solution of the dual equation

$$i\partial_t\varphi(t) = (H_0 + \sum_{l=1}^{L} E(t)_l H_l)\varphi(t), \quad \varphi(T) = \mathcal{O}\psi(T). \tag{4.2}$$

This means that for all $t \in [0,T]$ we have

$$\varphi(t) = G(T-t)^*\mathcal{O}\psi(T) + \int_t^T G(s-t)^* E(s)\cdot\tilde{H}^*\varphi(s)\,\mathrm{d}s, \tag{4.3}$$

which can equivalently be written as

$$\varphi(t) = \mathfrak{G}(T,t)^*\mathcal{O}\psi(T). \tag{4.4}$$

We also introduce the solution ψ' and φ' of the inhomogeneous equations, respectively,

$$\begin{aligned} i\partial_t\psi'(t) &= (H_0 + \sum_{l=1}^{L} E(t)_l H_l)\psi'(t) + \sum_{l=1}^{L}\delta E(t)_l H_l\psi(t),\\ \psi'(0) &= 0, \end{aligned} \tag{4.5}$$

and

$$\begin{aligned} i\partial_t\varphi'(t) &= (H_0 + \sum_{l=1}^{L} E(t)_l H_l)\varphi'(t) + \sum_{l=1}^{L}\delta E(t)_l H_l\varphi(t),\\ \varphi'(T) &= \mathcal{O}\psi'(T). \end{aligned} \tag{4.6}$$

In mild form they read

$$\psi'(t) = \int_0^t G(t-s)E(s)\cdot\tilde{H}\psi'(s)\,\mathrm{d}s + \int_0^t G(t-s)\delta E(s)\cdot\tilde{H}\psi(s)\,\mathrm{d}s \tag{4.7}$$

and

$$\varphi'(t) = G(T-t)^*\mathcal{O}\psi'(T) + \int_t^T G(s-t)^* E(s)\cdot\tilde{H}^*\varphi'(s)\,\mathrm{d}s + \int_t^T G(s-t)^*\delta E(s)\cdot\tilde{H}\varphi(s)\,\mathrm{d}s \tag{4.8}$$

for $\delta E \in L^1(0,T;\mathbb{R}^L)$.

Using ψ, φ, ψ' and φ', we can represent the derivatives of f in the following way.

Lemma 9. *Let $E, \delta E, \tau E \in L^1(0,T;\mathbb{R}^L)$, and let ψ, ψ', φ and φ' be the corresponding solutions of* (2.6), (4.7), (4.3) *and* (4.8)*, respectively. Then the map f defined by* (4.1) *is two times continuously differentiable with derivatives*

$$f'(E)(\delta E) = \int_0^T \delta E(t)\cdot\langle\varphi(t),\tilde{H}\psi(t)\rangle_{\mathcal{H}}\,\mathrm{d}t \tag{4.9}$$

and

$$f''(E)(\delta E,\tau E) = \int_0^T \tau E\cdot\Big(\langle\varphi'(t),\tilde{H}\psi(t)\rangle_{\mathcal{H}} + \langle\varphi(t),\tilde{H}\psi'(t)\rangle_{\mathcal{H}}\Big)\,\mathrm{d}t. \tag{4.10}$$

Proof. Proposition 4 and the product rule [Lan93, XIII,§3] give continuous differentiability of f and using $\mathcal{O}^* = \mathcal{O}$, we obtain

$$f'(u)(\delta u) = \langle\mathcal{O}\psi(T),\psi'(T)\rangle.$$

By Proposition 4 and the discussion after Proposition 1, the derivative of ψ in the direction δE is given by ψ' and can be written as

$$\psi'(T) = \int_0^T \mathfrak{G}(T,t)[\delta E(t)\cdot\tilde{H}\psi(t)]\,\mathrm{d}t.$$

Together with (4.4) this implies

$$\begin{aligned} f'(E)(\delta E) &= \langle\mathcal{O}\psi(T),\psi'(T)\rangle_{\mathcal{H}} \\ &= \langle\mathcal{O}\psi(T),\int_0^T \mathfrak{G}(T,t)[\delta E(t)\cdot\tilde{H}\psi(t)]\,\mathrm{d}t\rangle_{\mathcal{H}} \\ &= \int_0^T \delta E(t)\cdot\langle\mathfrak{G}(T,t)^*\mathcal{O}\psi(T),\tilde{H}\psi(t)\rangle_{\mathcal{H}}\,\mathrm{d}t \\ &= \int_0^T \delta E\cdot\langle\varphi(t),\tilde{H}\psi(t)\rangle_{\mathcal{H}}\,\mathrm{d}t, \end{aligned}$$

which is (4.9).

For the result on the second derivative, we will again use the product rule. By Corollary 5 and the chain rule [Lan93, XIII,§3], the function φ' indeed is the derivative of φ in the direction δE. Therefore we have

$$f''(E)(\tau E, \delta E) = \int_0^T \tau E \cdot \Big(\langle \varphi'(t), \tilde{H}\psi(t) \rangle_{\mathcal{H}} + \langle \varphi(t), \tilde{H}\psi'(t) \rangle_{\mathcal{H}} \Big) \, \mathrm{d}t.$$

Since f'' is continuous, a consequence of continuous differentiability of ψ and φ, we can exchange the order of the directions [Lan93, XIII, Theorem 5.3] and obtain (4.10). □

In the next section we will use the representation of the derivatives to give optimality conditions for the Hilbert space case.

4.1.2 Optimality Conditions for the Hilbert Space Case

In this subsection we present well known results in the Hilbert space case. Using the results from the preceding section, we give optimality conditions for the Hilbert space problem (3.2) and provide representations of the gradient and Hessian of the corresponding reduced cost functional. We will see that the different choices of the control field space $\mathcal{E}$ will lead to a different qualitative behavior of the local optimizers. The explicit representations of the derivatives of the reduced functional can be used to set up an optimization method using a first-optimize-then-discretize approach.

For a Hilbert space $\mathcal{E}$ continuously embedded into $L^1(0, T; \mathbb{R}^L)$, we define the reduced functional $j \colon \mathcal{E} \to \mathbb{R}$ by

$$j_{\mathrm{hilb}}(E) = \frac{1}{2} \langle \psi(T), \mathcal{O}\psi(T) \rangle + \frac{\alpha}{2} \|E\|_{\mathcal{E}}^2,$$

where ψ is the solution of (2.6) for the field $E \in \mathcal{E}$ and $\alpha > 0$. The derivatives of j_{hilb} can be expressed in terms of ψ, φ, ψ' and φ'.

Proposition 10. *Let $E, \delta E, \tau E \in \mathcal{E}$. The first and second derivatives of j_{hilb} are given by*

$$j'_{hilb}(E)(\delta E) = \int_0^T \delta E(t) \cdot \langle \varphi(t), \tilde{H}\psi(t) \rangle_{\mathcal{H}} \, \mathrm{d}t + \alpha \langle E, \delta E \rangle_{\mathcal{E}} \tag{4.11}$$

and

$$j''_{hilb}(E)(\delta E, \tau E) = \int_0^T \tau E(t) \cdot \Big(\langle \varphi'(t), \tilde{H}\psi(t) \rangle_{\mathcal{H}} + \langle \varphi(t), \tilde{H}\psi'(t) \rangle_{\mathcal{H}} \Big) \, \mathrm{d}t + \alpha \langle \delta E, \tau E \rangle_{\mathcal{E}}.$$

Proof. The result follows from Lemma 9 and the fact that norms of Hilbert spaces are differentiable with directional derivatives given by scalar products. □

The form of the gradient and the action of the Hessian depend on the Hilbert space structure of $\mathcal{E}$. Let us first consider the case $\mathcal{E} = L^2(0,T;\mathbb{R}^L)$. Then (4.11) implies

$$\nabla j_{\text{hilb}}(E) = \alpha E + \langle \varphi, \tilde{H}\psi \rangle_{\mathcal{H}} \tag{4.12}$$

and in the optimum $\bar{E}$ we have

$$\bar{E} = -\frac{1}{\alpha}\langle \bar{\varphi}, \tilde{H}\bar{\psi} \rangle_{\mathcal{H}}. \tag{4.13}$$

For the action of the Hessian, we obtain

$$\nabla^2 j_{\text{hilb}}(E) \cdot \delta E = \alpha \delta E + \langle \varphi', \tilde{H}\psi \rangle_{\mathcal{H}} + \langle \varphi, \tilde{H}\psi' \rangle_{\mathcal{H}}. \tag{4.14}$$

For the case $\mathcal{E} = H_0^1(0,T;\mathbb{R}^L)$, we get

$$\int_0^T \partial_t(\nabla j_{\text{hilb}}(E) - \alpha E) \cdot \partial_t \delta E \,\mathrm{d}t = \int_0^T \langle \varphi, \tilde{H}\psi \rangle_{\mathcal{H}} \cdot \delta E \,\mathrm{d}t$$

for all $\delta E \in H_0^1(0,T)$. This means that $z = \nabla j_{\text{hilb}}(E) - \alpha E$ is the weak solution of

$$\begin{aligned} -\Delta z &= \langle \varphi, \tilde{H}\psi \rangle_{\mathcal{H}}, \\ z(0) &= z(T) = 0, \end{aligned} \tag{4.15}$$

which is to be understood component wise, and in the optimum $\bar{E}$ we have

$$\begin{aligned} -\Delta \bar{E} &= -\frac{1}{\alpha}\langle \bar{\varphi}, \tilde{H}\bar{\psi} \rangle_{\mathcal{H}}, \\ \bar{E}(0) &= \bar{E}(T) = 0. \end{aligned} \tag{4.16}$$

For the action of the Hessian, we obtain

$$\int_0^T \partial_t(\nabla^2 j_{\text{hilb}}(E) \cdot \delta E - \alpha \delta E) \cdot \partial_t \tau E \,\mathrm{d}t = \int_0^T \Big(\langle \varphi', \tilde{H}\psi \rangle_{\mathcal{H}} + \langle \varphi, \tilde{H}\psi' \rangle_{\mathcal{H}} \Big) \cdot \tau E \,\mathrm{d}t$$

for all $\tau E \in \mathcal{H}_0^1(0,T)$. This means that $z = \nabla^2 j_{\text{hilb}}(E) \cdot \delta E - \alpha \delta E$ is the weak solution of

$$\begin{aligned} -\Delta z &= \langle \varphi', \tilde{H}\psi \rangle_{\mathcal{H}} + \langle \varphi, \tilde{H}\psi' \rangle_{\mathcal{H}}, \\ z(0) &= z(T) = 0. \end{aligned} \tag{4.17}$$

Comparing (4.13) and (4.16), we see the smoothing effect of the H_0^1 regularization compared to the L^2 regularization. In numerical experiments we see that the function $\langle \varphi, \tilde{H}\psi \rangle_{\mathcal{H}}$ typically oscillates in time. In (4.13) we can see that the optimal control for $\mathcal{E} = L^2$ inherits those oscillations. Equation (4.16) implies that for $\mathcal{E} = H_0^1$ the optimal control is a smoothed version of the oscillating function. In our application where oscillating controls are expected and necessary, this can lead to undesirable optimal controls that have large low frequency contributions. We will discuss this effect further in Chapter 7.

In the Hilbert space case, one can also formulate second order conditions for optimality for sufficiently large α. Using Proposition 10 and the stability estimates from Corollary 3, we obtain comparable results to [WBV10, Thm. A.9].

Proposition 11. *For $\alpha > 0$ large enough, there exists a $\gamma > 0$ such that for all $E, \delta E \in \mathcal{E}$ one has*

$$j''_{hilb}(E)(\delta E, \delta E) \geq \gamma \|\delta E\|_{\mathcal{E}}^2. \tag{4.18}$$

Proof. By Proposition 10 we have

$$j''_{\text{hilb}}(E)(\delta E, \delta E) = \int_0^T \delta E(t) \cdot \Big(\langle \varphi'(t), \tilde{H}\psi(t) \rangle_{\mathcal{H}} + \langle \varphi(t), \tilde{H}\psi'(t) \rangle_{\mathcal{H}} \Big) \, \mathrm{d}t + \alpha \|\delta E\|_{\mathcal{E}}^2. \tag{4.19}$$

To give an estimate for the time integral, we use Corollary 3. We immediately obtain the estimates

$$\|\psi(t)\|_{\mathcal{H}} = \|\psi_0\|_{\mathcal{H}}$$

and

$$\|\varphi(t)\|_{\mathcal{H}} = \|\mathcal{O}\psi(T)\|_{\mathcal{H}} \leq \|\mathcal{O}\|_{\mathcal{B}(\mathcal{H})} \|\psi_0\|_{\mathcal{H}}$$

for all $t \in [0, T]$. For ψ' and φ', we obtain the estimates

$$\|\psi'(t)\|_{\mathcal{H}} \leq \int_0^t \|\delta E(s) \cdot \tilde{H}\psi(s)\|_{\mathcal{H}} \, \mathrm{d}s \leq \|\tilde{H}\| \|\psi_0\|_{\mathcal{H}} \|\delta E\|_{L^1(0,T;\mathbb{R}^L)}$$

and

$$\begin{aligned} \|\varphi'(t)\|_{\mathcal{H}} &\leq \|\mathcal{O}\psi'(T)\|_{\mathcal{H}} + \|\delta E\|_{L^1(0,T;\mathbb{R}^L)} \|\tilde{H}\| \|\mathcal{O}\|_{\mathcal{B}(\mathcal{H})} \|\psi_0\|_{\mathcal{H}} \\ &\leq 2 \|\mathcal{O}\|_{\mathcal{B}(\mathcal{H})} \|\tilde{H}\| \|\psi_0\|_{\mathcal{H}} \|\delta E\|_{L^1(0,T;\mathbb{R}^L)} \end{aligned}$$

for all $t \in [0,T]$. Therefore we have

$$\begin{aligned}\left|\int_0^T \delta E(t)\cdot\Big(\langle\varphi'(t),\tilde{H}\psi(t)\rangle_{\mathcal{H}}+\langle\varphi(t),\tilde{H}\psi'(t)\rangle_{\mathcal{H}}\Big)\,\mathrm{d}t\right| \\ \leq \|\delta E\|_{L^1(0,T;\mathbb{R}^L)}\|\tilde{H}\|\left(\|\varphi'(t)\|_{\mathcal{H}}\|\psi(t)\|_{\mathcal{H}}+\|\varphi(t)\|_{\mathcal{H}}\|\psi'(t)\|_{\mathcal{H}}\right) \\ \leq 3\|\mathcal{O}\|_{\mathcal{B}(\mathcal{H})}\|\tilde{H}\|^2\|\psi_0\|_{\mathcal{H}}^2\|\delta E\|^2_{L^1(0,T;\mathbb{R}^L)}.\end{aligned}$$

Pluging this into (4.19) yields

$$f''(E)(\delta E,\delta E)\geq\alpha\|\delta E\|_{\mathcal{E}}^2-3\|\mathcal{O}\|_{\mathcal{B}(\mathcal{H})}\|\tilde{H}\|^2\|\psi_0\|_{\mathcal{H}}^2\|\delta E\|^2_{L^1(0,T;\mathbb{R}^L)}.$$

Due to the continuous embedding of $\mathcal{E}$ into $L^1(0,T;\mathbb{R}^L)$, there exists a constant c with

$$\|\delta E\|_{\mathcal{E}}\leq c\|\delta E\|_{L^1(0,T;\mathbb{R}^L)}$$

for all $\delta E\in\mathcal{E}$. This implies

$$f''(E)(\delta E,\delta E)\geq(\alpha-3c^2\|\mathcal{O}\|_{\mathcal{B}(\mathcal{H})}\|\tilde{H}\|^2\|\psi_0\|_{\mathcal{H}}^2)\|\delta E\|^2_{L^1(0,T;\mathbb{R}^L)}.$$

Choosing

$$\alpha>3c^2\|\mathcal{O}\|_{\mathcal{B}(\mathcal{H})}\|\tilde{H}\|^2\|\psi_0\|_{\mathcal{H}}^2$$

gives (4.18) with $\gamma=\alpha-3c^2\|\mathcal{O}\|_{\mathcal{B}(\mathcal{H})}\|\tilde{H}\|^2\|\psi_0\|_{\mathcal{H}}^2>0$. □

4.2 Optimality Conditions for the Measure Space Case

Next, we study optimality conditions for the new control framework proposed in Chapter 3. The conditions are useful to understand the structure of optimal controls. First we derive necessary optimality conditions for the general case. They lead to a condition on the support of optimal measures. Then we will look at the case of discrete sparsity domain, where a special reformulation of the optimality conditions in terms of a shrinkage operation can be given. We will close the chapter with a discussion of the optimality conditions for some of the concrete realizations of the general framework for time-frequency control.

4.2.1 Optimality Conditions for General Sparsity Domains

We will derive an optimality system for optimal controls of problem (3.3). As in Section 4.1.2, we will use the reduced cost functional. The reduced cost functional j can be

decomposed as

$$j(u) = f(Bu) + g(u) \tag{4.20}$$

with the expectation value f as in Section 4.1.1 and the scaled norm

$$g(u) = \alpha \|u\|_{\mathcal{M}(\Omega;\mathcal{U})}.$$

This is a decomposition into a differential but nonconvex part and a convex but non-differentiable part. Utilizing this decomposition we prove that optimal solutions satisfy the optimality system below. The proof is done along the lines of the proof of Theorem 2.11 in [KPV14] with modifications due to the nonlinearity of the problem, compare [CHW12] and the references therein.

Proposition 12. *Let $\bar{u}$ be a minimizer of problem (3.6) and let $\bar{\psi}$ and $\bar{\varphi}$ be the solutions of (2.6) and (4.3), respectively, for the control field $B\bar{u}$. Then*

$$\alpha \|\bar{u}\|_{\mathcal{M}} = -\langle B^* \langle \bar{\varphi}, \tilde{H}\bar{\psi} \rangle_{\mathcal{H}}, \bar{u} \rangle_{C_0(\Omega;\mathcal{U}),\mathcal{M}(\Omega;\mathcal{U})} \tag{4.21}$$

and

$$\|B^* \langle \bar{\varphi}, \tilde{H}\bar{\psi} \rangle_{\mathcal{H}}\|_{C_0(\Omega;\mathcal{U})} \leq \alpha. \tag{4.22}$$

Proof. Let $\bar{u}$ be a minimizer of problem (3.6) and let $\bar{\psi}$ and $\bar{\varphi}$ be the corresponding solutions of (2.6) and (4.3). We first show the variational inequality

$$g(\bar{u}) - f'(B\bar{u})(Bu - B\bar{u}) \leq g(u). \tag{4.23}$$

Since $\bar{u}$ is optimal, we have

$$\frac{1}{h}\big(j(\bar{u} + h(u - \bar{u})) - j(\bar{u})\big) \geq 0$$

for $u \in \mathcal{M}(\Omega;\mathcal{U})$ and $h \in (0, 1)$. Using the decomposition (4.20) and convexity of g, this implies

$$\frac{1}{h}\big(f(B\bar{u} + h(Bu - B\bar{u})) - f(B\bar{u})\big) + g(u) - g(\bar{u}) \geq 0.$$

Since f is differentiable, taking the limit $h \to 0$ yields (4.23).

Testing (4.23) with $u = 0$ and $u = 2\bar{u}$ gives

$$g(\bar{u}) + f'(B\bar{u})(B\bar{u}) = 0. \tag{4.24}$$

Substituting (4.24) into (4.23) gives

$$-f'(B\bar{u})(Bu) \leq g(u) \tag{4.25}$$

for all $u \in \mathcal{M}(\Omega;\mathcal{U})$.

Using Lemma 9 on the derivative of f, equation (4.24) gives

$$g(\bar{u}) = -\langle\langle\bar{\varphi}, \tilde{H}\bar{\psi}\rangle_{\mathcal{H}}, B\bar{u}\rangle_{L^q,L^p} = -\langle B^*\langle\bar{\varphi}, \tilde{H}\bar{\psi}\rangle_{\mathcal{H}}, \bar{u}\rangle_{C_0,\mathcal{M}}$$

which proves (4.21). From (4.25) we obtain

$$-\langle B^*\langle\bar{\varphi}, \tilde{H}\bar{\psi}\rangle_{\mathcal{H}}, u\rangle_{C_0,\mathcal{M}} \leq \alpha\|u\|_{\mathcal{M}}.$$

Testing this inequality with $u = -\delta_\omega(B^*\langle\bar{\varphi}, \tilde{H}\bar{\psi}\rangle_{\mathcal{H}})(\omega)$ for some $\omega \in \Omega$ yields

$$\|(B^*\langle\bar{\varphi}, \tilde{H}\bar{\psi}\rangle_{\mathcal{H}})(\omega)\|_{\mathcal{U}}^2 \leq \alpha\|(B^*\langle\bar{\varphi}, \tilde{H}\bar{\psi}\rangle_{\mathcal{H}})(\omega)\|_{\mathcal{U}}$$

which gives (4.22). □

Remark. Proposition 12 provides only a necessary condition for local optimality. Due to the nonlinear structure of the problem (3.3), we expect that there also exist non-optimal critical points of j as well as local optima that are not global.

Remark. The proof of Proposition 12 can be decomposed into three steps. First one shows (4.23), then the two statements (4.24) and (4.25), and finally the claims (4.21) and (4.22). Each of those steps can be carried out in a more general setting. The first step only uses differentiability of $f \circ B$ and convexity of g. For the second step one additionally uses that g is positive homogeneous of degree one. The estimate (4.22) in the third step uses that g is a norm. It then follows from the duality of C_0 and $\mathcal{M}$ and the resulting characterization of a norm through the dual.

Remark. In view of the preceding remark, one can also prove Proposition 12 using more general tools from convex analysis. The proof is much more involved, but provides insights into the underlying structures and translates directly to a more general setting. First we show

$$-(f \circ B)'(u) \in \partial g(\bar{u}). \tag{4.26}$$

The inclusion (4.26) can formally be derived as

$$\bar{u} \text{ optimal} \Rightarrow 0 \in \partial j(\bar{u}) = (f \circ B)'(\bar{u}) + \partial g(\bar{u}) \Rightarrow -(f \circ B)'(\bar{u}) \in \partial g(\bar{u}).$$

This argument can be made rigorous using the correct notion of differentiability. Since we split our functional into a sum of a nonconvex and a nonsmooth part, we need to fall back on the very general differential calculus of Clarke, see [Cla90]. The optimality condition $0 \in \partial_c j(\bar{u})$ ([Cla90, Prop. 2.3.2]) can be rewritten as $0 \in \partial_c(f \circ B)(\bar{u}) + \partial_c g(\bar{u})$ ([Cla90, Cor. 1 of Prop. 2.3.3]). Since the differential of Clarke reduces to the derivative and the convex subgradient for differentiable and convex functions, respectively ([Cla90, Prop. 2.2.4 and Prop. 2.2.7]), we can rewrite this as (4.26). We can now proceed with duality arguments, see [ET99, Chap. I, Prop. 5.1]. Equation (4.26) is equivalent to

$$g(\bar{u}) + g^*(-(f \circ B)'(\bar{u})) = -(f \circ B)'(\bar{u})(\bar{u}). \tag{4.27}$$

Here the convex conjugate g^* of g is given by the convex indicator function of the set $\{\, u^* \in \mathcal{M}(\Omega;\mathcal{U})^* \mid \|u^*\|_{\mathcal{M}^*} \leq \alpha \,\}$. By Proposition 9 we have $(f \circ B)'(\bar{u}) = B^*\langle \bar{\varphi}, \tilde{H}\bar{\psi}\rangle_{\mathcal{H}}$. Since we have the additional regularity $B^*\langle \varphi, \tilde{H}\psi\rangle_{\mathcal{H}} \in C_0(\Omega;\mathcal{U})$, we obtain $g^*(-(f \circ B)'(\bar{u})) = 0$ if $\|B^*\langle \varphi, \tilde{H}\psi\rangle_{\mathcal{H}}\|_{C_0(\mathcal{U})} \leq \alpha$ and $g^*(-(f \circ B)'(\bar{u})) = \infty$ if $\|B^*\langle \varphi, \tilde{H}\psi\rangle_{\mathcal{H}}\|_{C_0(\mathcal{U})} > \alpha$. In the latter case, (4.27) cannot be satisfied. Therefore (4.22) holds and (4.27) implies (4.21).

Proposition 12 implies the following interesting conditions for the support and direction of the optimal measure.

Proposition 13. *Let $\bar{u}$, $\bar{\psi}$ and $\bar{\varphi}$ be as in Proposition 12. Then we have*

$$\operatorname{supp}|\bar{u}| \subset \{\, \omega \in \Omega \mid \|(B^*\langle \bar{\varphi}, \tilde{H}\bar{\psi}\rangle_{\mathcal{H}})(\omega)\|_{\mathcal{U}} = \alpha \,\}, \tag{4.28}$$

$$\alpha \bar{u}'(\omega) = -(B^*\langle \bar{\varphi}, \tilde{H}\bar{\psi}\rangle_{\mathcal{H}})(\omega), \quad |\bar{u}|\text{-almost everywhere.} \tag{4.29}$$

Proof. Writing equation (4.21) as an integral yields

$$\int_\Omega \Big(\alpha + \langle (B^*\langle \bar{\varphi}, \tilde{H}\bar{\psi}\rangle_{\mathcal{H}})(\omega), \bar{u}'(\omega)\rangle_{\mathcal{U}}\Big)\, \mathrm{d}|\bar{u}|(\omega) = 0. \tag{4.30}$$

For the integrand we obtain by the Cauchy–Schwarz inequality, using $\|\bar{u}'(\omega)\|_{\mathcal{U}} = 1$ and (4.22),

$$\alpha + \langle (B^*\langle \bar{\varphi}, \tilde{H}\bar{\psi}\rangle_{\mathcal{H}})(\omega), \bar{u}'(\omega)\rangle_{\mathcal{U}} \geq \alpha - \|(B^*\langle \bar{\varphi}, \tilde{H}\bar{\psi}\rangle_{\mathcal{H}})(\omega)\|_{\mathcal{U}} \geq 0. \tag{4.31}$$

Thus (4.30) says that the integral of a non-negative function vanishes. This yields

$$\alpha + \langle (B^*\langle \bar{\varphi}, \tilde{H}\bar{\psi}\rangle_{\mathcal{H}})(\omega), \bar{u}'(\omega)\rangle_{\mathcal{U}} = 0 \tag{4.32}$$

for $|\bar{u}|$-almost all $\omega \in \Omega$. For those ω the Cauchy–Schwarz inequality in (4.31) was sharp. This implies

$$(B^*\langle \bar{\varphi}, \tilde{H}\bar{\psi}\rangle_{\mathcal{H}})(\omega) = c\bar{u}'(\omega)$$

for some $c \in \mathbb{R}$. Using (4.32) we obtain $c = -\alpha$ which gives (4.29) and

$$\|(B^*\langle \bar{\varphi}, \tilde{H}\bar{\psi}\rangle_{\mathcal{H}})(\omega)\|_{\mathcal{U}} = \alpha$$

for $|\bar{u}|$-almost all $\omega \in \Omega$. Denote by h the map $\omega \mapsto \|(B^*\langle \bar{\varphi}, \tilde{H}\bar{\psi}\rangle_{\mathcal{H}})(\omega)\|_{\mathcal{U}} - \alpha$. Then h is continuous since $B^*\langle \bar{\varphi}, \tilde{H}\bar{\psi}\rangle_{\mathcal{H}} \in C_0(\Omega;\mathcal{U})$ and $h(\omega) = 0$ for $|u|$-almost all ω. The inclusion (4.28) then follows from the measure theoretic observation that a continuous function h that vanishes $|u|$-almost everywhere satisfies $\operatorname{supp}|u| \subset h^{-1}(\{0\})$. Otherwise there would without loss of generality be $\omega' \in \operatorname{supp}|u|$ with $h(\omega') > 0$. Since h is continuous, there is an open neighborhood V of ω' with $f(\omega) > 0$ for all $\omega \in V$, and $x \in V$ implies $|u|(V) > 0$. This implies $\int_V h(\omega)\,\mathrm{d}|u|(\omega) > 0$, in contradiction to $h(\omega) = 0$ for $|u|$-almost all ω. This concludes the proof. □

The relation (4.28) for the support of the optimal measure gives us the following corollary.

Corollary 14. *Let $\bar{u}$ be a local minimizer of* (3.6)*. Then* $\operatorname{supp}|\bar{u}|$ *is compact.*

Proof. Since $B^*\langle \bar{\varphi}, \tilde{H}\bar{\psi}\rangle \in C_0(\Omega;\mathcal{U})$ we know that there is a compact set $K \subset \Omega$ such that $B^*\langle \bar{\varphi}, \tilde{H}\bar{\psi}\rangle \leq \alpha/2$ for all $\omega \notin K$. Using (4.28) this implies $\operatorname{supp}|\bar{u}| \subset K$. Therefore $\operatorname{supp}|\bar{u}|$ is compact as a closed subset of the compact set K. □

This corollary has interesting implications in the case of unbounded sparsity domains. It says that although the domain Ω might be unbounded, optimal solutions will always have bounded support. This corresponds to a constraint on the implementation of lasers, where arbitrary fast oscillations are not realizable. Unfortunately, we could not prove stronger localization properties for the support. In the ideal case, one would like to obtain something like discreteness of the support. For applications estimates on the number of points in the support would be interesting.

4.2.2 Optimality Conditions for Discrete Sparsity Domains

For a discrete sparsity domain, an alternative reformulation of the first order optimality condition can be given. A discrete sparsity domain is useful to model a case of fixed discrete frequencies or coefficients in a discrete basis set. Another important case of discrete sparsity domains is given by discretized versions of optimization problems on

continuous sparsity domains. We will see that the optimality condition given in this section gives rise to an algorithm similar to the methods using sifting [Ren+06].

In this section we will use the following modification of assumption (A3).

(A3') The sparsity domain $\Omega \subset \mathbb{R}^n$ is discrete.

By a discrete set we mean a set that does only contain isolated points. We do not assume that the set is finite. Under assumption (A3') we can write elements $u \in \mathcal{M}(\Omega;\mathcal{U})$ as sums of Dirac measures,

$$u = \sum_{\omega} u_\omega \delta_\omega.$$

The mapping $u \mapsto (u_\omega)_\omega$ defines an isometric isomorphism between $\mathcal{M}(\Omega;\mathcal{U})$ and $\ell^1(\Omega;\mathcal{U})$. Then elements of $\mathcal{M}(\Omega;\mathcal{U})$ and $C_0(\Omega;\mathcal{U})$ can be considered as elements of $\ell^\infty(\Omega;\mathcal{U})$. To rewrite the optimality conditions, we use the following shrinkage operation. Let the shrinkage operator $S_c \colon \mathcal{U} \to \mathcal{U}$ be defined by

$$S_c(z) := \begin{cases} 0 & \text{if } \|z\|_{\mathcal{U}} \leq c, \\ z - c\frac{z}{\|z\|_{\mathcal{U}}} & \text{if } \|z\|_{\mathcal{U}} > c, \end{cases}$$

and extend the operator pointwise to $S_c \colon \ell^\infty(\Omega;\mathcal{U}) \to \ell^\infty(\Omega;\mathcal{U})$ by $(S_c(u))_\omega = S_c(u_\omega)$. Alternatively, S_c can also be written as

$$(S_c(u))_\omega = \max\left(0, 1 - \frac{c}{\|u_\omega\|_{\mathcal{U}}}\right) u_\omega.$$

One can now reformulate the optimality condition in Proposition 13 as follows.

Proposition 15. *Let $\bar{u}$, $\bar{\psi}$ and $\bar{\varphi}$ be as in Proposition 12. Then we have*

$$\bar{u} = S_{\gamma\alpha}(\bar{u} - \gamma B^* \langle \bar{\varphi}, \tilde{H}\bar{\psi}\rangle_{\mathcal{H}}) \tag{4.33}$$

for every $\gamma > 0$.

Proof. Let $\gamma > 0$ and $\omega \in \Omega$. We will show that

$$\bar{u}_\omega = S_{\gamma\alpha}\left(\bar{u}_\omega - \gamma(B^* \langle \bar{\varphi}, \tilde{H}\bar{\psi}\rangle_{\mathcal{H}})(\omega)\right) \tag{4.34}$$

which implies (4.33) since $\omega \in \Omega$ was arbitrary. We consider the two cases $\bar{u}_\omega = 0$ and $\bar{u}_\omega \neq 0$.

If $\bar{u}_\omega = 0$, then

$$S_{\gamma\alpha}\left(\bar{u}_\omega - \gamma(B^*\langle\bar{\varphi}, \tilde{H}\bar{\psi}\rangle_{\mathcal{H}})(\omega)\right) = S_{\alpha\gamma}\left(-\gamma(B^*\langle\bar{\varphi}, \tilde{H}\bar{\psi}\rangle_{\mathcal{H}})(\omega)\right).$$

Since $\|\gamma(B^*\langle\bar{\varphi}, \tilde{H}\bar{\psi}\rangle_{\mathcal{H}})(\omega)\|_{\mathcal{U}} \leq \gamma\alpha$ by (4.22), we obtain

$$S_{\alpha\gamma}\left(-\gamma(B^*\langle\bar{\varphi}, \tilde{H}\bar{\psi}\rangle_{\mathcal{H}})(\omega)\right) = 0 = \bar{u}_\omega.$$

Together this gives (4.34).

Now suppose $\bar{u}_\omega \neq 0$. By (4.29) we have

$$\bar{u}_\omega - \gamma(B^*\langle\bar{\varphi}, \tilde{H}\bar{\psi}\rangle_{\mathcal{H}})(\omega) = \left(1 + \frac{\gamma\alpha}{\|\bar{u}_\omega\|_{\mathcal{U}}}\right)\bar{u}_\omega,$$

and taking the norm gives

$$\left\|\left(1 + \frac{\gamma\alpha}{\|\bar{u}_\omega\|_{\mathcal{U}}}\right)\bar{u}_\omega\right\|_{\mathcal{U}} = \|\bar{u}_\omega\|_{\mathcal{U}} + \gamma\alpha > \gamma\alpha.$$

Thus the definition of $S_{\gamma\alpha}$ implies

$$\begin{aligned} S_{\gamma\alpha}\left(\bar{u}_\omega - \gamma(B^*\langle\bar{\varphi}, \tilde{H}\bar{\psi}\rangle_{\mathcal{H}})(\omega)\right) &= S_{\gamma\alpha}\left(\left(1 + \frac{\gamma\alpha}{\|\bar{u}_\omega\|_{\mathcal{U}}}\right)\bar{u}_\omega\right) \\ &= \left(1 + \frac{\gamma\alpha}{\|\bar{u}_\omega\|_{\mathcal{U}}}\right)\bar{u}_\omega - \gamma\alpha\frac{\bar{u}_\omega}{\|\bar{u}_\omega\|_{\mathcal{U}}} \\ &= \bar{u}_\omega \end{aligned}$$

which is (4.34). □

An alternative proof idea using the theory of monotone operators can be found in [GL08].

The optimality condition (4.33) gives rise to an iterative procedure. For a control u_{old} together with its state ψ_{old} and adjoint state φ_{old}, we obtain a new control by

$$u_{\text{new}} = S_{\alpha\gamma}(u_{\text{old}} - \gamma B^*\langle\varphi_{\text{old}}, \tilde{H}\psi_{\text{old}}\rangle_{\mathcal{H}}). \tag{4.35}$$

Here γ can be interpreted as a step size. This method is a proximal gradient method [Roc76]. Condition (4.33) can also serve as a basis for a semismooth Newton method [Ulb11; Pie15]. The proximal gradient method bears some similarity to the approach in [Ren+06].

There, an update of the form

$$u_{\text{new}} = u_{\text{old}} + s(u_{\text{update}}) \cdot u_{\text{update}}$$

is applied, where u_{update} is the original update and s can be thought of as a filter. In contrast, equation (4.35) can also be written using a filter as

$$u_{\text{new}} = s(u_{\text{old}} + u_{\text{update}}) \cdot (u_{\text{old}} + u_{\text{update}}).$$

Here, the filter is applied to the whole right-hand side and not just to the update. In our case, the filter is given by

$$s_\omega = \max\left(0, 1 - \frac{\alpha\gamma}{\|u_\omega\|_{\mathcal{U}}}\right).$$

In the case of sifting, we have, for example,

$$s_\omega = \frac{1}{2}\left(1 + \tanh\left(\frac{90}{\max_\omega \|u_\omega\|_{\mathcal{U}}}\left(\|u_\omega\|_{\mathcal{U}} - \frac{\max_\omega \|u_\omega\|_{\mathcal{U}}}{3}\right)\right)\right).$$

In Figure 4.1 we have plotted the two functions for suitable values of α, γ and $\max_\omega \|u_\omega\|_{\mathcal{U}}$. We will not go into detail here since we will follow a different approach for the numerics.

4.2.3 Additional Regularity of Optimal Controls

In this section implications of the optimality system for the regularity of optimal controls in the case of concrete time-frequency controls are presented. We will discuss additional regularity of optimal controls and optimal control fields for the setups of the Examples 3 and 4.

Using the optimality condition, one can often deduce additional regularity of optimal controls. In the case of control with function-valued measures, we are mainly interested in additional regularity in the function part of the measure. Utilizing the decomposition

$$\mathrm{d}u = u' \,\mathrm{d}|u|$$

this means additional regularity of the Radon–Nikodym derivative $u'(\omega)$ for fixed ω. Proposition 12 gives

$$\alpha\bar{u}'(\omega) = -(B^*\langle\bar{\varphi}, \tilde{H}\bar{\psi}\rangle_{\mathcal{H}})(\omega), \quad |\bar{u}|\text{-almost everywhere.}$$

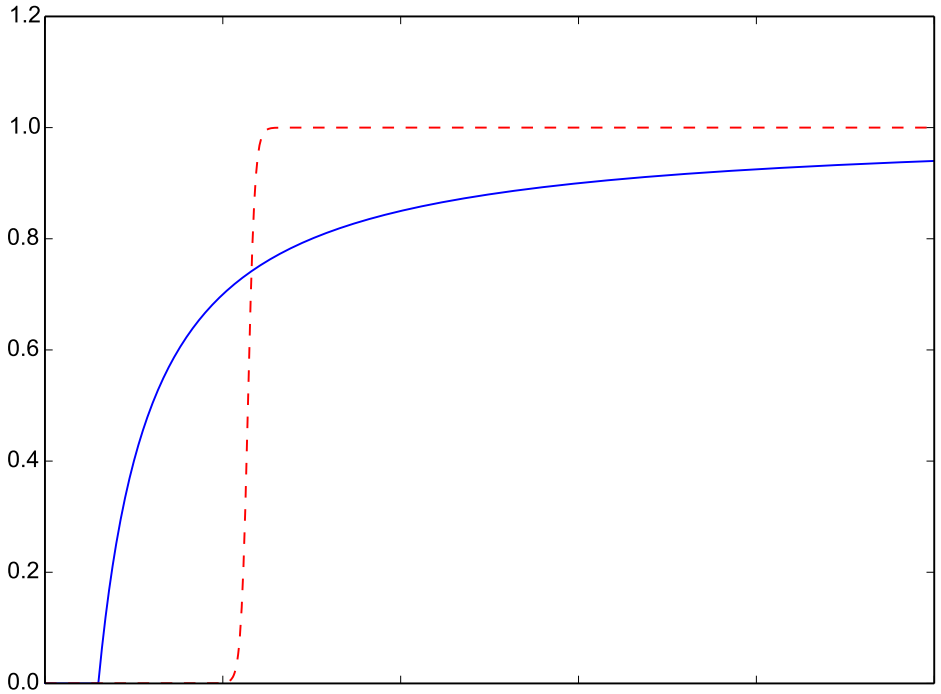


Figure 4.1: The value of the shrinkage (solid) and sifting (dashed) filters plotted against the norm of $\|u_\omega\|_{\mathcal{U}}$.

For the product $\langle \bar{\varphi}, \tilde{H}\bar{\psi}\rangle_{\mathcal{H}}$, we always obtain the regularity $C([0,T];\mathbb{R})$ since $\bar{\psi}, \bar{\varphi} \in C([0,T];\mathcal{H})$. Additional regularity for u' now follows from regularity properties of the predual control operator B^*. For general control operators and vector measure spaces, we might not obtain any additional regularity. We will discuss the concrete realizations of our general framework given by Examples 3 and 4.

For the setup in Example 3, the operator B^* is given by solutions of Poisson equations, see (3.9). Since the right hand side of the equation satisfies $\langle \bar{\varphi}, \tilde{H}\bar{\psi}\rangle_{\mathcal{H}} e^{-i\omega t} \in L^2(0,T;\mathbb{C})$, we immediately obtain the additional regularity $u'(\omega) \in H^2(0,T;\mathbb{C})$ and

$$\|u'(\omega)\|_{H^2(0,T;\mathbb{C})} \leq c$$

with a constant c independent of ω. Together with the compact support due to Corollary 14, this will give H^2 regularity of the control field Bu. For discrete sparsity domains, we can easily deduce this additional regularity since compactness implies finiteness of the support. Then the integral expression in the definition of B collapses into a finite sum. For continuous Ω the additional regularity of the control field follows from the following lemma.

Lemma 16. *Let $u \in \mathcal{M}(\mathbb{R}; L^2(0,T;\mathbb{C}))$ with $\operatorname{supp}|u|$ compact and let $k \in \mathbb{N}$ such that*

$$\|u'(\omega)\|_{H^k(0,T;\mathbb{C})} \leq c$$

for $|u|$-almost all $\omega \in \Omega$ with a constant independent of ω. Then Bu defined by (3.8) satisfies

$$Bu \in H^k(0,T;\mathbb{R}).$$

Proof. Formally exchanging differentiation and integration gives

$$\partial_t^k (Bu)(t) = \operatorname{Re} \int_\Omega \partial_t^k (u'(\omega,t) e^{i\omega t}) \,\mathrm{d}|u|(\omega).$$

We need to show that this equality holds and defines a function in $L^2(0,T;\mathbb{R})$. We will make this rigorous using Fubini's theorem [Lan93, Thm. 8.4, Thm. 8.7] and the compactness of $\operatorname{supp}|u|$.

Define $v\colon \Omega \to L^2(0,T;\mathbb{C})$ by

$$v(\omega) = \partial_t^k (u'(\omega) e^{i\omega \cdot}).$$

By the product rule we indeed have $v(\omega) \in L^2(0,T;\mathbb{C})$ with

$$\|v(\omega)\|^2_{L^2(0,T;\mathbb{C})} \leq \max\left\{ \binom{k}{l}\omega^{2l} \mid l = 0,\dots,k \right\} \|u'(\omega)\|_{H^k(0,T;\mathbb{C})} \leq c_k(1+\omega^{2k})$$

with a constant c_k depending on k. The compactness of $\operatorname{supp}|u|$ then implies $v \in L^\infty(\Omega,|u|;L^2(0,T;\mathbb{C}))$. By the boundedness of u, we then have $v \in L^2(\Omega,|u|;L^2(0,T;\mathbb{C}))$ which also implies $v \in L^1(\Omega,|u|;L^1(0,T;\mathbb{C}))$. Fubini's theorem applied to v gives us that the map $V\colon [0,T] \to \mathbb{R}$ given by

$$V(t) = \operatorname{Re}\int_\Omega v(\omega,t)\,\mathrm{d}|u|(\omega)$$

is well defined. Applying the theorem to v^2, we see that

$$\begin{aligned}\int_0^T V(t)^2\,\mathrm{d}t &= \int_0^T \left|\int_\Omega v(\omega,t)\,\mathrm{d}|u|(\omega)\right|^2 \mathrm{d}t \\ &\leq \|u\|_{\mathcal{M}(\Omega;L^2(0,T;\mathbb{R}))}\int_0^T v(\omega,t)^2\,\mathrm{d}|u|(\omega) < \infty\end{aligned}$$

where we used the Cauchy–Schwarz inequality in $L^2(\Omega,|u|;\mathbb{C})$. This gives $V \in L^2(0,T;\mathbb{R})$. It remains to check that $\partial_t^k Bu = V$. Let $\phi\colon [0,T] \to \mathbb{R}$ be a compactly supported smooth function. We need to show

$$\int_0^T (Bu)(t)\partial_t^k\phi(t)\,\mathrm{d}t = (-1)^k\int_0^T V(t)\phi(t)\,\mathrm{d}t.$$

Using the definitions of B and V, and Fubini's theorem for v and $u' \in L^1(\Omega,|u|;L^2(0,T;\mathbb{C})) \hookrightarrow L^1(\Omega,|u|;L^1(0,T;\mathbb{C}))$, we obtain

$$\begin{aligned}\int_0^T (Bu)(t)\partial_t^k\phi(t)\,\mathrm{d}t &= \operatorname{Re}\int_0^T\int_\Omega u'(\omega,t)e^{i\omega t}\partial_t^k\phi(t)\,\mathrm{d}|u|(\omega)\,\mathrm{d}t \\ &= \operatorname{Re}\int_\Omega\int_0^T u'(\omega,t)e^{i\omega t}\partial_t^k\phi(t)\,\mathrm{d}t\,\mathrm{d}|u|(\omega) \\ &= (-1)^k\operatorname{Re}\int_\Omega\int_0^T v(\omega,t)\phi(t)\,\mathrm{d}t\,\mathrm{d}|u|(\omega) \\ &= (-1)^k\operatorname{Re}\int_0^T\int_\Omega v(\omega,t)\phi(t)\,\mathrm{d}|u|(\omega)\,\mathrm{d}t \\ &= (-1)^k\int_0^T V(t)\phi(t)\,\mathrm{d}t.\end{aligned}$$

Thus $\partial_t^k Bu = V \in L^2(0,T;\mathbb{R})$. This gives $Bu \in H^k(0,T;\mathbb{R})$. □

Under stronger assumptions on initial data and the operators H_l and $\mathcal{O}$ such that the solutions ψ and φ lie in the space $C([0,T];\mathcal{V}) \cap H^1(0,T;\mathcal{V}^*)$, as discussed in Section 2.2.1, one obtains $\langle \bar{\varphi}, \tilde{H}\bar{\psi}\rangle_{\mathcal{H}} \in H^1(0,T;\mathbb{R})$. We then obtain the regularity $u'(\omega) \in H^3(0,T;\mathbb{C})$. This regularity then also translates into additional regularity for Bu.

For Example 4, the operator B^* is given by the integral representation (3.12). Additional regularity depends on the kernel k. For the case of a Gaussian kernel, B^* corresponds to a convolution with a Gaussian. This results in the additional regularity $u'(\omega) \in C^\infty([0,T];\mathbb{C})$. An argument as the one above then implies $Bu \in C^\infty([0,T];\mathbb{R})$.

5 A Generalized Suzuki–Trotter Type Method in Optimal Control

In this chapter a solver for the time-dependent quantum systems is analyzed in the context of optimal quantum control. We first introduce the generalized Suzuki–Trotter (GST) method and discuss structural properties of it. Then we derive explicit expressions for derivatives of discrete expectation values in the direction of the discrete control field. The content of this chapter is joint work with Manfred Liebmann [HL15].

The bottleneck for the numerical solution of the optimal quantum control problem is the solutions of the time-dependent quantum system. The evaluation of the expectation value in the cost functional, the associated gradient, and the evaluation of the action of the Hessian require multiple solutions of the time-dependent quantum system. Therefore, a fast solution method for time-dependent quantum systems is of major importance. There are several methods for the solution of time-dependent quantum systems, see [Lub08] for an overview. Most methods exploit that the time evolution operator of a quantum system is essentially the exponential of the Hamiltonian of the quantum system. The classical Suzuki–Trotter method approximates an exponential operator with a product ansatz of computationally simple exponential operators [Suz91; Suz90]. The GST method instead uses a product ansatz of lower order approximations of the exponential operator for the construction of the higher order approximation of the exponential operator [Lie00]. Thus computationally inexpensive linear approximations for the exponential operator can be used as building blocks for the generalized Suzuki–Trotter method. The resulting explicit scheme then generates a polynomial approximation of the exponential operator. In practice only the action of the Hamiltonian is required to construct an approximation for the time evolution operator. Therefore the GST method is easily applicable to a wide variety of quantum systems.

5.1 Structural Properties of the GST Method

In this section structural properties of the GST method are discussed. The definition and basic properties are presented as in [Lie00], with minor modifications. Then the

behavior with respect to differentiation is studied. We derived expressions for the first and second order derivatives of the time stepping scheme with respect to the control. These expressions fit naturally into the GST framework and thus can be computed efficiently. This is important in the context of optimal control to give efficiently computable representations for the gradient and Hessian of discrete expectation values.

5.1.1 Definition and Basic Properties

In this section we give an overview of the definition and basic properties of the generalized Suzuki–Trotter (GST) method. The GST method was introduced in [Lie00] and is an extension of the approximation scheme for the exponential operator presented in [Suz91; Suz90]. The main idea of the method is to approximate the exponential operator in a recursive scheme as a product of lower order approximations. With exception of Lemma 21, the results presented in this section can be found in [Lie00]. There the method is formulated in the context of a Banach algebra. We will restrict ourself to the case of algebras of bounded linear operators on a Hilbert space. Throughout this section $\mathfrak{C}$ denotes such an algebra. We present the results in dependence of a discrete control field, which will not be used until Section 5.1.2. The original proofs translate to our setting without modification. We denote by $\mathcal{E}^h$ a discrete space for the control fields. The set of functions from the space X into the space Y is denoted by Y^X.

Definition 17. Let $C \in \mathfrak{C}^{\mathbb{C}\times\mathcal{E}^h}, E^h \in \mathcal{E}^h, z \in \mathbb{C}$ and $m \in \mathbb{N}$ then the generalized Suzuki–Trotter (GST) operator $\mathcal{Q}_m \colon \mathfrak{C}^{\mathbb{C}\times\mathcal{E}^h} \to \mathfrak{C}^{\mathbb{C}\times\mathcal{E}^h}$ is recursively defined by

$$\begin{aligned}\mathcal{Q}_1(C)(z,E^h) &= C(z,E^h),\\ \mathcal{Q}_m(C)(z,E^h) &= \mathcal{Q}_{m-1}(C)(p_m z,E^h)\cdot\mathcal{Q}_{m-1}(C)(\bar{p}_m z,E^h), \quad m>1,\end{aligned}$$

where the coefficients $p_m \in \mathbb{C}$ satisfy the relations

$$p_m + \bar{p}_m = 1, \quad p_m^m + \bar{p}_m^m = 0\ .$$

Here $\bar{p}_m$ denotes the complex conjugate of p_m. The equations for the coefficients p_m in Definition 17 can be solved analytically by

$$p_m = \frac{1}{1+e^{-i\pi/m}} = \frac{1}{2} + \frac{i}{2}\tan(\pi/2m)\ .$$

The operator $\mathcal{Q}_m$ can also be defined on $\mathfrak{C}^{\mathbb{C}}$ and then extended pointwise to $\mathfrak{C}^{\mathbb{C}\times\mathcal{E}^h}$ by $\mathcal{Q}_m(C)(z,E^h) = \mathcal{Q}_m(C(\cdot,E^h))(z)$. In practice we will not use the GST method with

$z \in \mathbb{C}$ but only for $t \in \mathbb{R}$. However, for the theory it is sometimes useful to allow complex time steps. The GST operator can be written as a product,

$$\mathcal{Q}_m(C)(z, E^h) = \prod_{k=1}^{2^{m-1}} C(\alpha_k z, E^h), \tag{5.1}$$

where α_k is an $(m-1)$-fold product of the coefficients p_j or $\overline{p_j}$ with $1 < j \leq m$. The following result states the fundamental approximation property of the GST method [Lie00, Satz 1.1].

Theorem 18. *Let $A \in \mathfrak{C}^{\mathcal{E}^h}$, $C \in \mathfrak{C}^{\mathbb{C}\times\mathcal{E}^h}$, $E^h \in \mathcal{E}^h$, $z \in \mathbb{C}$ and $m \in \mathbb{N}$. If the map $z \mapsto C(z, E^h)$ is analytic and*

$$C(z, E^h) - \exp(zA(E^h)) = o(z),$$

then the map $z \mapsto \mathcal{Q}_m(C)(z, E^h)$ is analytic and

$$\mathcal{Q}_m(C)(z, E^h) - \exp(zA(E^h)) = o(z^m).$$

A good choice for the first order approximation operator C is obviously the linear approximation

$$C(z, E^h) := I + zA(E^h). \tag{5.2}$$

For the linear approximation, we obtain an explicit error bound for the approximation of the exponential, as well as a restriction on the norm of the argument for the polynomial approximation [Lie00, Satz 1.5].

Proposition 19. *Let $A \in \mathcal{C}^{\mathcal{E}^h}$, $C \in \mathfrak{C}^{\mathbb{C}\times\mathcal{E}^h}$ and $C(z, E^h) := I + zA(E^h)$ a linear approximation and $m \in \mathbb{N}$. If $s := \frac{3}{2^{m-1}}\|zA(E^h)\| < 1$, then*

$$\|\mathcal{Q}_m(C)(z, E^h) - \exp(zA(E^h))\| \leq \left(\exp\left(2^{m-1}\sum_{k=m+1}^{\infty}\frac{s^k}{k}\right) - 1\right)(1+s)^{m-1}.$$

For specific operators A, this leads to a restriction of the step size $t \in \mathbb{R}$ of the explicit time stepping scheme as well as on the size of the control E^h.

Since $\mathcal{Q}_m$ generates approximations of the exponential operator, it also approximately satisfies the group property. For even m we obtain an enhanced approximation [Lie00, Satz 1.4].

Lemma 20. *Under the assumptions of Theorem 18 with m even, we have*

$$\mathcal{Q}_m(C)(-z, E^h) \cdot \mathcal{Q}_m(C)(z, E^h) - I = o(z^{m+1}).$$

The next lemma states that the equivalence of taking the adjoint and reversing time is preserved by the GST scheme.

Lemma 21. *Let $C \in \mathfrak{C}^{\mathbb{C}\times\mathcal{E}^h}$, $E^h \in \mathcal{E}^h$ and $m \in \mathbb{N}$. If C satisfies the symmetry condition*

$$C(z, E^h)^* = C(-\bar{z}, E^h) \tag{5.3}$$

for all $z \in \mathbb{C}$, then

$$\mathcal{Q}_m(C)(z, E^h)^* = \mathcal{Q}_m(C)(-\bar{z}, E^h).$$

Proof. The result follows by induction over m using Definition 17. □

This result is important for the computation of the adjoint state. The adjoint action of the time stepping scheme can be computed by using negative time steps. For the linear approximation (5.2), the symmetry condition (5.3) for C is equivalent to the condition $A(E^h)^* = -A(E^h)$ for A. For our application we have $A(E^h) = -iH(E^h)$ for some self-adjoint $H(E^h)$, thus the condition is satisfied. For general approximations the condition on C is stronger than the skew-adjointness of A. In general $\mathcal{Q}_m(C)(t, E^h)$ with $t \in \mathbb{R}$ will not be unitary. However, under the assumptions of Lemma 20 and 21, we obtain

$$\mathcal{Q}_m(C)(t, E^h)^* \cdot \mathcal{Q}_m(C)(t, E^h) - I = o(t^{m+1}) \ .$$

5.1.2 GST Method and Differentiation

In the context of optimal control, we need to differentiate the time stepping scheme with respect to the control. A naive differentiation of the product in (5.1) leads to the inefficient expression

$$\mathcal{Q}'_m(C)(z, E^h)(\delta E^h) = \sum_{k=1}^{2^{m-1}} \left(\prod_{l=1}^{k-1} C(\alpha_l z, E^h) \right) C'(\alpha_k z, E^h)(\delta E^h) \left(\prod_{l=k+1}^{2^{m-1}} C(\alpha_l z, E^h) \right)$$

with complexity $\mathcal{O}(2^{2m-2})$. Exploiting the multiplicative structure of the algorithm, we will follow a different approach that will give the same complexity $\mathcal{O}(2^{m-1})$ as the original product. The approach can be understood as a modification of the result from [NH95], where it is shown that for an analytic function F and square matrices C and

δC, we have

$$F\begin{pmatrix} C & 0 \\ \delta C & C \end{pmatrix} = \begin{pmatrix} F(C) & 0 \\ F'(C)(\delta C) & F(C) \end{pmatrix}.$$

That is, the directional derivative of F can be obtained by applying the function to a matrix with special structure. For general directional derivatives with respect to a parameter, this matrix gets more complicated. We introduce the notation

$$D^s_{\delta E^h}(C)(z, E^h) = C^{(s)}(z, E^h)(\delta E^h)^s$$

for the directional derivative of order s in direction δE^h, the space

$$C^r_{\delta E^h}(\mathfrak{C}) = \{\, C \in \mathfrak{C}^{\mathbb{C}\times\mathcal{E}^h} \mid z \mapsto D^s_{\delta E^h}(C)(z, E^h) \text{ analytic for each } 0 \le s \le r, E^h \in \mathcal{E}^h \,\}$$

of differentiable functions, and the nilpotent $(r+1)\times(r+1)$-matrix N, given by $N_{s,s+1} = 1$, $s = 1, \ldots, r$, and zero otherwise.

Definition 22. The operator $\mathcal{D}^{(r)}_{\delta E^h} : C^r_{\delta E^h}(\mathfrak{C}) \to C^0_{\delta E^h}(\mathfrak{C}^{\otimes r+1})$ is defined by

$$\mathcal{D}^{(r)}_{\delta E^h}(C) = \sum_{s=0}^{r} \frac{1}{s!} N^s \otimes D^s_{\delta E^h} C.$$

This operator generates a matrix of derivatives in the direction δE^h up to order r. It can also formally be understood as $\mathcal{D}^{(r)}_{\delta E^h}(C) = \exp(N \otimes D_{\delta E^h})(I \otimes C)$. More explicitly, we have

$$\mathcal{D}^{(r)}_{\delta E^h}(C)(z, E^h) = \begin{pmatrix} C(z,E^h) & 0 & \cdots & \cdots & \cdots & 0 \\ \frac{1}{1!}D^1_{\delta E^h}C(z,E^h) & C(z,E^h) & 0 & \cdots & \cdots & 0 \\ \frac{1}{2!}D^2_{\delta E^h}C(z,E^h) & \frac{1}{1!}D^1_{\delta E^h}C(z,E^h)(\delta E^h) & C(z,E^h) & 0 & \ldots & 0 \\ \vdots & \vdots & & \ddots & \ddots & \vdots \\ \vdots & \vdots & & & \ddots & 0 \\ \frac{1}{r!}D^r_{\delta E^h}C(z,E^h) & \frac{1}{(r-1)!}D^{r-1}_{\delta E^h}C(z,E^h) & \cdots & \cdots & \cdots & C(z,E^h) \end{pmatrix}.$$

Alternatively, the operator $\mathcal{D}^{(r)}_{\delta E^h}$ could also be defined on differentiable functions in $\mathfrak{C}^{\mathcal{E}^h}$ first and then extended pointwise to $C^r_{\delta E^h}(\mathfrak{C})$ by $\mathcal{D}^{(r)}_{\delta E^h}(C)(z, E^h) = \mathcal{D}^{(r)}_{\delta E^h}(C(z,\cdot))(E^h)$. Note that $\mathcal{C}^{\otimes r+1}$ can be considered as a space of bounded linear operators. We obtain the following crucial property for $D^{(r)}_{\delta E^h}$.

Lemma 23. *For $A, B \in C^r_{\delta E^h}(\mathfrak{C})$,*

$$\mathcal{D}^{(r)}_{\delta E^h}(A \cdot B) = \mathcal{D}^{(r)}_{\delta E^h}(A) \cdot \mathcal{D}^{(r)}_{\delta E^h}(B).$$

Proof. Using the general Leibniz rule, we obtain

$$\mathcal{D}^{(r)}_{\delta E^h}(A \cdot B) = \sum_{s=0}^{r} \sum_{k+l=s} N^s \otimes \frac{1}{k!l!} D^k_{\delta E^h} A \cdot D^l_{\delta E^h} B.$$

On the other hand

$$\mathcal{D}^{(r)}_{\delta E^h}(A) \cdot \mathcal{D}^{(r)}_{\delta E^h}(B) = \sum_{k=0}^{r} \sum_{l=0}^{r} N^{k+l} \otimes \frac{1}{k!l!} D^k_{\delta E^h} A \cdot D^l_{\delta E^h} B.$$

Since $N^{k+l} = 0$ for $k + l > r$, the two double sums on the right coincide. □

This result can be interpreted as multiplicativity of the differential operator $D^{(r)}_{\delta E^h}$, which is of course not true for the directional derivative itself. The multiplicativity implies the following commutativity relation of differentiation and the GST scheme.

Theorem 24. *Let $\delta E^h \in \mathcal{E}^h$ and $r, m \in \mathbb{N}$. Then*

$$\mathcal{Q}_m \circ \mathcal{D}^{(r)}_{\delta E^h} = \mathcal{D}^{(r)}_{\delta E^h} \circ \mathcal{Q}_m$$

on $C^r_{\delta E^h}(\mathfrak{C})$.

Proof. We prove the result by induction over m. For $m = 1$ we have $\mathcal{Q}_1(C) = C$ for all $C \in C^r_{\delta E^h}(\mathfrak{C})$ and the claim holds. Let the result be true for $m - 1$. We will use the preceding lemma and the multiplicative structure of $\mathcal{Q}_m$. Let $z \in \mathbb{C}$, $E^h \in \mathcal{E}^h$ and $C \in C^r_{\delta E^h}(\mathfrak{C})$. Using the definition of $\mathcal{Q}_m$, Lemma 23, the induction hypothesis, and the definition of $\mathcal{Q}_m$ again, we obtain

$$\begin{aligned}
\mathcal{D}^{(r)}_{\delta E^h}(\mathcal{Q}_m(C))(z, E^h) &= \mathcal{D}^{(r)}_{\delta E^h}\Big(\mathcal{Q}_{m-1}(C)(p_m z, \cdot) \cdot \mathcal{Q}_{m-1}(C)(\overline{p_m} z, \cdot)\Big)(E^h) \\
&= \mathcal{D}^{(r)}_{\delta E^h}\Big(\mathcal{Q}_{m-1}(C)(p_m z, \cdot)\Big)(E^h) \cdot \mathcal{D}^{(r)}_{\delta E^h}\Big(\mathcal{Q}_{m-1}(C)(\overline{p_m} z, \cdot)\Big)(E^h) \\
&= \mathcal{Q}_m(\mathcal{D}^{(r)}_{\delta E^h}(C))(p_m z, \delta E^h) \cdot \mathcal{Q}_m(\mathcal{D}^{(r)}_{\delta E^h}(C))(\overline{p_m} z, \delta E^h) \\
&= \mathcal{Q}_m(\mathcal{D}^{(r)}_{\delta E^h}(C))(z, E^h).
\end{aligned}$$

Since z, E^h and C were arbitrary, we get $\mathcal{D}^{(r)}_{\delta E^h} \circ \mathcal{Q}_m = \mathcal{Q}_m \circ \mathcal{D}^{(r)}_{\delta E^h}$. □

The theorem says that we can obtain a matrix of derivatives of the higher order time stepping operators by applying the GST operator to a matrix of derivatives of the first order time stepping operator. In the case $r = 1$, we get

$$\mathcal{Q}_m \begin{pmatrix} C(E^h) & 0 \\ C'(E^h)(\delta E^h) & C(E^h) \end{pmatrix} = \begin{pmatrix} \mathcal{Q}_m(C)(E^h) & 0 \\ \mathcal{Q}_m(C)'(E^h)(\delta E^h) & \mathcal{Q}_m(C)(E^h) \end{pmatrix}.$$

Second derivatives can be obtained in two ways. For derivatives in the same direction, one can use Theorem 24 for $r = 2$. This yields

$$\mathcal{Q}_m \begin{pmatrix} C(E^h) & 0 & 0 \\ C'(E^h)(\delta E^h) & C(E^h) & 0 \\ \frac{1}{2}C''(E^h)(\delta E^h, \delta E^h) & C'(E^h)(\delta E^h) & C(E^h) \end{pmatrix}$$
$$= \begin{pmatrix} \mathcal{Q}_m(C)(E^h) & 0 & 0 \\ \mathcal{Q}_m(C)'(E^h)(\delta E^h) & \mathcal{Q}_m(C)(E^h) & 0 \\ \frac{1}{2}\mathcal{Q}_m(C)''(E^h)(\delta E^h, \delta E^h) & \mathcal{Q}_m(C)'(E^h)(\delta E^h) & \mathcal{Q}_m(C)(E^h) \end{pmatrix}.$$

For second derivatives in different directions, one can use

$$\mathcal{Q}_m \circ \mathcal{D}^{(r)}_{\delta E^h} \circ \mathcal{D}^{(r)}_{\tau E^h} = \mathcal{D}^{(r)}_{\delta E^h} \circ \mathcal{D}^{(r)}_{\tau E^h} \circ \mathcal{Q}_m \tag{5.4}$$

which follows immediately from Theorem 24. We obtain

$$\mathcal{Q}_m \begin{pmatrix} C(E^h) & 0 & 0 & 0 \\ C'(E^h)(\delta E^h) & C(E^h) & 0 & 0 \\ C'(E^h)(\tau E^h) & 0 & C(E^h) & 0 \\ C''(E^h)(\delta E^h, \tau E^h) & C'(E^h)(\tau E^h) & C'(E^h)(\delta E^h) & C(E^h) \end{pmatrix}$$
$$= \begin{pmatrix} \mathcal{Q}_m(C)(E^h) & 0 & 0 & 0 \\ \mathcal{Q}_m(C)'(E^h)(\delta E^h) & \mathcal{Q}_m(C)(E^h) & 0 & 0 \\ \mathcal{Q}_m(C)'(E^h)(\tau E^h) & 0 & \mathcal{Q}_m(C)(E^h) & 0 \\ \mathcal{Q}_m(C)''(E^h)(\delta E^h, \tau E^h) & \mathcal{Q}_m(C)'(E^h)(\tau E^h) & \mathcal{Q}_m(C)'(E^h)(\delta E^h) & \mathcal{Q}_m(C)(E^h) \end{pmatrix}.$$

Using this approach we do not compute the derivatives themselves, but always actions of a lower triangular operator matrix as in

$$\begin{pmatrix} \mathcal{Q}_m(C)(E^h) & 0 \\ \mathcal{Q}_m(C)'(E^h)(\delta E^h) & \mathcal{Q}_m(C)(u_h) \end{pmatrix} \begin{pmatrix} x^h \\ y^h \end{pmatrix}.$$

Often the whole expression is needed, as we will see in Lemma 27. The derivative $\mathcal{Q}_m(C)'(E^h)(\delta E^h)x^h$ together with $\mathcal{Q}_m(C)(E^h)y^h$ can be obtained by computing the right-hand side of

$$\begin{pmatrix} \mathcal{Q}_m(C)(E^h)x^h \\ \mathcal{Q}_m(C)'(E^h)(\delta E^h)y^h \end{pmatrix} = \begin{pmatrix} I & 0 \\ -I & I \end{pmatrix} \begin{pmatrix} \mathcal{Q}_m(C)(E^h) & 0 \\ \mathcal{Q}_m(C)'(E^h)(\delta E^h) & \mathcal{Q}_m(C)(u_h) \end{pmatrix} \begin{pmatrix} x^h \\ y^h \end{pmatrix},$$

see Algorithm 5 and 6 for specific implementations.

5.2 Computing the Derivative of Expectation Values

In this section the derivatives of discrete expectation values with respect to the discrete control field are studied. First we give a short overview of how the state is discretized. This is then used to formulate a discrete analogue of the expectation value in the continuous cost functional. We present expressions for the derivative of the discrete expectation value similar to the results for the continuous expectation values in Section (4.1.1).

When solving optimal control problems, one has to distinguish two approaches. One possibility is a first-optimize-then-discretize (OTD) approach where we discretize the continuous functional as well as the continuous equations for the derivatives. Another possibility is a first-discretize-then-optimize (DTO) approach where we first write down a discrete version of the cost functional and then the correct discrete derivatives for this function. From the continuous case, we know that second derivatives of the expectation value result in solutions of inhomogeneous quantum systems. This poses a problem for some solution methods since inhomogeneous quantum systems might not be supported. We shall outline a simple approach to tackle this problem for the GST method in a DTO approach.

5.2.1 Discretization of the State Equation

In order to obtain a discrete optimization problem, we need to approximate the expectation value f by a discrete version f^h. The main ingredient will be the approximation of the state ψ by a discrete state ψ^h. We will just sketch how to obtain such a discrete state in the case of a time-dependent problem with possibly unbounded operators. We proceed in three steps.

First, we approximate the action of the evolution operator of the state equation (2.1) by a product of exponential functions. We introduce a time grid $(t_n)_{n=0}^N$ and the space $\mathcal{E}^h$ of linear finite elements on this grid. We approximate $\mathfrak{G}(t+t_{n-1}, t_{n-1})$ by $\exp(tA_n(E^h))$.

Here A_n is obtained by a Magnus expansion for the time-dependent Hamiltonian, see [Bla+09] for an overview the Magnus expansion.

As a second step, we approximate the exponential of the possibly unbounded operator A_n by an exponential of a bounded operator. This corresponds to substituting (2.1) by a semi-discrete equation. To this end we approximate the Hilbert space $\mathcal{H}$ by a discrete space $\mathcal{H}^h$. We define discrete versions $H_0^h, H_l^h \in \mathcal{B}(\mathcal{H}^h)$ of the operators H_0 and H_l to give a discrete version A_n^h of A_n. We also approximate the initial state ψ_0 by $\psi_0^h \in \mathcal{H}^h$. If the quantum system was already discrete, the discretization of it can of course be skipped. The exponential $\exp(tA_n(E^h))$ is then approximated by $\exp(tA_n^h(E^h))$. Examples for the discretization of the state include approximation with finite differences or finite elements, as well as spectral representations.

The third step consists in approximating the exponential with the GST method. That is $\exp(tA_n^h(E^h)) \approx \mathcal{Q}_m(C_n)(t, E^h)$ for some first order approximation C_n of $\exp(tA_n(E^h))$, which is sufficiently smooth.

The discrete state ψ^h is then given by the explicit formula

$$\psi_n^h = \psi_n^h(E^h) = \prod_{p=1}^{n} Q_p(E^h)\psi_0^h, \quad n = 1, \dots, N, \tag{5.5}$$

where

$$Q_p(E^h) = \mathcal{Q}_m(C_p)(t_p - t_{p-1}, E^h) \tag{5.6}$$

is the discrete time stepping operator. Here and in the following, we use the convention $\prod_{i=j}^{k} Q_i = Q_k \cdots Q_j$, that is, the order in product expressions is reversed.

In the approximation of ψ by ψ^h, each of the three steps introduces some error. We will not give an error analysis here. The first step needs results on the convergence of the Magnus expansion, compare [HL03]. For the second step, one would need to use a suitable version of the Kato–Trotter theorem, compare [IK98]. A justification for the third step is given in Section 5.1.1.

5.2.2 Derivative of the Discrete Expectation Value

In addition to the discrete state, we also need a discrete version $\mathcal{O}^h \in B(\mathcal{H}^h)$ of the observable $\mathcal{O}$ to define the discrete expectation value. We then set

$$f^h(E^h) = \frac{1}{2}\langle \psi_N^h, \mathcal{O}^h \psi_N^h \rangle_{\mathcal{H}^h}$$

where ψ_N^h is the discrete state for the control field E^h according to (5.5).

Following an OTD approach, we could discretize equations (2.1), (4.2), (4.5), and (4.6) to compute approximations of the derivatives of f according to the continuous equations (4.9) and (4.10). For the adjoint equation (4.2), we obtain

$$\varphi_n^h = \varphi_j^h(E^h) = \left(\prod_{i=j+1}^{N} Q_i(E^h) \right)^* \mathcal{O}^h \psi_N^h(E^h), \quad j = 1, \dots, N. \tag{5.7}$$

But it is not obvious how the inhomogeneous equations (4.5) and (4.6) can be discretized. Another problem with a naive OTD approach is that the gradient will not be consistent and in addition the approximated Hessian will in general not be symmetric. Following a DTO approach, we have to derive expressions for the derivatives of f^h. This typically requires more work in setting up the equations, but it is a constructive approach in the sense that we know how to correctly discretize (4.5) and (4.6). We will see that the correct discretization of the scalar products in (4.11) leads to additional computations. In Section 5.2.2 we will numerically compare the results of the two approaches.

We give representations of the derivatives of f^h. We denote by $\psi_j^{h\prime}$ and $\varphi_j^{h\prime}$ the derivatives of ψ_j^h and φ_j^h, respectively, for the control field E^h and in the direction δE^h. The discrete version of Proposition 9 then reads as follows.

Lemma 25. *Let $E^h, \delta E^h, \tau E^h \in \mathcal{E}^h$. The first and second derivatives of f^h are given by*

$$f^{h\prime}(E^h)(\delta E^h) = \sum_{n=1}^{N} \langle Q_n'(E^h)(\delta E^h)^* \varphi_n^h, \psi_{n-1}^h \rangle_{\mathcal{H}^h} \tag{5.8}$$

and

$$\begin{aligned} f^{h\prime\prime}(E^h)(\delta E^h, \tau E^h) = \sum_{n=1}^{N} \; & \langle Q_n'(E^h)(\tau E^h)^* \varphi_n^{h\prime}, \psi_{n-1}^h \rangle_{\mathcal{H}^h} \\ & + \langle Q_n'(E^h)(\tau E^h)^* \varphi_n^h, \psi_{n-1}^{h\prime} \rangle_{\mathcal{H}^h} \\ & + \langle Q_n''(E^h)(\delta E^h, \tau E^h)^* \varphi_n^h, \psi_{n-1}^h \rangle_{\mathcal{H}^h}. \end{aligned} \tag{5.9}$$

Proof. By the product rule applied to (5.5), we obtain

$$\psi_N^{h\prime} = \sum_{n=1}^{N} \prod_{p=n+1}^{N} Q_p(E^h) Q_n'(E^h)(\delta E^h) \prod_{p=1}^{n-1} Q_p(E^h) \psi_0^h.$$

Therefore, using (5.5) and (5.7),

$$\begin{aligned} f^{h\prime}(E^h)(\delta E^h) &= \langle \mathcal{O}^h \psi_N^h, \psi_N^{h\prime} \rangle_{\mathcal{H}^h} \\ &= \sum_{n=1}^{N} \langle Q_n'(E^h)(\delta E^h)^* \left(\prod_{p=n+1}^{N} Q_p(E^h) \right)^* \mathcal{O}^h \psi_N^h, \prod_{p=1}^{n-1} Q_p(E^h) \psi_0^h \rangle_{\mathcal{H}^h} \\ &= \sum_{n=1}^{N} \langle Q_n'(E^h)(\delta E^h)^* \varphi_n^h, \psi_{n-1}^h \rangle_{\mathcal{H}^h}. \end{aligned}$$

Differentiating this expression again and using the symmetry of the derivative in δE^h and τE^h concludes the proof. □

Let us compare the results of the preceding proposition to the approximation of derivatives in an OTD approach. For this bilinear problem the important difference between DTO and OTD is the discretization of the product $\langle \tilde{H}^* \varphi, \psi \rangle$. Naive discretization of (4.9) might result in an expression of the form

$$\sum_{n=1}^{N} (t_n - t_{n-1}) \langle \tilde{H}^{h*} \varphi_n^h, \psi_n^h \rangle_{\mathcal{H}^h}.$$

This means that $Q_n'^*$ in the DTO approach corresponds to $(t_n - t_{n-1}) Q_n^* \tilde{H}^{h*}$ in the OTD approach. The same effect can be observed for discretization of the second derivatives using (4.10). Additionally, we obtain in the discrete case another term involving Q''^*, which has no counterpart in the continuous case.

For a fixed basis (h_k) of $\mathcal{E}^h$, we can derive expressions for the coordinates of the gradient and Hessian-vector product. The coordinates Z of an element $E^h \in \mathcal{E}^h$ are determined by the equation $E^h = \sum_k Z_k h_k$. Let the mass matrix M of $\mathcal{E}^h$ with respect to the basis (h_k) be given by $M_{kl} = \langle h_k, h_l \rangle_{\mathcal{E}^h}$. Then the coordinates of the gradient and Hessian-vector product can be obtained from the directional derivatives by solving a linear system of equations.

Corollary 26. *Let $E^h, \delta E^h \in \mathcal{E}^h$, and let Z be the coordinates of $\nabla f^h(E^h)$. Then Z solves*

$$MZ = X$$

with

$$X_k = \sum_{n=1}^{N} \langle Q_n'(E^h)(h_k)^* \varphi_n^h, \psi_{n-1}^h \rangle_{\mathcal{H}^h}.$$

Let Z be the coordinates of $\nabla^2 f^h(E^h)\cdot\delta E^h$. Then Z solves

$$MZ = Y$$

with

$$\begin{aligned} Y_k = \sum_{n=1}^{N} \quad & \langle Q_n'(E^h)(h_k)^*\varphi_n^{h\prime}, \psi_{n-1}^h\rangle_{\mathcal{H}^h} \\ & + \langle Q_n'(E^h)(h_k)^*\varphi_n^h, \psi_{n-1}^{h\prime}\rangle_{\mathcal{H}^h} \\ & + \langle Q_n''(E^h)(\delta E^h, h_k)^*\varphi_n^h, \psi_{n-1}^h\rangle_{\mathcal{H}^h}. \end{aligned}$$

Proof. Test (5.8) with $\delta E^h = h_k$ and (5.9) with $\tau E^h = h_k$. □

The main cost in computing gradients or Hessian actions of f^h is in assembling X and Y. This splits into the computation of discrete solutions of PDEs and the action of the operators Q_j' and Q_j''. For the computation of the derivatives of Q_n, we can use Theorem 24. The following lemma addresses the iterative computation of the states and their derivatives.

Lemma 27. *Let $E^h, \delta E^h \in \mathcal{E}^h$. Then ψ_n^h, $\psi_n^{h\prime}$, φ_n^h, $\varphi_n^{h\prime}$ satisfy the forward and backward recursions*

$$\begin{pmatrix}\psi_0^h\\ \psi_0^{h\prime}\end{pmatrix} = \begin{pmatrix}\psi_0^h\\ 0\end{pmatrix}, \quad \begin{pmatrix}\psi_n^h\\ \psi_n^{h\prime}\end{pmatrix} = \begin{pmatrix} Q_n(E^h) & 0\\ Q_n'(E^h)(\delta E^h) & Q_n(E^h)\end{pmatrix}\begin{pmatrix}\psi_{n-1}^h\\ \psi_{n-1}^{h\prime}\end{pmatrix}$$

and, respectively,

$$\begin{pmatrix}\varphi_N^h\\ \varphi_N^{h\prime}\end{pmatrix} = \begin{pmatrix}\mathcal{O}^h\psi_N^h\\ \mathcal{O}^h\psi_N^{h\prime}\end{pmatrix}, \quad \begin{pmatrix}\varphi_n^h\\ \varphi_n^{h\prime}\end{pmatrix} = \begin{pmatrix} Q_{n+1}(E^h)^* & 0\\ Q_{n+1}'(E^h)(\delta E^h)^* & Q_{n+1}(E^h)^*\end{pmatrix}\begin{pmatrix}\varphi_{n+1}^h\\ \varphi_{n+1}^{h\prime}\end{pmatrix}.$$

Proof. The definitions of $\psi_n^h(E^h)$ and $\varphi_n^h(E^h)$ imply

$$\psi_n^h(E^h) = Q_n(E^h)\psi_{n-1}^h(E^h)$$

and

$$\varphi_n^h(E^h) = Q_{n+1}(E^h)^*\varphi_{n+1}^h(E^h).$$

Differentiation of the equations in the direction δE^h yields

$$\psi_n^{h\prime}(E^h)(\delta E^h) = Q_n'(E^h)(\delta E^h)\psi_n^h(E^h) + Q_n(E^h)\psi_{n-1}^{h\prime}(E^h)(\delta E^h)$$

and

$$\varphi_n^{h\prime}(E^h)(\delta E^h) = Q'_{n+1}(E^h)(\delta E^h)^* \varphi_{n+1}^h(E^h) + Q_{n+1}(E^h)^* \varphi_{n+1}^{h\prime}(E^h)(\delta E^h).$$

Additionally we have the initial and terminal conditions $\psi_0^{h\prime}(E^h) = 0$, $\varphi_N^h(E^h) = \mathcal{O}^h \psi_N^h(E^h)$ and $\varphi_N^{h\prime}(E^h)(\delta E^h) = \mathcal{O}^h \psi_N^{h\prime}(E^h)(\delta E^h)$. Writing those equations in matrix and vector form completes the proof. □

To set up the vectors X and Y for the computations of the derivatives of f^h, the whole time history of ψ^h and φ^h is needed. Saving the states for all time steps poses a restriction on the size of the problem. To reduce the memory requirements, we use a trick already used in the beginning of quantum control. We will not save the whole history of quantum states $(\psi_n^h)_n$ and $(\varphi_n^h)_n$ and their derivatives computed in a forward and backward iteration, and thereafter set up the vectors X and Y to compute the gradient and Hessian actions. Instead, we will save only the current iterates and assemble X and Y on the fly during the backwards iteration. This method introduces an additional backwards solve of the forward equation for the state and its derivative and an additional error due to the non-unitarity of the time stepping scheme. The following lemma tells us that we can indeed compute solutions of the (inhomogeneous) forward equations for ψ^h (and $\psi^{h\prime}$) backwards in time with a reasonable error.

Lemma 28. *Let $E^h, \delta E^h \in \mathcal{E}^h$, and let C_n satisfy the symmetry condition* (5.3) *and the additional approximation property*

$$D^1_{\delta E^h}\big(C_n(z, E^h) - \exp(zA_n(E^h))\big) = o(z)\ .$$

Then ψ_n^h and $\psi_n^{h\prime}$ satisfy the backward recursion

$$\begin{pmatrix} \psi_n^h \\ \psi_n^{h\prime} \end{pmatrix} = \begin{pmatrix} Q_{n+1}(E^h)^* & 0 \\ Q'_{n+1}(E^h)(\delta E^h)^* & Q_{n+1}(E^h)^* \end{pmatrix} \begin{pmatrix} \psi_{n+1}^h \\ \psi_{n+1}^{h\prime} \end{pmatrix} + o(\Delta t_{n+1}^{m+1}),$$

where $\Delta t_n = t_n - t_{n-1}$ and m is the order of approximation as in (5.6).

Proof. We will use the result for the forward recursion from Lemma 27 and combine it with the approximate group property in Lemma 20 and the relation between taking adjoints and inverting time from Lemma 21. We have

$$\begin{pmatrix} \psi_{n+1}^h \\ \psi_{n+1}^{h\prime} \end{pmatrix} = \begin{pmatrix} Q_{n+1} & 0 \\ Q'_{n+1} & Q_{n+1} \end{pmatrix} \begin{pmatrix} \psi_n^h \\ \psi_n^{h\prime} \end{pmatrix}.$$

Therefore we get

$$\begin{pmatrix} Q^*_{n+1} & 0 \\ Q'^*_{n+1} & Q^*_{n+1} \end{pmatrix} \begin{pmatrix} \psi^h_{n+1} \\ \psi^{h\prime}_{n+1} \end{pmatrix} = \begin{pmatrix} Q^*_{n+1} & 0 \\ Q'^*_{n+1} & Q^*_{n+1} \end{pmatrix} \begin{pmatrix} Q_{n+1} & 0 \\ Q'_{n+1} & Q_{n+1} \end{pmatrix} \begin{pmatrix} \psi^h_n \\ \psi^{h\prime}_n \end{pmatrix}. \tag{5.10}$$

Under the additional assumption on A_{n+1} and C_{n+1}, we have

$$\begin{pmatrix} C_{n+1}(\Delta t_{n+1}, E^h) & 0 \\ C'_{n+1}(\Delta t_{n+1}, E^h)(\delta E^h) & C_{n+1}(\Delta t_{n+1}, E^h) \end{pmatrix} \\ = I + \Delta t_{n+1} \begin{pmatrix} A_{n+1}(E^h) & 0 \\ A'_{n+1}(E^h)(\delta E^h) & A_{n+1}(E^h) \end{pmatrix} + o(\Delta t_{n+1}).$$

This implies

$$\begin{pmatrix} C_{n+1}(\Delta t_{n+1}, E^h) & 0 \\ C'_{n+1}(\Delta t_{n+1}, E^h)(\delta E^h) & C_{n+1}(\Delta t_{n+1}, E^h) \end{pmatrix} - \\ \exp\left(\Delta t_{n+1} \begin{pmatrix} A_{n+1}(E^h) & 0 \\ A'_{n+1}(E^h)(\delta E^h) & A_{n+1}(E^h) \end{pmatrix} \right) = o(\Delta t_{n+1}).$$

Thus, we can use Lemma 21 and 20 for even m to obtain

$$\begin{aligned} &\begin{pmatrix} Q^*_{n+1} & 0 \\ Q'^*_{n+1} & Q^*_{n+1} \end{pmatrix} \begin{pmatrix} Q_{n+1} & 0 \\ Q'_{n+1} & Q_{n+1} \end{pmatrix} \\ &\quad = \left[\mathcal{Q}_m \begin{pmatrix} C_{n+1} & 0 \\ C'_{n+1} & C_{n+1} \end{pmatrix} (-\Delta t_{n+1}, E^h) \right] \left[\mathcal{Q}_m \begin{pmatrix} C_{n+1} & 0 \\ C'_{n+1} & C_{n+1} \end{pmatrix} (\Delta t_{n+1}, E^h) \right] \\ &\quad = I + o(\Delta t^{m+1}_{n+1}). \end{aligned}$$

Plugging this into (5.10) concludes the proof. □

The additional approximation property in the preceding lemma is satisfied for the linear approximation since there we have

$$\begin{aligned} D^1_{\delta E^h}\big(C_n(z, E^h) - \exp(zA_n(E^h))\big) &= zA'_n(E^h)(\delta E^h) - \exp(zA_n(E^h))zA'_n(E^h)(\delta E^h) \\ &= o(z)\ . \end{aligned}$$

In fact it is satisfied for all polynomial approximations. It might be that the additional assumption already is implied by the assumption $C \in C^r_{\delta E^h}(\mathfrak{C})$.

We will refer to the approaches that save the whole history as exact and memory

inefficient DTO. Approaches based on Lemma 28 will be referred to as memory efficient. If we do not specify the approach, we always mean the memory efficient version.

The scheme for computations of gradients and Hessian actions of f^h are given by Algorithm 1 and 2. The subroutines for the assembly of the right-hand side are in

Data: E^h
Result: v^h
$\psi^h \leftarrow \psi_0^h$;
$X \leftarrow 0$;
for $n = 1, \dots, N$ **do**
  $\psi^h \leftarrow Q_n(E^h)\psi^h$;
end
$\varphi^h \leftarrow \mathcal{O}^h\psi^h$;
for $n = N, \dots, 1$ **do**
  ψ^h, φ^h, $X \leftarrow$ assemble_rhs_gradient(E^h, n; ψ^h, φ^h, X);
end
solve for Z: $MZ = X$;
$v^h \leftarrow \sum_{k=1}^{K} Z_k h_k$;

Algorithm 1: Compute the gradient of f^h

their most general form given by Algorithm 3 and 4. In every step they modify the vectors X and Y, respectively, and evolve the states.

Different discretizations of the control field give rise to different assemble_rhs_gradient and assemble_rhs_hessian routines. For piecewise linear controls and a Magnus expansion of order two using the midpoint rule, we obtain Algorithm 5 and 6. There the loop over k reduces to a common update of just two entries of X and Y, respectively. Notice how these algorithms compute the derivatives of the time stepping via Theorem 24 and reuse results from the action of the large operator matrices. The second derivatives for the Hessian action are computed using equation (5.4). The updates of X and Y in a ODT approach are given in Algorithm 7 and 8. In contrast to the DTO approach, for the OTD approach the updates are computed in the same way as the continuous derivatives in Section 4.1.1.

Comparing the computational complexity of assemble_rhs_gradient for the DTO and OTD approach, we see that the DTO approach has roughly 1.5 times the cost of the OTD approach. We need an application of a 2×2 time stepping matrix instead of just an application of the ordinary time stepping operator. For assemble_rhs_hessian the cost increases roughly by a factor 2. Here we need an additional application of the 4×4 time stepping matrix.

Data: E^h, δE^h
Result: w^h
$\begin{pmatrix}\psi^h\\ \psi^{h\prime}\end{pmatrix} \leftarrow \begin{pmatrix}\psi_0^h\\ 0\end{pmatrix}$;
$Y \leftarrow 0$;
for $n = 1, \dots, N$ **do**
$\quad \begin{pmatrix}\psi^h\\ \psi^{h\prime}\end{pmatrix} \leftarrow \begin{pmatrix} Q_n(E^h) & 0 \\ Q_n'(E^h)(\delta E^h) & Q_n(u_h)\end{pmatrix}\begin{pmatrix}\psi^h\\ \psi^{h\prime}\end{pmatrix}$;
end
$\begin{pmatrix}\varphi^h\\ \varphi^{h\prime}\end{pmatrix} \leftarrow \begin{pmatrix}\mathcal{O}^h\psi^h\\ \mathcal{O}^h\psi^{h\prime}\end{pmatrix}$;
for $n = N, \dots, 1$ **do**
$\quad \psi^h, \psi^{h\prime}, \varphi^h, \varphi^{h\prime}, Y \leftarrow$ assemble_rhs_hessian(E^h, δE^h, n; ψ^h, $\psi^{h\prime}$, φ^h, $\varphi^{h\prime}$, Y);
end
solve for Z: $MZ = Y$;
$w^h \leftarrow \sum_{k=1}^K Z_k h_k$;

Algorithm 2: Apply the Hessian of f^h to a vector δE^h

Data: E^h, n, ψ^h, φ^h, X
Result: ψ^h, φ^h, X
$\psi^h \leftarrow Q_n(E^h)^*\psi^h$;
for $k = 1, \dots, K$ **do**
$\quad X_k \leftarrow X_k + \langle Q_n'^*(E^h)(h_k)\varphi^h, \psi^h\rangle_{\mathcal{H}^h}$;
end
$\varphi^h \leftarrow Q_n(E^h)^*\varphi^h$;

Algorithm 3: assemble_rhs_gradient for DTO and general controls

Data: E^h, n, ψ^h, $\psi^{h\prime}$, φ^h, $\varphi^{h\prime}$, Y
Result: ψ^h, $\psi^{h\prime}$, φ^h, $\varphi^{h\prime}$, Y
$\begin{pmatrix}\psi^h\\ \psi^{h\prime}\end{pmatrix} \leftarrow \begin{pmatrix} Q_n(E^h)^* & 0 \\ Q_n'(E^h)(\delta E^h)^* & Q_n(u_h)^*\end{pmatrix}\begin{pmatrix}\psi^h\\ \psi^{h\prime}\end{pmatrix}$;
for $k = 1, \dots, K$ **do**
$\quad Y_k \leftarrow Y_k + \langle Q_n'(E^h)(h_k)^*\varphi^{h\prime}, \psi^h\rangle_{\mathcal{H}^h} + \langle Q_n'(E^h)(h_k)^*\varphi^h, \psi^{h\prime}\rangle_{\mathcal{H}^h}$;
$\qquad\qquad + \langle Q_n''(E^h)(\delta E^h, h_k)^*\varphi^h, \psi^h\rangle_{\mathcal{H}^h}$
end
$\begin{pmatrix}\varphi^h\\ \varphi^{h\prime}\end{pmatrix} \leftarrow \begin{pmatrix} Q_n(E^h)^* & 0 \\ Q_n'(E^h)(\delta E^h)^* & Q_n(u_h)^*\end{pmatrix}\begin{pmatrix}\varphi^h\\ \varphi^{h\prime}\end{pmatrix}$;

Algorithm 4: assemble_rhs_hessian for DTO and general controls

Data: E^h, n, ψ^h, φ^h, X
Result: ψ^h, φ^h, X
$\psi^h \leftarrow Q_n(E^h)^* \psi^h$;
$\chi^h \leftarrow \varphi^h$;
$\begin{pmatrix} \varphi^h \\ \chi^h \end{pmatrix} \leftarrow \begin{pmatrix} Q_n(E^h)^* & 0 \\ Q_n'(E^h)(h_n)^* & Q_n(E_h)^* \end{pmatrix} \begin{pmatrix} \varphi^h \\ \chi^h \end{pmatrix}$;
$\chi^h \leftarrow \chi^h - \varphi^h$;
$\delta X \leftarrow \langle \chi^h, \psi^h \rangle_{\mathcal{H}^h}$;
$X_n \leftarrow X_n + \delta X$;
$X_{n-1} \leftarrow X_{n-1} + \delta X$;

Algorithm 5: assemble_rhs_gradient for DTO piecewise linear controls

Data: E^h, δE^h, n, ψ^h, $\psi^{h\prime}$, φ^h, $\varphi^{h\prime}$, Y
Result: ψ^h, $\psi^{h\prime}$, φ^h, $\varphi^{h\prime}$, Y
$\begin{pmatrix} \psi^h \\ \psi^{h\prime} \end{pmatrix} \leftarrow \begin{pmatrix} Q_n(E^h)^* & 0 \\ Q_n'(E^h)(\delta E^h)^* & Q_n(E_h)^* \end{pmatrix} \begin{pmatrix} \psi^h \\ \psi^{h\prime} \end{pmatrix}$;
;
$\chi_2^h \leftarrow \varphi^{h\prime}$;
$\chi_3^h \leftarrow \varphi^{h\prime}$;
$\begin{pmatrix} \varphi^{h\prime} \\ \chi_2^h \end{pmatrix} \leftarrow \begin{pmatrix} Q_n(E^h)^* & 0 \\ Q_n'(E^h)(h_n)^* & Q_n(E_h)^* \end{pmatrix} \begin{pmatrix} \varphi^{h\prime} \\ \chi_2^h \end{pmatrix}$;
$\chi_2^h \leftarrow \chi_2^h - \varphi^{h\prime}$;
$\varphi^{h\prime} \leftarrow \chi_3^h$;
;
$\chi_1^h \leftarrow \varphi^h$;
$\begin{pmatrix} \varphi^h \\ \varphi^{h\prime} \\ \chi_1^h \\ \chi_3^h \end{pmatrix} \leftarrow$
$\begin{pmatrix} Q_n(E^h)^* & 0 & 0 & 0 \\ Q_n'(E^h)(\delta E^h)^* & Q_n(E^h)^* & 0 & 0 \\ Q_n'(E^h)(h_n)^* & 0 & Q_n(E^h)^* & 0 \\ Q_n''(E^h)(\delta E^h, h_n)^* & Q_n(E^h)(h_n)^* & Q_n'(E^h)(\delta E^h)^* & Q_n(u^h)^* \end{pmatrix} \begin{pmatrix} \varphi^h \\ \varphi^{h\prime} \\ \chi_1^h \\ \chi_3^h \end{pmatrix}$;
$\chi_1^h \leftarrow \chi_1^h - \varphi^h$;
$\chi_3^h \leftarrow \chi_3^h - \varphi^{h\prime} - \chi_2^h$;
;
$\delta Y \leftarrow \langle \chi_1^h, \psi^{h\prime} \rangle_{\mathcal{H}^h} + \langle \chi_2^h, \psi^h \rangle_{\mathcal{H}^h} + \langle \chi_3^h, \psi^h \rangle_{\mathcal{H}^h}$;
$Y_n \leftarrow Y_n + \delta Y$;
$Y_{n-1} \leftarrow Y_{n-1} + \delta Y$;

Algorithm 6: assemble_rhs_hessian for DTO and piecewise linear controls

Data: E^h, n, ψ^h, φ^h, X
Result: ψ^h, φ^h, X
$\delta X \leftarrow \langle \tilde{H}^{h*}\varphi^h, \psi^h \rangle_{\mathcal{H}^h}$;
$X_n \leftarrow X_n + \delta X$;
$\psi^h \leftarrow Q_n(E^h)^* \psi^h$;
$\varphi^h \leftarrow Q_n(E^h)^* \varphi^h$;

Algorithm 7: assemble_rhs_gradient for OTD and piecewise linear controls

Data: E^h, δE^h, n, ψ^h, $\psi^{h\prime}$, φ^h, $\varphi^{h\prime}$, Y
Result: ψ^h, $\psi^{h\prime}$, φ^h, $\varphi^{h\prime}$, Y
$\delta Y \leftarrow \langle \tilde{H}^{h*}\varphi^h, \psi^{h\prime} \rangle_{\mathcal{H}^h} + \langle \tilde{H}^{h*}\varphi^{h\prime}, \psi^h \rangle_{\mathcal{H}^h}$;
$Y_n \leftarrow Y_n + \delta Y$;

$$\begin{pmatrix} \psi^h \\ \psi^{h\prime} \end{pmatrix} \leftarrow \begin{pmatrix} Q_n(E^h)^* & 0 \\ Q_n'(E^h)(\delta E^h)^* & Q_n(E^h)^* \end{pmatrix} \begin{pmatrix} \psi^h \\ \psi^{h\prime} \end{pmatrix};$$

$$\begin{pmatrix} \varphi^h \\ \varphi^{h\prime} \end{pmatrix} \leftarrow \begin{pmatrix} Q_n(E^h)^* & 0 \\ Q_n'(E^h)(\delta E^h)^* & Q_n(E^h)^* \end{pmatrix} \begin{pmatrix} \varphi^h \\ \varphi^{h\prime} \end{pmatrix};$$

Algorithm 8: assemble_rhs_hessian for OTD and piecewise linear controls

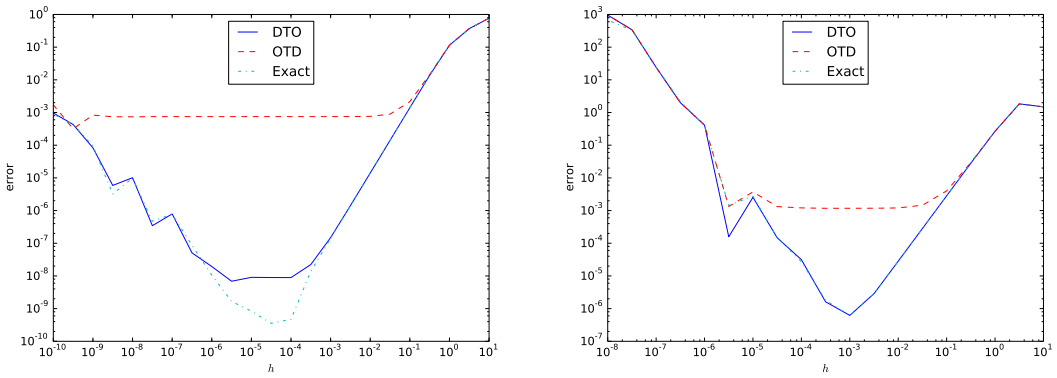


Figure 5.1: The difference between finite differences and the memory efficient DTO (solid), the memory efficient OTD (dashed) and the exact memory inefficient DTO (dotted) representations of the gradient (left) and Hessian action (right).

Let us compare the consistency of the derivatives generated by the memory efficient DTO approach with the derivatives generated by the OTD approach and the memory inefficient DTO. That means we study how well the approximations of the derivatives match the actual derivatives of the discrete expectation value f^h. To this end we compare the representations obtained through the calculus in this section to finite difference approximations of the derivatives with varying step size h. In Figure 5.1 we plotted the absolute value of the difference divided by h against h. Comparing finite differences to a good approximation of the derivatives results in a V shape. The right-hand side of the V corresponds to the convergence of the finite differences with decreasing h and the left-hand side results from numerical cancellations for very small h. The gradient and Hessian computed by the exact but memory inefficient DTO approach show this behavior since they are the exact derivatives up to rounding errors. We see that the memory efficient DTO approach leads to an approximation comparable to the exact DTO approach and is much better than the OTD approach. A better approximation of the derivatives can lead to more robustness of the optimization method. With the OTD gradient, we often got stuck at suboptimal controls, because the computed gradient direction did not lead to a decrease of the functional. This justifies using the memory efficient DTO implementation in the context of the GST method.

5.2.3 Discrete Optimization Problem in the Hilbert Space Case

As in the continuous case, it is easy to discuss the discrete optimization problem in the Hilbert space case. The discrete reduced cost functional is given by

$$j^h_{\mathrm{hilb}}(E^h) = f^h(E^h) + \frac{\alpha}{2}\|E^h\|^2_{\mathcal{E}^h}.$$

The discrete version of (3.2) in reduced form then reads

$$\text{Minimize} \quad j^h_{\mathrm{hilb}}(E^h), \quad E^h \in \mathcal{E}^h.$$

Using the results on the derivatives of f^h, we immediately obtain the following result for the derivatives of j^h_{hilb}.

Corollary 29. *Let $E^h, \delta E^h \in \mathcal{E}^h$, and let Z be the coordinates of $\nabla j^h_{hilb}(E^h) - \alpha E^h$. Then Z solves*

$$MZ = X \tag{5.11}$$

with

$$X_k = \sum_{n=1}^{N} \langle Q_n'(E^h)(h_k)^* \varphi_n^h, \psi_{n-1}^h \rangle_{\mathcal{H}^h}.$$

Let Z be the coordinates of $\nabla^2 j_{hilb}^h(E^h)(\delta E^h) - \alpha \delta E^h$. Then Z solves

$$MZ = Y \tag{5.12}$$

with

$$\begin{aligned} Y_k = \sum_{n=1}^{N} \quad & \langle Q_n'(E^h)(h_k)^* \varphi_n^{h\prime}, \psi_{n-1}^h \rangle_{\mathcal{H}^h} \\ & + \langle Q_n'(E^h)(h_k)^* \varphi_n^h, \psi_{n-1}^{h\prime} \rangle_{\mathcal{H}^h} \\ & + \langle Q_n''(E^h)(\delta E^h, h_k)^* \varphi_n^h, \psi_{n-1}^h \rangle_{\mathcal{H}^h}. \end{aligned}$$

For different spaces $\mathcal{E}^h$, we have different mass matrices M. This affects the gradient and Hessian-vector product for j_{hilb}^h in a similar way as the choice of $\mathcal{E}$ on gradient and Hessian in the continuous case in Section 4.1.2. In the case of $\mathcal{E}^h$ being a finite element space with the L^2 scalar product, M is the standard mass matrix for finite elements. Then equation (5.11) and (5.12) are discrete versions of the continuous equations (4.12) and (4.15). In the case of linear finite elements equipped with the H_0^1 norm, the matrix M is the stiffness matrix for the Poisson equation with homogeneous Dirichlet boundary conditions. Then equations (5.11) and (5.12) are discrete analogs of the continuous equations (4.14) and (4.17).

6 The Discrete Optimization Problem

In this chapter we will formulate the discrete optimization problem and discuss how it can be solved. To obtain an optimal control problem that can be solved numerically, we first discretize the control space and control operator. The resulting discrete problem is then regularized. This regularized discrete problem will be used for the numerical experiments in the following chapter.

6.1 The Discrete Optimization Problem in the Measure Space Case

The goal of this section is to give a precise formulation of the discrete optimal control problem that we will solve in the following chapter. We proceed in two steps. First we discretize the measure space and the control operator to write down a discrete version of (3.3). Then we will Huber-regularize the resulting problem to make standard methods for smooth optimization applicable for its solution.

6.1.1 Discretization of the Measure Space

In this section we discretize the measure space $\mathcal{M}(\Omega;\mathcal{U})$ to obtain a finite dimensional space $\mathcal{M}^h$. We then give a discrete version $B^h\colon \mathcal{M}^h \to \mathcal{E}^h$ of the control operator B.

The discretized measure space $\mathcal{M}^h$ is given by a measure space on a finite sparsity domain Ω^h and with values in a finite dimensional function space $\mathcal{U}^h$,

$$\mathcal{M}^h = \mathcal{M}(\Omega^h;\mathcal{U}^h).$$

The choice of Ω^h and $\mathcal{U}^h$ of course depends on their continuous counterparts Ω and $\mathcal{U}$. If Ω and $\mathcal{U}$ are already discrete, then it might be possible to choose $\Omega^h = \Omega$ and $\mathcal{U}^h = \mathcal{U}$.

For continuous sparsity domains Ω we will just use a uniform and finite grid of points $\Omega^h \subset \Omega$. The restriction to finite and therefore bounded Ω^h is partly justified by Corollary 14. We use uniform grids for simplicity. Since optimal controls ideally have very localized frequency supports, adaptive grids might significantly reduce the number of

degrees of freedom in the discretization. Since we choose Ω to be finite, we are in the setting of discrete sparsity domains explained in Section 4.2.2.

If $\mathcal{U}$ is a function space like $H_0^1(0,T;\mathbb{C})$ or $L^2(0,T;\mathbb{C})$ we will use linear finite element spaces for $\mathcal{U}^h$. There we will use the same time discretization as for the space $\mathcal{E}^h$. The main difference between $\mathcal{U}^h$ and $\mathcal{E}^h$ is that the former contains complex-valued finite element functions, the latter only real-valued ones. Since $\mathcal{U}^h$ contains the discrete envelopes for the different frequencies, which should vary slowly in time, we could also choose a much coarser time discretization for $\mathcal{U}^h$ than for $\mathcal{E}^h$. Since finite element spaces form finite dimensional Hilbert spaces, we remain in the theoretical setting of Section 4.2.2. In fact, most of the results in the current Section 6.1.1 can be generalized to the setting of discrete sparsity domains.

We will always choose conforming discretization, that is $\mathcal{M}^h$ fulfills the relation

$$\mathcal{M}^h \subset \mathcal{M}(\Omega;\mathcal{U}).$$

We will not discuss in which sense the discrete space $\mathcal{M}^h$ approximates the continuous space $\mathcal{M}(\Omega;\mathcal{U})$.

We will identify measures $u^h \in \mathcal{U}^h$ with the collection (u^h_ω) in $\mathcal{U}^h$ where

$$u = \sum_{\omega\in\Omega^h} u^h_\omega \delta_\omega.$$

The natural norm on $\mathcal{M}^h$ is given by

$$\|u^h\|_{\mathcal{M}^h} = \sum_\omega \|u^h_\omega\|_{\mathcal{U}^h}. \tag{6.1}$$

But since $\mathcal{M}^h$ is finite dimensional we can equip $\mathcal{M}^h$ with a Hilbert space structure,

$$\langle u, v\rangle = \sum_\omega \langle u_\omega, v_\omega\rangle_{\mathcal{U}^h}, \tag{6.2}$$

which gives rise to an equivalent norm. With respect to this structure we will be able to talk about gradients and Hessians, and can apply standard optimization methods.

To formulate a fully discrete version of the optimal control problem 3.3, we also need to give a discrete control operator B^h. Since we use conforming discretizations we can define the restriction $B\restriction_{\mathcal{M}^h}\colon \mathcal{M}^h \to \mathcal{E}$. But this restriction does often not map to $\mathcal{E}^h$. For this we would need the additional property $B(\mathcal{M}^h) \subset \mathcal{E}^h$. Even for matching discretizations of $\mathcal{U}^h$ and $\mathcal{E}^h$ this will not be satisfied for the control operators defined in

the examples in Chapter 3. Therefore, an additional discretizatoin step is needed. We will use discrete control operators $B^h \colon \mathcal{M}^h \to \mathcal{E}^h$ of the form

$$B^h u^h = \operatorname{Re} \sum_{\omega} \Lambda_\omega^h K^h u_\omega^h. \tag{6.3}$$

Here $K^h \colon \mathcal{U}^h \to \mathcal{U}^h$ is a suitable discrete version of K in Example 4, or can also just be the identity. The map $\Lambda_\omega \colon \mathcal{U}^h \to \mathcal{E}^h \oplus i\mathcal{E}^h$ is a discrete version of the multiplication with the function $e^{i\omega t}$ in Examples 3, 4 and 6 or a modulated version of this function in Example 7. The real part then acts as a map $\operatorname{Re} \colon \mathcal{E}^h \oplus i\mathcal{E}^h \to \mathcal{E}^h$.

Using the discrete measure space and the discrete control operator, and together with the discrete expectation value defined in the preceding chapter, we can define a discrete optimal control problem. For the discrete reduced cost functional

$$j^h(u^h) = f^h(B^h u^h) + \alpha \|u^h\|_{\mathcal{M}^h}$$

we obtain the optimization problem

$$\text{Minimize} \quad j^h(u^h), \quad u^h \in \mathcal{M}^h. \tag{6.4}$$

This is a nonsmooth and nonconvex optimization problem, as the original problem (3.3). We will not solve this problem directly, but instead regularize it to make it smooth.

6.1.2 The Regularized Discrete Optimization Problem

For the numerical experiments we will not solve the discrete problem (6.4) directly. Instead of applying a method adapted to the non-smoothness of the problem, we Huber-regularize the discrete problem and make it smooth. This means that in the discrete cost functional we replace the norm of $\mathcal{U}^h$ in (6.1) by the function $H^\theta \colon \mathcal{U}^h \to \mathbb{R}$ given by

$$H^\theta(z) = \begin{cases} \|z\|_{\mathcal{U}^h} - \frac{\theta}{2}, & \text{if } \|z\|_{\mathcal{U}^h} > \theta, \\ \frac{1}{2\theta}\|z\|_{\mathcal{U}^h}^2, & \text{if } \|z\|_{\mathcal{U}^h} \leq \theta, \end{cases}$$

for some regularization parameter $\theta > 0$. The function H^θ has the following two important properties: it is differentiable, and the derivatives of H^θ and the norm of $\mathcal{U}^h$ have the same behavior away from zero. The first property makes the optimization problem differentiable. The second property preserves the possibility of strongly peaked solutions. The smoothness, however, will come at the cost of relaxing the support property.

Other smooth functions with the two properties can also be used.

We define the regularized reduced cost functional

$$j^{h\theta}(u^h) = f^h(B^h u^h) + \alpha \sum_\omega H^\theta(u^h_\omega)$$

and the corresponding regularized optimal control problem

$$\text{Minimize} \quad j^{h\theta}(u^h), \quad u^h \in \mathcal{M}^h. \tag{6.5}$$

Since $j^{h\theta}$ is differentiable we can easily derive expressions for its gradient. The gradient is given with respect to the Hilbert space structure generated by the scalar product (6.2). This product is also used to construct B^{h*}.

Lemma 30. *Let $u^h \in \mathcal{M}^h$ and $\theta > 0$. The gradient of $j^{h\theta}$ is given by*

$$\nabla j^{h\theta}(u^h) = B^{h*} \nabla f^h(B^h u^h) + \alpha \sum_\omega \nabla H^\theta(u^h_\omega),$$

where

$$\nabla H^\theta(z) = \begin{cases} \frac{z}{\|z\|_{\mathcal{U}^h}}, & \text{if } \|z\|_{\mathcal{U}^h} > \theta, \\ \frac{z}{\theta}, & \text{if } \|z\|_{\mathcal{U}^h} \leq \theta. \end{cases}$$

Proof. The only thing that needs to be checked is the differentiability of H^θ on the sphere $\{ z \in \mathcal{U}^h \mid \|z\|_{\mathcal{U}^h} = 1 \} = \theta$. This is a routine calculation. □

We can also formally derive an expression for the second derivative. It can be made rigorous in the context of semismooth derivatives [Ulb11]. We then have

$$\nabla^2 j^{h\theta}(u^h) \cdot \delta u^h = B^*(\nabla^2 f^h(B^h u^h) \cdot \delta u^h) + \alpha \sum_\omega \nabla^2 H^\theta(u^h_\omega) \cdot \delta u^h_\omega,$$

where

$$\nabla^2 H^\theta(z) \cdot \delta z = \begin{cases} \frac{\delta z}{\|z\|_{\mathcal{U}^h}} - \frac{\langle z, \delta z \rangle_{\mathcal{U}^h} z}{\|z\|^3_{\mathcal{U}^h}}, & \text{if } \|z\|_{\mathcal{U}^h} > \theta, \\ \frac{\delta z}{\theta}, & \text{if } \|z\|_{\mathcal{U}^h} \leq \theta. \end{cases}$$

These expressions for the first and second derivative will be used to set up optimization routines.

We will not go into detail on the choice of the regularization parameter θ and on the limit $\theta \to 0$. In our numerical examples θ is chosen at least two orders of magnitude smaller than $\max_\omega \|u_\omega\|_{\mathcal{U}^h}$. In Section 7.1 we will numerically observe the effect of the regularization on optimal controls.

6.2 Optimization Methods

Let us discuss optimization methods to solve the optimal quantum control problem (6.5). Since we regularized the problem, we can apply methods for smooth optimization. Since the problem is nonconvex we need to use a globalized method and might also need a reasonable initial guess.

For the computation of solutions to the regularized problem (6.5) in the next chapter we will user a limited memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) method with a strong Wolfe line search [NW06]. This method is based on the computations of gradients of $j^{h\theta}$ and uses the history of controls and gradients to generate an approximation of the Hessian. To compute the gradient of $j^{h\theta}$ we use Lemma 30 and Corollary 26. The stopping criterion is a relative decrease of norm of the gradient by 10^{-6}. For small α, the initial guess u_0^h for the control uses a fixed element $\tilde{u}_0^h$ in $\mathcal{U}^h$ for all $\omega \in \Omega^h$ with a phase θ_ω randomly chosen from a uniform distribution on $[0, 2\pi[$, $u_{0,\omega}^h = e^{i\theta_\omega}\tilde{u}_0^h$. For larger α, where such a generic initial guess leads to convergence of the method to suboptimal critical points near the origin, we use optimal solutions for smaller α as initial guess.

In numerical experiments we compared the performance of different optimization methods. A similar comparison for the Hilbert space case is given in [HL15]. Since using the calculus for the GST method from Chapter 5 makes second derivatives of j^h easily accessible, we are able to formally apply second order methods based on the Hessian. Since we solve problems with a large number of degrees of freedom, in the experiments in the next section up to 409200, it is not feasible to save the whole history of controls and gradients as in the original BFGS method. We compared the L-BFGS method a Barzilai–Borwein (BB) method with a non-monotone line search [Ray97] and a trust region Newton (TRN) method [NW06]. Plots of the convergence history are given in Figure 6.1. The L-BFGS is the fastest of the methods, with respect to the number of steps as well as the needed computational time. It is interesting that it outperforms the TRN method. This might have several reasons. First, the functional we use is not actually two times differentiable. The discontinuity in the second derivative might cause troubles. A second reason might be the nonlinearity of the problem. For most of the time the TRN method is not close to an optimum, where fast convergence can be expected. Another reason might be that the Hessian generated using the GST method in its memory efficient implementation is in general not symmetric. The trust region globalization we used for the Newton method uses a Steihaug conjugate gradient (CG) method for the solution of the trust region subproblems [NW06, Alg. 7.2]. This

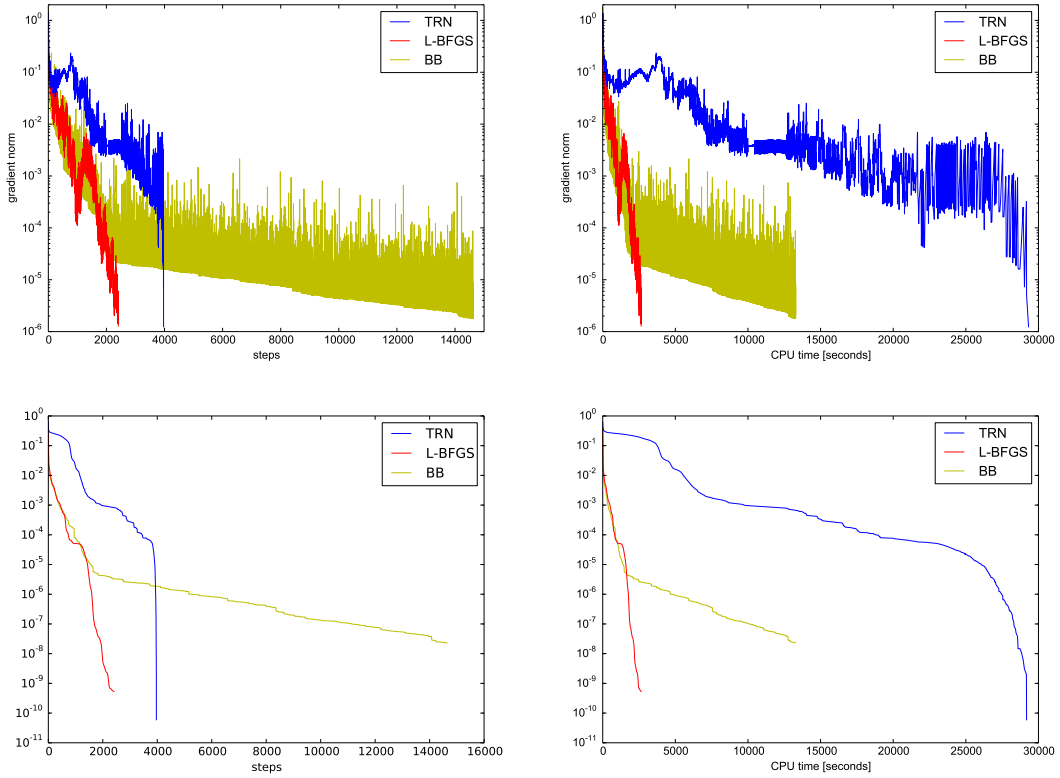


Figure 6.1: Convergence history for different optimization methods of the gradient norm (top) and the difference of the cost functional to the minimal computed value (bottom), plotted against the number of steps (left) and the computational time (right).

CG method relies on the symmetry of the Hessian. The only preconditioning we use for the CG method is the implicit preconditioning through the scalar product of $\mathcal{M}^h$. The BB method performs much worse than the L-BFGS method. We also do not observe its possible upside of being more robust with respect to ending up in undesirable local optima. For the numerical examples in the next chapter we therefore chose to use the L-BFGS method.

An interesting alternative for the solution of the control problem would be to not regularize it but tackle (6.4) directly. One way to do this is to use a fixed point iteration for the optimality system as described in Section 4.2.2. The optimality system could also be solved using a Newton type method, see [GL08] in the not vector-valued case. An alternative way to regularize the problem is to add an $L^2(\Omega;\mathcal{U})$ Tikhonov term to the cost functional. This leads to problems of the form as in [HSW12; Pie15].

As a reference for our new control approach, we also compute solutions for the Hilbert space case. There we use a TRN method, which performs considerably better in that setting [HL15]. The optimal quantum control problem in the Hilbert space case is typically solved using a gradient based method [PDR88; Hoh+07; KHK10]. Second order methods like Newtons method are less frequently used [WBV10; Hin+13].

7 Applications

In this chapter, we will apply the framework for time-frequency control in two examples. We consider a finite dimensional system with three levels and a system of two coupled one-dimensional Schrödinger equations. In the first examples, we will give a detailed analysis of optimal controls with respect to optimality conditions and structure of the field. In the more challenging second example, we will compare optimal controls for different realizations of our general control framework.

7.1 Three Level System

As our first example, we look at a finite dimensional quantum system. Here, we choose a problem with three levels to obtain the simplest model where more than one transition frequency is needed. For this example, we use the matrices

$$H_0 = \begin{pmatrix} -2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \quad H_1 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

as drift and coupling Hamiltonians, respectively, on the Hilbert space $\mathcal{H} = \mathbb{C}^3$. The control objective is to reach the second eigenstate starting from the first eigenstate. The initial condition ψ_0 and observable $\mathcal{O}$ are given by

$$\psi_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathcal{O} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The first and second levels are not directly coupled and so the population transfer must happen trough level three. The expected transition frequencies corresponding to the differences in the eigenvalues are $\omega_1 = 4$ for the transition between levels one and three, and $\omega_2 = 3$ for the transition between levels three and two. We chose a time horizon of $T = 100$ to allow for sufficiently many oscillations with the transition frequencies.

For the control space we choose $\Omega = [2, 5]$ and $\mathcal{U} = H_0^1(0, T; \mathbb{C})$. That means we look

control problem	α	$j^{h\theta}$	f^h
measure space case	10^{-01}	$2.8818 * 10^{-03}$	$7.8098 * 10^{-07}$
Hilbert space case	10^{-04}	$7.4283 * 10^{-06}$	$1.9942 * 10^{-11}$

Table 7.1: Comparison of the cost functional and expectation value for the measure space case in $\mathcal{M}(\Omega; H_0^1(0,T;\mathbb{C}))$ and the Hilbert space case $L^2(0,T)$.

for controls $u \in \mathcal{M}([2,5]; H_0^1(0,T;\mathbb{C}))$. The expected transition frequencies ω_1 and ω_2 are contained in Ω. The control operator B is given by (3.8). Therefore we are in the setting of Example 3. We chose a cost parameter of $\alpha = 10^{-1}$. As a comparison we also solve the Hilbert space problem for a control field $E \in \mathcal{E} = L^2(0,T;\mathbb{R})$ and $\alpha = 10^{-4}$.

The state only needs to be discretized in time since $\mathcal{H}$ already is finite dimensional. In time we use a uniform grid $(t_n)_{n=0}^N$ with $N = 4096$ points. The space $\mathcal{E}^h$ is given by linear finite elements on this grid. For the time evolution of the state we use the GST method for the linear approximation

$$C_n(z, E^h) = I + zA_n(E^h)$$

where A_n is given by the midpoint rule,

$$A_n(E^h) = -i\left(H_0 + \frac{E^h(t_{n-1}) + E^h(t_n)}{2} H_1\right).$$

Then C_n is an analytical first order approximation of the exponential function and satisfies the symmetry condition $C_n(z, E^h)^* = C_n(-\bar{z}, E^h)$ and the additional assumption $D^1_{\delta E^h}(C_n(z, E^h) - \exp(zA_n(E^h))) = o(z)$ from Lemma 27. Using the GST scheme we increased the order of the approximation to $m = 4$. The discrete state is then given by (5.5).

For the discretization of the frequency domain Ω we use a uniform grid with 100 grid points. The frequencies ω_1 and ω_2 are contained in the discrete frequency domain. The space $\mathcal{U}$ is discretized by complex valued linear finite elements on the same uniform grid as $\mathcal{E}^h$. This results in a discrete measure space $\mathcal{M}^h$ with $100 \cdot 2 \cdot 4096 = 819200$ real degrees of freedom. The control operator is discretized as in (6.3).

In Table 7.1 we give the values of the discrete cost functional for the measure space case and the Hilbert space case. Looking at the discrete expectation value, we see that the values below 10^{-6} are attained. This value corresponds to a probability of more than 99.9998% to achieve the control goal to enter the second eigenspace. The value for the L^2 control is smaller than the value for the $\mathcal{M}$ control. This might be due to the

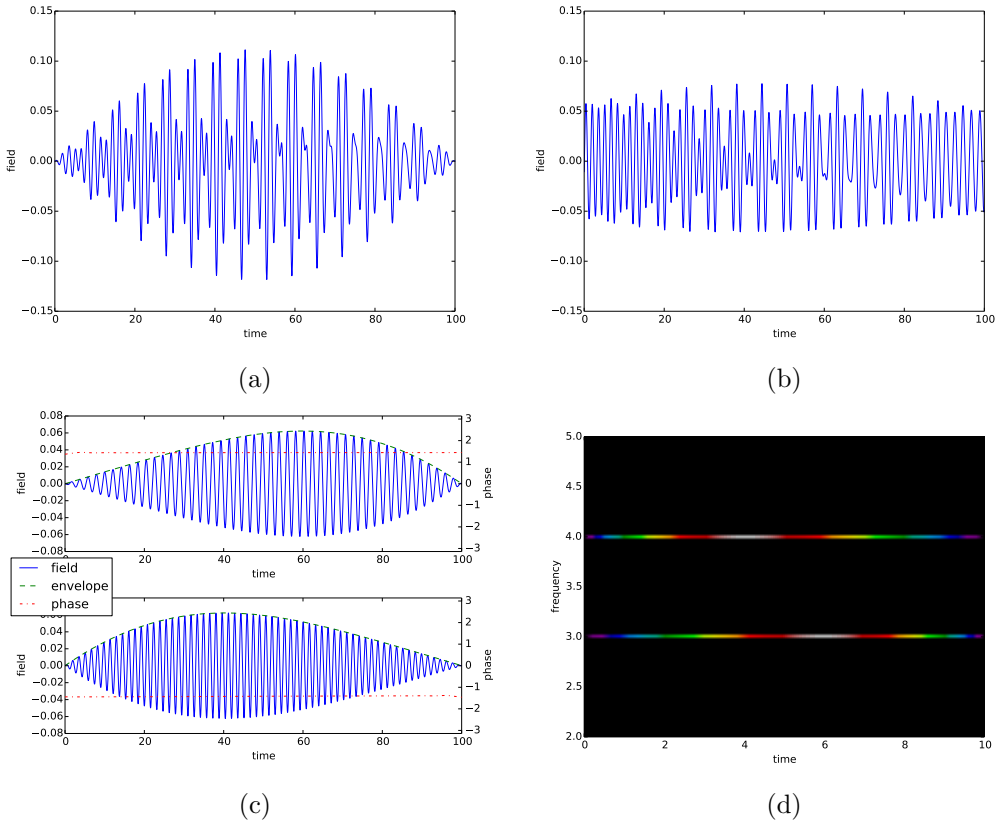


Figure 7.1: In (a) the optimal control field for the measure space case in $\mathcal{M}(\Omega; H_0^1(0,T;\mathbb{C}))$ and in (b) the optimal field for the Hilbert space case in $L^2(0,T;\mathbb{R})$ are plotted. In (c) we see the two main contributions of the field in (a). In (d) we give the absolute values of the coefficients of the optimal measure in the time-frequency plane.

smaller regularization in the former case.

In Figure 7.1a we show the control field generated by the optimal control $\bar{u}^h \in \mathcal{M}^h$. Figure 7.1b shows the optimal control field corresponding to the $L^2(0,T;\mathbb{R})$ problem. The great advantage of the measure space control is that we can decompose the field into easy components. Figure 7.1d shows the absolute values of the coefficients of the optimal measure in the time-frequency plane. We see that only two frequencies have a visible contribution. They correspond to the two transition frequencies ω_1 and ω_2. Figure 7.1c shows the decomposition of the optimal field in the two contributions of those frequencies. These two fields have a clear structure. The profile in time looks smooth and both fields are switched on during the whole time interval. This is expected due to the choice $\mathcal{U} = H_0^1(0,T;\mathbb{C})$, which promotes smoothness and non-locality in time.

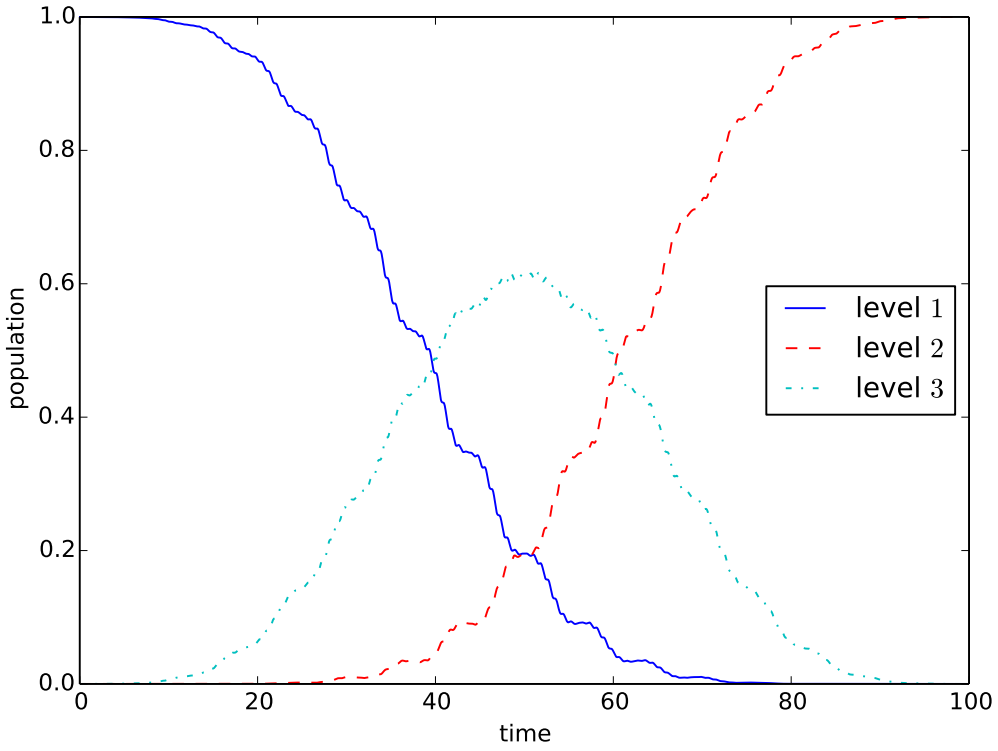


Figure 7.2: Time evolution of the populations for the first (solid), second (dashed) and third (dash-dotted) level.

The field for the first transition reaches its maximum before the field for the second transition. This corresponds to the intuitive understanding that we have to induce the transition between the levels one and three before the transition between the levels three and two. This can be observed in the plot for the time evolution of the populations of the different levels in Figure 7.2. The structure of the phase looks also very simple except for some oscillations at the boundaries. Those oscillations can be seen in Figure 7.3. Those boundary effects are not desired in applications since a time dependent phase can lead to an effective change of the frequency. These effects are probably caused by the fact that using the space $H_0^1(0,T;\mathbb{C})$ we penalize oscillations of the phase only via the natural norm in $\mathbb{C}$. This penalization for oscillations in the phase is weaker if the absolute value of the envelope is small. This is the case near the boundaries. However, experiments with simple post processing show that the influence of this effect is not large. As post processing we first set all frequency contributions in $\bar{u}^h$ to zero,

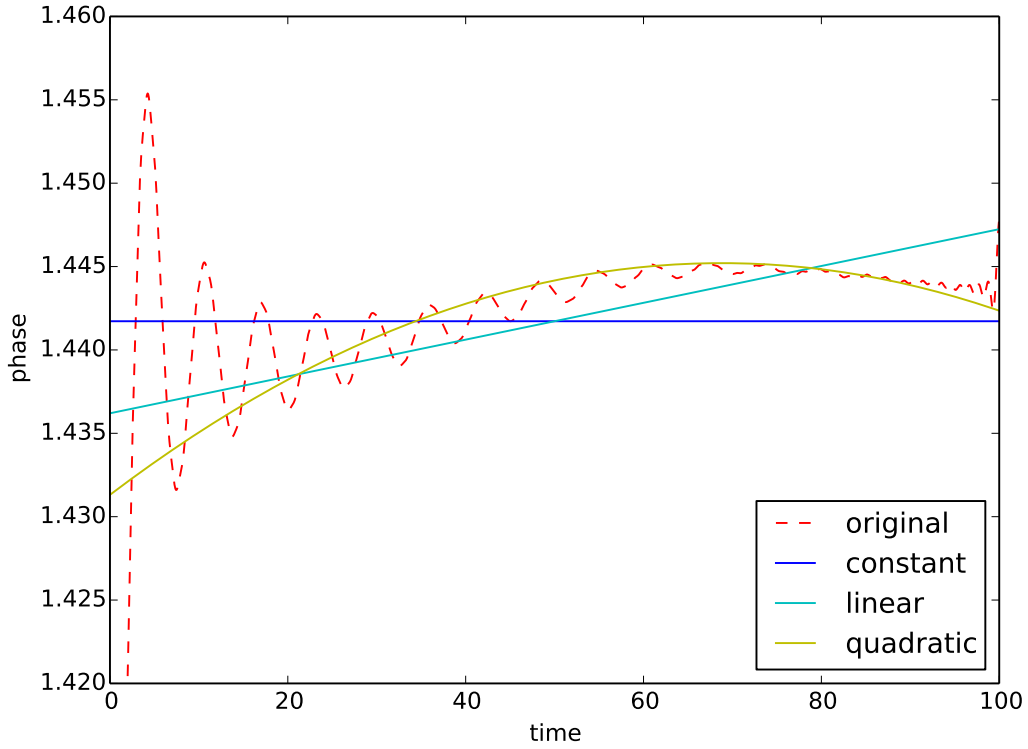
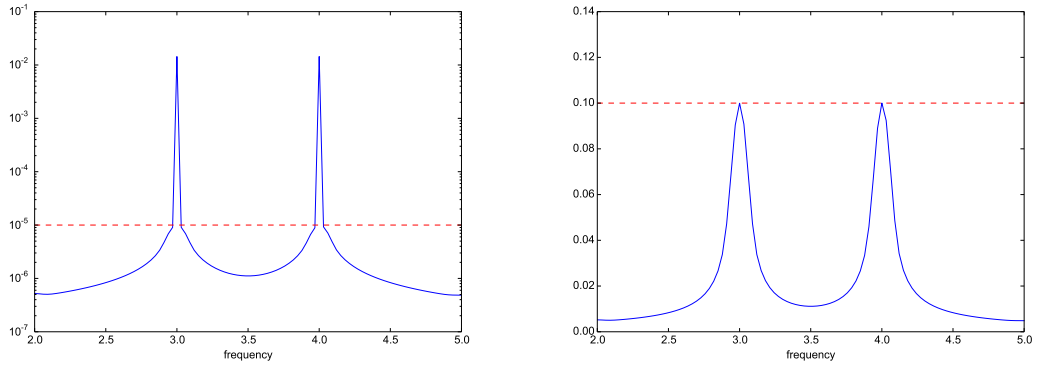


Figure 7.3: The phase of the envelope for the frequency ω_2 (dashed) and polynomial approximations of the phase for different orders (solid).

ansatz for phase	$j^{h\theta}$	f^h
original	$2.9056 * 10^{-03}$	$2.8515 * 10^{-05}$
zero	$3.1029 * 10^{-03}$	$2.2734 * 10^{-04}$
constant	$3.0224 * 10^{-03}$	$1.4676 * 10^{-04}$
linear	$2.9274 * 10^{-03}$	$5.1612 * 10^{-05}$
quadratic	$2.9215 * 10^{-03}$	$4.5579 * 10^{-05}$

Table 7.2: The value of the cost functional and expectation value after simple post-processing.



(a) (b)

Figure 7.4: In (a), the norms $\|\bar{u}^h\|_{\mathcal{U}^h}$ (solid) and the regularization parameter θ (dashed) are plotted. In (b), the norms $\|(B^{h*}\nabla f^h(B^h\bar{u}))_\omega\|_{\mathcal{U}^h}$ (solid) and the cost parameter α (dashed) are plotted.

except for the two frequencies ω_1 and ω_2. Then we approximated the phase for the two frequencies using polynomials. Approximation by the zero polynomial corresponds to neglecting the phase. Constant, linear and quadratic polynomials correspond to a phase shift, an additional constant frequency shift and an additional frequency shift linear in time, respectively. The resulting values for the discrete expectation value are given in Table 7.2. We see that we still achieve the control goal reasonably well after only using two frequencies, but there is a considerable gap to the unprocessed case. We lose another order of magnitude by neglecting the phase. Increasing the polynomial degree leads to smaller expectation values. The moderate decrease going from linear to quadratic polynomials suggests that the large phase oscillations on the boundary have a nonnegligible effect. But even in the worst case of neglicting the phase altogether, the expectation value still corresponds to a probability of more than 99.95% to achieve the control goal.

In Figure 7.4 we can see that our optimal control is consistent with the norm estimate in Proposition 12 and the support condition in Proposition 13. We plotted $\|B^{h*}\nabla f^h(B^h\bar{u}^h)\|_{\mathcal{U}^h}$ in Figure 7.4b. We see that $\|B^{h*}\nabla f^h(B^h\bar{u}^h)\|_{\mathcal{U}^h}$ is below α for all frequencies. In Figure 7.4a we plotted the values of the norms $\|\bar{u}^h_\omega\|_{\mathcal{U}^h}$. One can see that $\|\bar{u}^h_\omega\|_{\mathcal{U}^h}$ is large for ω with $\|(B^{h*}\nabla f^h(B^h\bar{u}^h))_\omega\|_{\mathcal{U}^h} = \alpha$. But for $\|(B^{h*}\nabla f^h(B^h\bar{u}^h))_\omega\|_{\mathcal{U}^h} < \alpha$ we do not have $\|\bar{u}^h_\omega\|_{\mathcal{U}^h} = 0$. This is an artifact of Huber-regularization. As in Proposition 12 we obtain the norm estimate

$$\|(B^*\nabla f^h(B^h\bar{u}^h))_\omega\|_{\mathcal{U}^h} \leq \alpha$$

for all $\omega \in \Omega$. But instead of the support conditions in Proposition 13 we have

$$\|(B^*\nabla f^h(B^h\bar{u}^h))_\omega\|_{\mathcal{U}^h} < \alpha \Rightarrow \|\bar{u}^\theta_\omega\|_{\mathcal{U}^h} < \theta$$

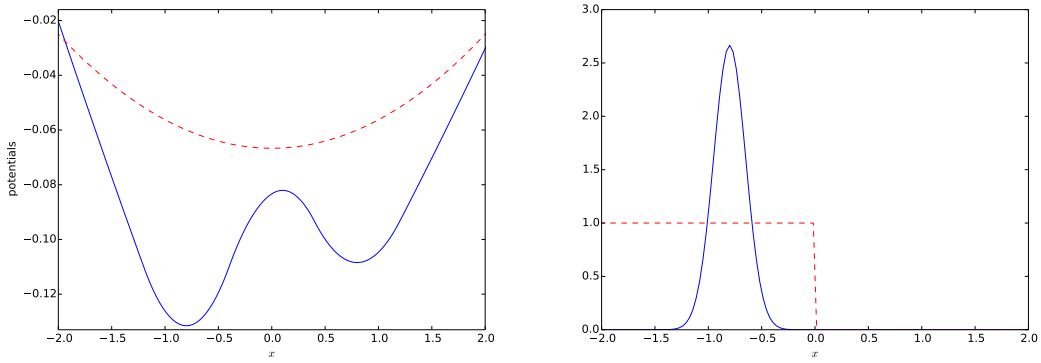
for all $\omega \in \Omega$. This follows from Lemma 30 together with the optimality condition $\nabla j^{h\theta} = 0$. Those two conditions are exactly what we see in Figure 7.4. In slight abuse of notation we will, therefore, refer to the super level set $\{\, \omega \in \Omega^h \mid \|u^h_\omega\|_{\mathcal{U}^h} > \theta \,\}$ as the support of u^h. In this example, we obtain a small support of the desired size two and a very small expectation value $7.8 * 10^{-7}$ for a cost parameter $\alpha = 10^{-1}$. For the second example we will see that a small support and a small expectation value are conflicting goals, which makes the choice of a good parameter α more important.

7.2 One-Dimensional Schrödinger Equation on Two Surfaces

As second example to test our framework, we consider the optimal control of a system of one-dimensional Schrödinger equations on two potential energy surfaces. In this example, much more frequencies are potentially useful in the control. We also expect an additional time structure in the controls due to the movement of the densities in space. Therefore it is much more challenging to obtain simple controls for this example. We focus on a comparison between controls generated for different choices of Ω, $\mathcal{U}$ and B.

We consider the model quantum system presented in Example 2 with $d = 1$ and $M = 2$. This toy problem resembles a 1D version of the problem studied in [Kos+89] and has also been used in [HL15]. We assume the particle mass of a proton m_p. All quantities are given in atomic units. The drift Hamiltonian is given by

$$H_0 = \begin{pmatrix} -\frac{1}{2m_p}\partial_x^2 & 0 \\ 0 & -\frac{1}{2m_p}\partial_x^2 \end{pmatrix} + \begin{pmatrix} V_1 & 0 \\ 0 & V_2 \end{pmatrix}.$$



(a) (b)

Figure 7.5: In (a), the two potential energy surfaces. In (b), the initial state (solid) and the coefficients for the observable (dashed).

Figure 7.6

The potential energy surfaces V_1 and V_2 are plotted in Figure 7.5a. The coupling Hamiltonian is given by

$$H_1 = \begin{pmatrix} x & I \\ I & x \end{pmatrix}$$

with the dipole operator on the diagonal and constant time independent coupling on the off-diagonal. The state Hilbert space is given by $\mathcal{H} = L^2([-4, 4]; \mathbb{C}^2)$. We choose the spatial domain to be bounded, which makes H_1 a bounded operator. The initial state ψ_0 is given by a Gaussian located in the potential lower well on the left, depicted in Figure 7.5b. The control objective is to reach the potential well on the right starting from the potential well on the left. This is modeled by the observation operator $\mathcal{O}$, which is the projection on the complement of functions with support on the lower energy surface to the right of the potential barrier. This observable is a multiplication operator. The corresponding function is equal to one on the upper surface and is plotted in Figure 7.5b on the lower surface. The energy differences between the two potential energy surfaces measured at the local minima of the lower surface are around 0.074 and 0.048. Considering the relation between energy differences and transition frequencies, we expect the control to contain large frequency contributions around the two frequencies $\omega_1 \approx 0.071$ and $\omega_2 \approx 0.048$. We choose a time horizon of $T = 3000$. The time horizon is large enough to allow for sufficiently many oscillations with the transition frequencies ω_1 and ω_2, and for sufficient movement of the wave packet in space.

We discretize the state space $\mathcal{H}$ by a nodal basis on a uniform grid with 256 points.

The discrete drift Hamiltonian H_0^h is given by a simple finite difference stencil and a pointwise multiplication operator at the grid points, assuming homogeneous Dirichlet boundary conditions. The coupling Hamiltonian H_1^h is also a pointwise multiplication operator at the grid points. For the time-discretization, we define a uniform grid $(t_n)_{n=0}^N$ with $N = 2048$ points. For the discrete control fields $\mathcal{E}^h$, we use linear finite elements on this grid. For the time stepping we again use the GST method of order $m = 4$ based on a linear approximation and a Magnus expansion using the mid point rule.

For this problem we compare the resulting optimal controls for different choices of Ω and $\mathcal{U}$ and B. The different choices correspond to the Examples 3 through 7. We will also compare the resulting optimal fields to the optimal fields for the classical Hilbert space regularization with $L^2(0,T;\mathbb{R})$ or $H_0^1(0,T;\mathbb{R})$ norm. In particular, we study the following setups.

Example 3. $\Omega = [1/30, 1/10]$, $\mathcal{U} = H_0^1(0,T;\mathbb{C})$, B given by (3.8): The sparsity domain Ω is chosen such that it contains the expected transition frequencies ω_1 and ω_2. It is discretized with a uniform grid of 100 frequencies. We discretize $\mathcal{U}$ with complex-valued linear finite elements. The time grid for $\mathcal{U}^h$ corresponds to the grid of the time stepping and $\mathcal{E}^h$. This results in $100 \cdot 2 \cdot 2048 = 409600$ real degrees of freedom. The discrete control operator is discretized as in (6.3) with $K^h = I$ and Λ_ω^h being a pointwise multiplication of the basis coefficients with $(e^{i\omega t_n})_n$.

Example 4. $\Omega = [1/30, 1/10]$, $\mathcal{U} = L^2(0,T;\mathbb{C})$, B given by (3.10) with a Gaussian kernel, suitably adapted to generate homogeneous Dirichlet boundary conditions: We discretize $\Omega, \mathcal{U}$ and B as before. For the evaluation of B^h we explicitly construct the dense matrix K^h. As above, this results in $100 \cdot 2 \cdot 2048 = 409600$ real degrees of freedom.

Example 6. $\Omega = [1/30, 1/10]$, $\mathcal{U} = \mathbb{C}$, B given by (3.14): Again, the frequency band Ω is discretized with 100 grid points. There is no discretization necessary for $\mathcal{U}$. This results in $2 \cdot 100 = 200$ real degrees of freedom.

Example 7. $\Omega = [1/30, 1/10] \times [0,T]$, $\mathcal{U} = \mathbb{C}$, B given by (3.15): The time-frequency plane Ω is discretized by a tensor grid. In frequency direction we use a grid with 100 grid points. In time direction we use a grid of 14 points. This results in $100 \cdot 14 \cdot 2 = 2800$ real degrees of freedom.

As a reference we also compute solutions for the Hilbert space cases $L^2(0,T;\mathbb{R})$ and $H_0^1(0,T;\mathbb{R})$. For both cases we use linear finite elements on the time grid of the time stepping method. This results in 2048 real degrees of freedom.

For the simulations we chose α such that the probability to end up in the desired subspace is near 95%. In Figure 7.7 we plotted the fields generated by the optimal controls, their Fourier coefficients, time-frequency representation and the absolute values of the coefficients of the optimal measure arranged in the time-frequency plane. The time evolution of the populations of the two surfaces as well as the evolution of the expectation value are given in Figure 7.8. In Figure 7.9, we see snapshots of the time evolution of the densities of the state. As a first observation we see that the structure of optimal controls and the resulting influence on the quantum system heavily depend on the different cost functionals. This flexibility of our framework should be exploited in applications. In the following we will compare the optimal controls for the different control setups.

From the second and third columns of Figure 7.7 we see that all the fields, except for the $H_0^1(0,T;\mathbb{R})$ field, have two frequency regions around ω_1 and ω_2 that stand out. The regions correspond to the transitions up from the first well and down into the second well, respectively. This is desired and expected behavior. Interestingly, the frequency for the first transition is a little smaller than predicted. The $H_0^1(0,T;\mathbb{R})$ field, on the other hand, mostly has low frequency components. In this case, almost no transition of the state to the second surface is induced, as can be seen in Figure 7.8f. Closer examination of the time evolution shows that the state instead moves back and forth in the lower well until it is forced over the potential barrier into the second well. This shows that the choice of the control cost term can have an influence on the control mechanism triggered by the field. Since for this example the transitions between surfaces are of chemical interest, the use of the $H_0^1(0,T;\mathbb{R})$ norm as a cost leads to undesired control fields.

In the third column of Figure 7.7, we see that the other controls, with exception of the control for Example 6, all show an alternating time structure in the two important frequency regions. It corresponds to the expected pump-dump structure, which means that we first induce the transition up from the lower well with frequency ω_1 and then down into the upper well with frequency ω_2. This can also be seen in Figure 7.8. The time lag between the peaks for the frequencies ω_1 and ω_2 corresponds to the time it takes the wave package to travel an the upper surface between the positions of the two wells on the lower surface. The field from Example 7 corresponds closest to a pump-dump field of two pulses separated in time and frequency. The other controls induce transitions distributed throughout the whole time interval. For those cases, it is important to choose a suitable time horizon T. For the horizon we chose, we see that the controls from Example 3 and the L^2 control basically perform two pump-dump sequences. For the control for Example 4, we also see a hint of two sequences, with the second sequence

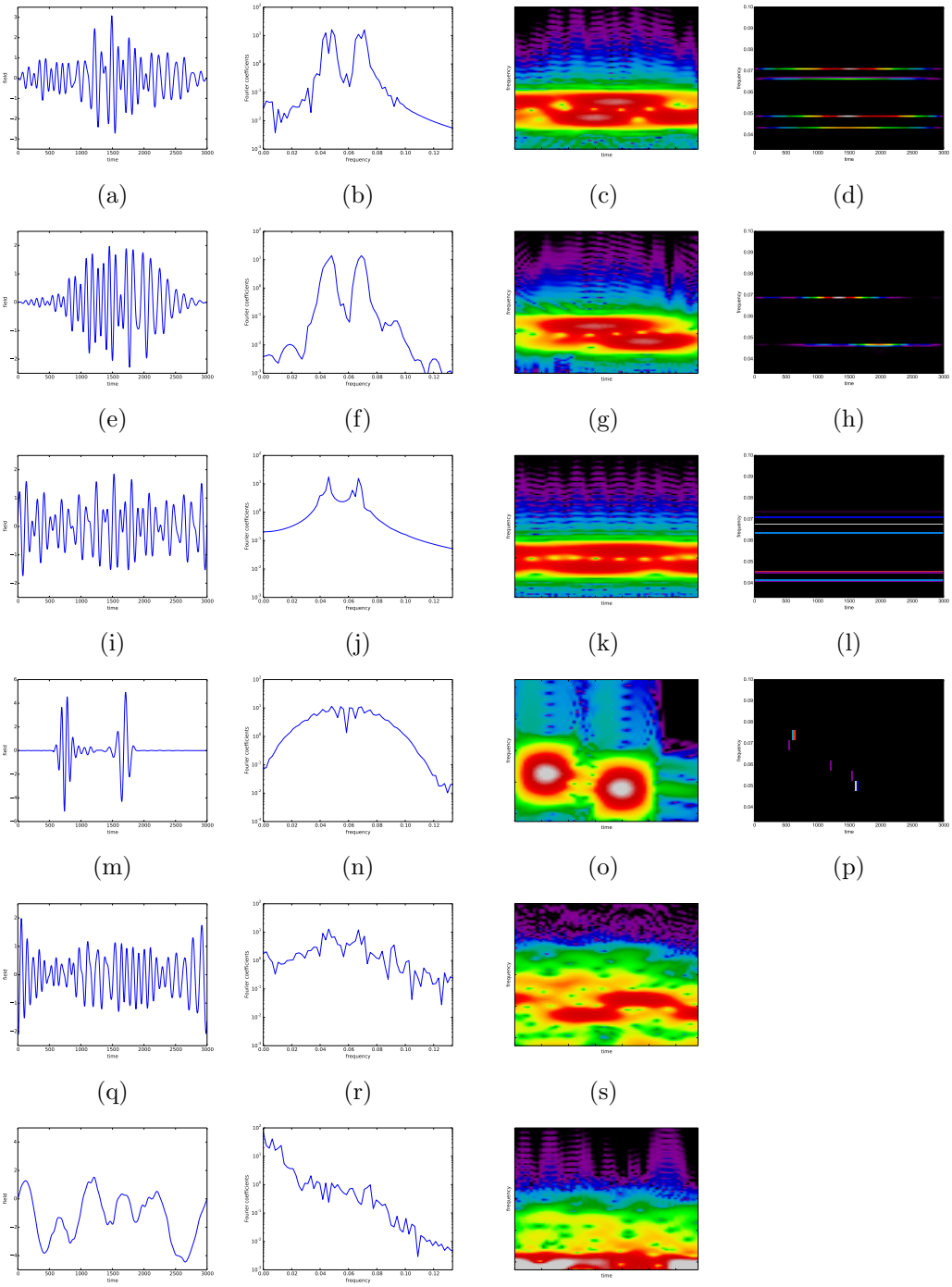
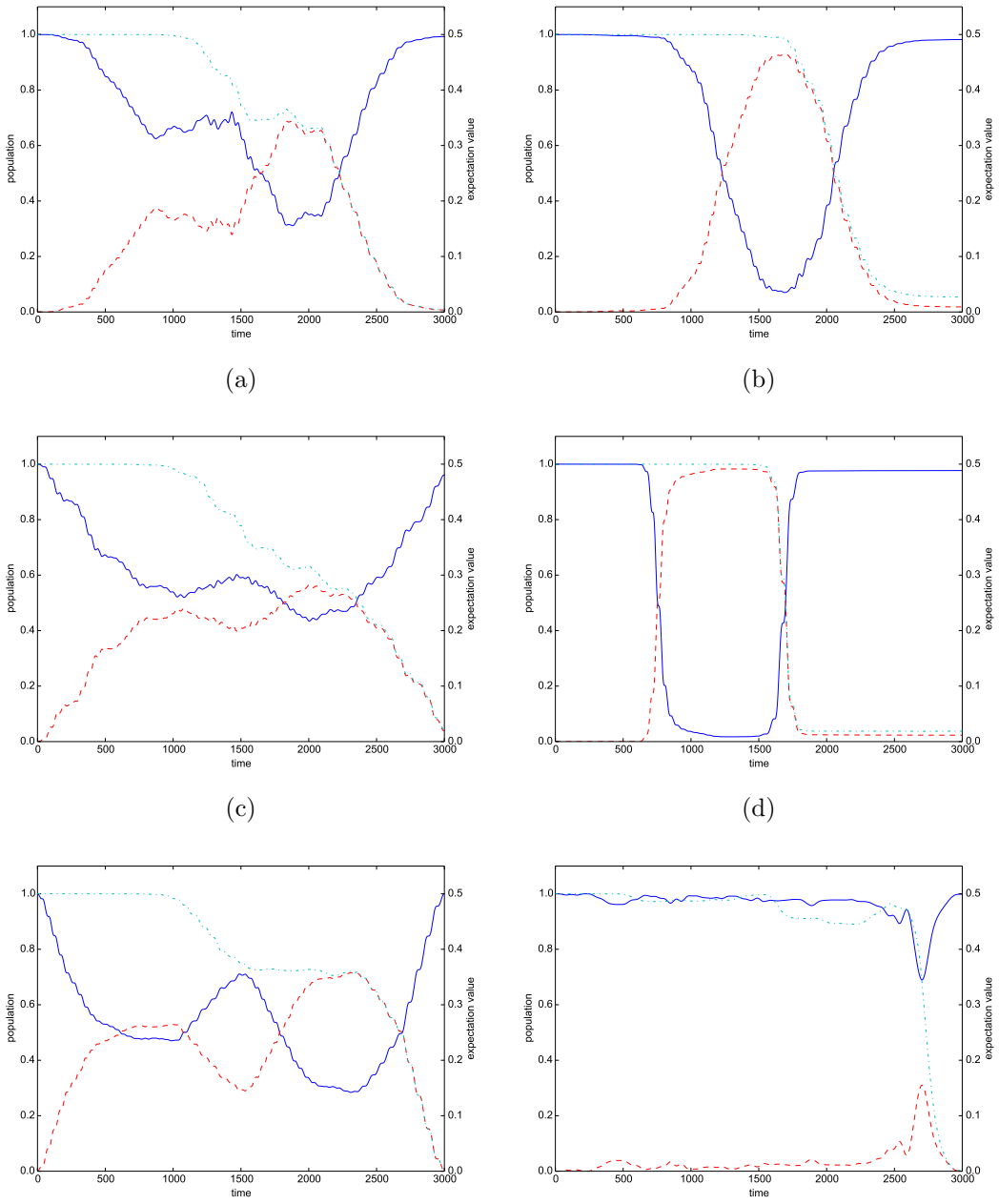


Figure 7.7: We plotted the different control fields in different representations. The rows, from top to bottom, correspond to the Examples 3, 4, 6 and 7, and the Hilbert space cases in $L^2(0,T;\mathbb{R})$ and $H_0^1(0,T;\mathbb{R})$. The columns, from left to right, correspond to the time, frequency and time-freqency representation. In the rightmost column, the absolute values of the optimal measures are plotted in the time-frequency plane.



(a) (b)

(c) (d)

(e) (f)

Figure 7.8: The evolution of the populations on the lower (solid) and upper (dashed) surface, and the evolution of the expectation value (dashdotted) for the different control spaces from Example 3 in (a), Example 4 in (b), Example 6 in (c), Example 7 in (d), $L^2(0,T;\mathbb{R})$ in (e) and $H_0^1(0,T;\mathbb{R})$ in (f).

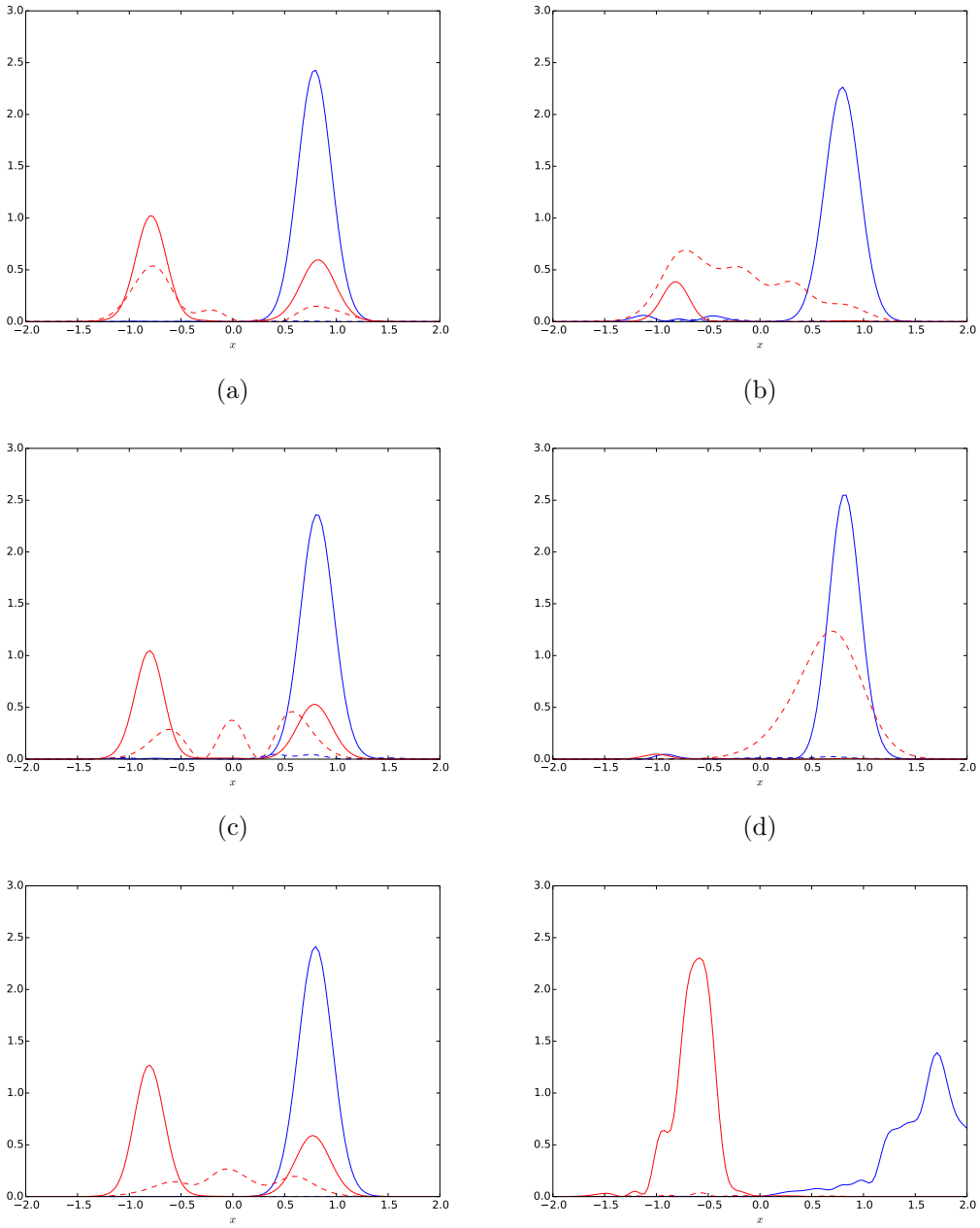


Figure 7.9: Snapshots of the densities $|\bar{\psi}(t)|$ on the lower (solid) and upper (dashed) surface at the times $t = T/2$ (blue) and $t = T$ (red) for the different control spaces as in Figure 7.8.

being much more pronounced. The control from Example 7 is the only approach that really only uses one sequence. It might depend the least on variations of the final time T, which might be beneficial if an appropriate final time is not known a priori.

The controls generated using the measure space setting all show a localized time-frequency structure. In particular, with exception of the control for Example 7, the Fourier coefficients and the time-frequency representation decay faster than for the Hilbert space controls. In comparison, the time-frequency representation of the $L^2(0,T;\mathbb{R})$ control looks very complicated and difficult to analyze. It also contains unwanted low frequency components. For the control from Example 7, we see a difficult to analyze frequency spectrum. This is probably do to the fact that the control field consists of two main pulses that are very localized in time. This localization in time leads to a more difficult frequency representation. It is interesting that the space $\mathcal{M}(\Omega; H_0^1(0,T;\mathbb{C}))$ leads to more active frequencies compared to the approach with space $\mathcal{M}(\Omega; L^2(0,T;\mathbb{C}))$ with the smoothing control operator. In the former case we see that the frequencies ω_1 and ω_2 do not have a visible contribution. Instead we have contributions shifted up and down around those two frequencies. This might be an artifact of forcing the envelope to have a small $H_0^1(0,T;\mathbb{C})$ norm in a case where more time structure in the control is necessary. The symmetric shift up and down then might exploit an effect similar to the equality $\cos((\omega+\Delta\omega)t) + \cos((\omega-\Delta\omega)t) = 2\cos(\Delta\omega t)\cos(\omega t)$ for the superposition of different frequencies resulting in a modulated middle frequency.

We will now take a closer look at the support of the different optimal controls. In Figure 7.10 we plotted the achievement of the control goal against the size of the support for different control spaces for varying cost parameter α. The vertical line marks the value $2.5 * 10^{-2}$ of the expectation value that corresponds to a 95% probability of ending up in the desired subspace. The figure illustrates the conflicting goals of minimizing the cost term, resulting in small supports, versus minimizing the expectation value. In contrast to the first example, we do not obtain almost perfect achievement of the control goal for the desired support size two. Using the control spaces from Examples 3 and 4 leads to good achievement of the control goal for a relatively small support. This is most likely due to the fact that those approaches separate frequency and time degrees of freedom. In frequency direction sparsity is favored, while we still have enough flexibility in the time direction. For Example 4, we achieve a remarkably low number of three contributing frequencies. Two of those frequencies lie on neighboring grid points, which could well be a consequence of coarse discretization. If we increase the cost parameter to a point where the size of the support actually is forced to become two, we see that that control goal is not achieved in a satisfactory way anymore. Examples 6 and 7 lead to

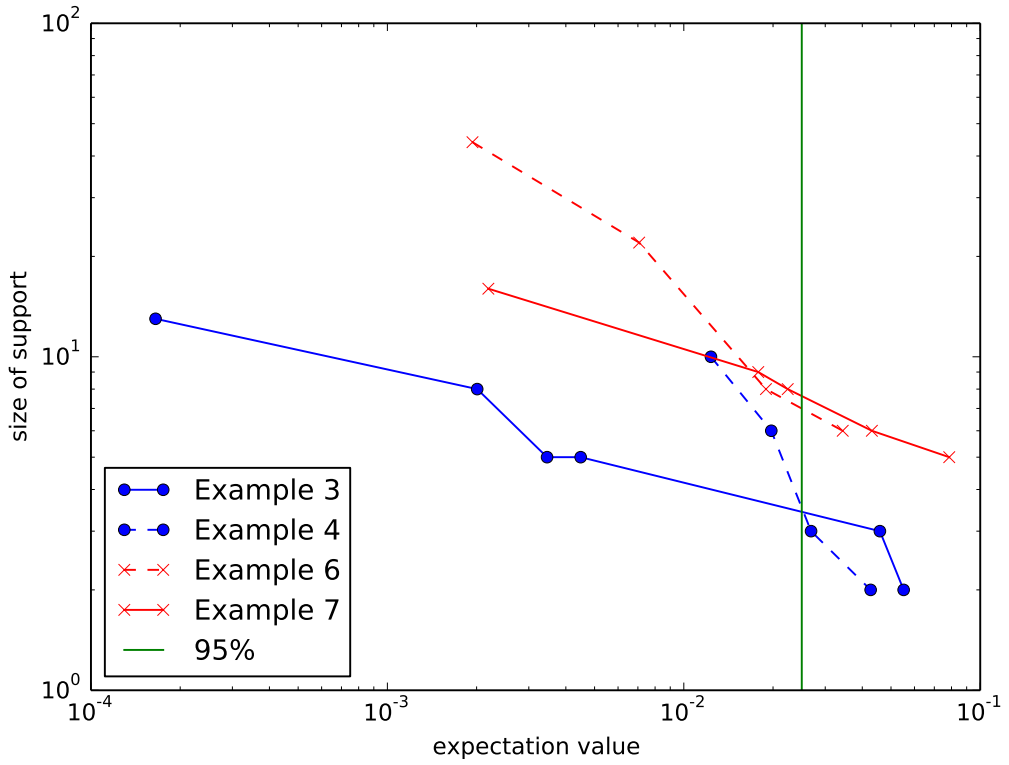


Figure 7.10: For different control spaces and cost terms, the size of the support of optimal measures is plotted against the corresponding expectation value. The vertical line marks the expectation value corresponding to a probability of 95% to achieve the control goal.

larger but still reasonably small supports. This is remarkable because for those examples the size of the support directly corresponds to the number of used degrees of freedom. Both approaches need way less degrees of freedom compared to the other approaches. We are, however, more interested in an easy frequency structure than in a low number of degrees of freedom. This is because the size of the support corresponds to degrees of freedom responsible for the oscillations that make the control fields difficult to analyze and realize in experiments. The controls for Example 7 does not exploit a separation of time and frequency structure. Since the shape of the pulses is fixed beforehand, it cannot adapt to the problem, leading to additional active coefficients. The additional frequencies in the control for Example 6 are probably necessary to obtain additional time structure in the resulting field.

Bibliography

[AB06] C. D. Aliprantis and K. C. Border. *Infinite dimensional analysis*. Third. A hitchhiker's guide. Springer, Berlin, 2006, pp. xxii+703.

[Abe+13] A. Abedi et al. "Dynamical Steps that Bridge Piecewise Adiabatic Shapes in the Exact Time-Dependent Potential Energy Surface." In: *Phys. Rev. Lett.* 110 (26 June 2013), p. 263001.

[AMG10] A. Abedi, N. T. Maitra, and E. K. U. Gross. "Exact Factorization of the Time-Dependent Electron-Nuclear Wave Function." In: *Phys. Rev. Lett.* 105 (12 Sept. 2010), p. 123002.

[Ass+13] E. Assémat et al. "On the application of geometric optimal control theory to nuclear magnetic resonance." In: *Math. Control Relat. Fields* 3.4 (2013), pp. 375–396.

[Bal+05] G. G. Balint-Kurti et al. "Quantum control of molecular motion including electronic polarization effects with a two-stage toolkit." In: *J. Chem. Phys.* 122.8, 084110 (2005), p. 084110.

[BCR10] C. Brif, R. Chakrabarti, and H. Rabitz. "Control of quantum phenomena: past, present and future." In: *New J. Phys.* 12.7 (2010), p. 075008.

[BEH08] J. Blank, P. Exner, and M. Havlíček. *Hilbert space operators in quantum physics*. Second. Theoretical and Mathematical Physics. Springer, New York; AIP Press, New York, 2008, pp. xviii+664.

[BKT01] A. Bartana, R. Kosloff, and D. J. Tannor. "Laser cooling of molecules by dynamically trapped states." In: *Chem. Phys.* 267.13 (2001), pp. 195–207.

[Bla+09] S. Blanes et al. "The Magnus expansion and some of its applications." In: *Phys. Rep.* 470.5-6 (2009), pp. 151–238.

[BMS82] J. M. Ball, J. E. Marsden, and M. Slemrod. "Controllability for distributed bilinear systems." In: *SIAM J. Control Optim.* 20.4 (1982), pp. 575–597.

[Bos+02] U. Boscain et al. "Optimal control in laser-induced population transfer for two- and three-level quantum systems." In: *J. Math. Phys.* 43.5 (2002), pp. 2107–2132.

[Bos+12] U. Boscain et al. "A weak spectral condition for the controllability of the bilinear Schrödinger equation with application to the control of a rotating planar molecule." In: *Comm. Math. Phys.* 311.2 (2012), pp. 423–455.

[Bou65] N. Bourbaki. *Éléments de mathématique. Fasc. XIII. Livre VI: Intégration. Chapitres 1, 2, 3 et 4: Inégalités de convexité, Espaces de Riesz, Mesures sur les espaces localement compacts, Prolongement d'une mesure, Espaces L^p*. Deuxième édition revue et augmentée. Actualités Scientifiques et Industrielles, No. 1175. Hermann, Paris, 1965, p. 283.

[BZB08] G. G. Balint-Kurti, S. Zou, and A. Brown. "Optimal Control Theory for Manipulating Molecular Processes." In: *Advances in Chemical Physics*. John Wiley & Sons, Inc., 2008, pp. 43–94.

[Cam76] M. Cambern. "Isomorphisms of spaces of continuous vector-valued functions." In: *Illinois J. Math.* 20.1 (1976), pp. 1–11.

[CB06] T. Cheng and A. Brown. "Pulse shaping for optimal control of molecular processes." In: *J. Chem. Phys.* 124.14, 144109 (2006),

[CCM11] T. Caneva, T. Calarco, and S. Montangero. "Chopped random-basis quantum optimization." In: *Phys. Rev. A* 84 (2 Aug. 2011), p. 022326.

[Cha12] T. Chambrion. "Periodic excitations of bilinear quantum systems." In: *Automatica* 48.9 (2012), pp. 2040–2046.

[CHW12] E. Casas, R. Herzog, and G. Wachsmuth. "Optimality Conditions and Error Analysis of Semilinear Elliptic Control Problems with L^1 Cost Functional." In: *SIAM J. Optim.* 22.3 (2012), pp. 795–820.

[Cla90] F. H. Clarke. *Optimization and nonsmooth analysis*. Second. Vol. 5. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1990, pp. xii+308.

[DAl08] D. D'Alessandro. *Introduction to quantum control and dynamics*. Chapman & Hall/CRC Applied Mathematics and Nonlinear Science Series. Chapman & Hall/CRC, Boca Raton, FL, 2008, pp. xiv+343.

[Dio+02] C. M. Dion et al. "Optimal laser control of orientation: The kicked molecule." In: *Phys. Rev. A* 65 (6 June 2002), p. 063408.

[DU77] J. Diestel and J. J. Uhl Jr. *Vector measures*. With a foreword by B. J. Pettis, Mathematical Surveys, No. 15. American Mathematical Society, Providence, R.I., 1977, pp. xiii+322.

[ET99] I. Ekeland and R. Témam. *Convex analysis and variational problems*. English. Vol. 28. Classics in Applied Mathematics. Translated from the French. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1999, pp. xiv+402.

[Fec+07] S. Fechner et al. "The von Neumann picture: a new representation for ultrashort laser pulses." In: *Opt. Express* 15.23 (Nov. 2007), pp. 15387–15401.

[FHK15] G. Friesecke, F. Henneke, and P. Kunisch. *Sparse control of quantum systems*. pre-print. arXiv:1507.00768. 2015.

[FR08] M. Fornasier and H. Rauhut. "Recovery algorithms for vector-valued data with joint sparsity constraints." In: *SIAM J. Numer. Anal.* 46.2 (2008), pp. 577–613.

[FZC14] B. Feng, D. Zhao, and P. Chen. "Optimal bilinear control of nonlinear Schrödinger equations with singular potentials." In: *Nonlinear Anal.* 107 (2014), pp. 12–21.

[GL08] R. Griesse and D. A. Lorenz. "A semismooth Newton method for Tikhonov functionals with sparsity constraints." In: *Inverse Probl.* 24.3 (2008), p. 035007.

[GNR92] P. Gross, D. Neuhauser, and H. Rabitz. "Optimal control of curve-crossing systems." In: *J. Chem. Phys.* 96.4 (1992), pp. 2834–2845.

[Grö01] K. Gröchenig. *Foundations of time-frequency analysis*. Applied and Numerical Harmonic Analysis. Birkhäuser Boston, Inc., Boston, MA, 2001, pp. xvi+359.

[Hin+13] M. Hintermüller et al. "Optimal bilinear control of Gross-Pitaevskii equations." In: *SIAM J. Control Optim.* 51.3 (2013), pp. 2509–2543.

[HL03] M. Hochbruck and C. Lubich. "On Magnus integrators for time-dependent Schrödinger equations." In: *SIAM J. Numer. Anal.* 41.3 (2003), 945–963 (electronic).

[HL15] F. Henneke and M. Liebmann. *A Generalized Suzuki–Trotter Type Method in Optimal Control of Coupled Schrödinger Equations*. pre-print. `http://igdk1754.ma.tum.de/foswiki/pub/IGDK1754/Preprints/HennekeLiebmann_2015.pdf`. 2015.

[Hof+12] P. v. d. Hoff et al. “Optimal control theory - closing the gap between theory and experiment.” In: *Phys. Chem. Chem. Phys.* 14 (42 2012), pp. 14460–14485.

[Hoh+07] U. Hohenester et al. “Optimal quantum control of Bose–Einstein condensates in magnetic microtraps.” In: *Phys. Rev. A* 75 (2 Feb. 2007), p. 023602.

[HSW12] R. Herzog, G. Stadler, and G. Wachsmuth. “Directional sparsity in optimal control of partial differential equations.” In: *SIAM J. Control Optim.* 50.2 (2012), pp. 943–963.

[HTC83] G. M. Huang, T. J. Tarn, and J. W. Clark. “On the controllability of quantum-mechanical systems.” In: *J. Math. Phys.* 24.11 (1983), pp. 2608–2618.

[IK07] K. Ito and K. Kunisch. “Optimal Bilinear Control of an Abstract Schrödinger Equation.” In: *SIAM J. Control Optim.* 46.1 (2007), pp. 274–287.

[IK09] K. Ito and K. Kunisch. “Asymptotic Properties of Feedback Solutions for a Class of Quantum Control Problems.” In: *SIAM J. Control Optim.* 48.4 (2009), pp. 2323–2343.

[IK98] K. Ito and F. Kappel. “The Trotter–Kato theorem and approximation of PDEs.” In: *Math. Comp* 67 (1998), pp. 21–44.

[KHK10] K. Kormann, S. Holmgren, and H. Karlsson. “A Fourier-Coefficient Based Solution of an Optimal Control Problem in Quantum Chemistry.” English. In: *J. Optim. Theory Appl.* 147 (3 2010), pp. 491–506.

[KN08] A. Kondorskiy and H. Nakamura. “Semiclassical guided optimal control of molecular processes of many degrees of freedom.” In: *Phys. Rev. A* 77 (4 Apr. 2008), p. 043407.

[Kos+89] R. Kosloff et al. “Wavepacket dancing: Achieving chemical selectivity by shaping light pulses.” In: *Chem. Phys.* 139.1 (1989), pp. 201–220.

[KPV14] K. Kunisch, K. Pieper, and B. Vexler. “Measure valued directional sparsity for parabolic optimal control problems.” In: *SIAM J. Control Optim.* 52.5 (2014), pp. 3078–3108.

[KW13] K. Kunisch and D. Wachsmuth. “On Time Optimal Control of the Wave Equation and Its Numerical Realization as Parametric Optimization Problem.” In: *SIAM J. Control Optim.* 51.2 (2013), pp. 1232–1262.

[Lan93] S. Lang. *Real and functional analysis*. Third. Vol. 142. Graduate Texts in Mathematics. Springer-Verlag, New York, 1993, pp. xiv+580.

[Lap+09] M. Lapert et al. "Monotonically convergent optimal control theory of quantum systems with spectral constraints on the control field." English. In: *Phys. Rev. A* 79 (2009), p. 063411.

[Lie00] M. Liebmann. "Ein effizienter Algorithmus zur numerischen Simulation von zeitabhängigen Problemen aus der Quantenmechanik." Diploma thesis. University of Graz, Institute for Mathematics and Scientific Computing, 2000.

[Lub08] C. Lubich. *From quantum to classical molecular dynamics: reduced models and numerical analysis*. Zurich Lectures in Advanced Mathematics. European Mathematical Society (EMS), Zürich, 2008, pp. x+144.

[LY95] X. J. Li and J. M. Yong. *Optimal control theory for infinite-dimensional systems*. Systems & Control: Foundations & Applications. Boston, MA: Birkhäuser Boston Inc., 1995, pp. xii+448.

[Mez09] L. Meziani. "On the dual space $C_0^*(S, X)$." In: *Acta Math. Univ. Comenian. (N.S.)* 78.1 (2009), pp. 153–160.

[MWC10] A. Monmayrant, S. Weber, and B. Chatel. "A newcomer's guide to ultrashort pulse shaping and characterization." In: *J. Phys. B At. Mol. Opt. Phys.* 43.10 (2010), p. 103001.

[Nak12] H. Nakamura. *Nonadiabatic Transition: Concepts, Basic Theories and Applications*. Second. World Scientific, Singapore, 2012.

[NH95] I. Najfeld and T. Havel. "Derivatives of the Matrix Exponential and Their Computation." In: *Adv. Appl. Math.* 16.3 (1995), pp. 321–375.

[NW06] J. Nocedal and S. J. Wright. *Numerical optimization*. Second. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2006, pp. xxii+664.

[PDR88] A. P. Peirce, M. A. Dahleh, and H. Rabitz. "Optimal control of quantum-mechanical systems: existence, numerical approximation, and applications." In: *Phys. Rev. A (3)* 37.12 (1988), pp. 4950–4964.

[Pie15] K. Pieper. "Finite element discretization and efficient numerical solution of elliptic and parabolic sparse control problems." PhD thesis. Technische Universität München, 2015.

[QG10] B. Qi and L. Guo. "Is measurement-based feedback still better for quantum control systems?" In: *Syst. Control Lett.* 59.6 (2010), pp. 333–339.

[Rab37] I. I. Rabi. "Space Quantization in a Gyrating Magnetic Field." In: *Phys. Rev.* 51 (8 Apr. 1937), pp. 652–654.

[Ray97] M. Raydan. "The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem." In: *SIAM J. Optim.* 7.1 (1997), pp. 26–33.

[Ren+06] Q. Ren et al. "Quantum control of molecular vibrational and rotational excitations in a homonuclear diatomic molecule: A full three-dimensional treatment with polarization forces." In: *J. Chem. Phys.* 124.1, 014111 (2006), p. 014111.

[Roc76] R. T. Rockafellar. "Monotone operators and the proximal point algorithm." In: *SIAM J. Control Optim.* 14.5 (1976), pp. 877–898.

[Rue+11] S. Ruetzel et al. "Adaptive coherent control using the von Neumann basis." In: *Phys. Chem. Chem. Phys.* 13 (19 2011), pp. 8627–8636.

[RWP09] R. Roloff, M. Wenin, and W. Pötz. "Optimal control for open quantum systems: Qubits and quantum gates." In: *J. Comput. Theor. Nanosci.* 6.8 (2009), pp. 1837–1863.

[Sch+14] J. Scheuer et al. "Precise qubit control beyond the rotating wave approximation." In: *New J. Phys.* 16.9 (2014), p. 093022.

[SSB10] S. Sharma, H. Singh, and G. G. Balint-Kurti. "Genetic algorithm optimization of laser pulses for molecular quantum state excitation." In: *J. Chem. Phys.* 132.6, 064108 (2010), p. 064108.

[Sta09] G. Stadler. "Elliptic optimal control problems with L^1-control cost and applications for the placement of control devices." In: *Comput. Optim. Appl.* 44.2 (2009), pp. 159–181.

[Suz90] M. Suzuki. "Fractal decomposition of exponential operators with applications to many-body theories and Monte Carlo simulations." In: *Phys. Lett. A* 146.6 (1990), pp. 319–323.

[Suz91] M. Suzuki. "General theory of fractal path integrals with applications to many-body theories and statistical physics." In: *J. Math. Phys.* 32.2 (1991), pp. 400–407.

[Teu03] S. Teufel. *Adiabatic perturbation theory in quantum dynamics*. Vol. 1821. Lecture Notes in Mathematics. Springer-Verlag, Berlin, 2003, pp. vi+236.

[TLR04] G. Turinici, C. Le Bris, and H. Rabitz. "Efficient algorithms for the laboratory discovery of optimal quantum controls." In: *Phys. Rev. E* 70 (1 July 2004), p. 016704.

[TR03] G. Turinici and H. Rabitz. "Wavefunction controllability for finite-dimensional bilinear quantum systems." In: *J. Phys. A* 36.10 (2003), pp. 2565–2576.

[Trö10] F. Tröltzsch. *Optimal control of partial differential equations*. Vol. 112. Graduate Studies in Mathematics. Theory, methods and applications, Translated from the 2005 German original by Jürgen Sprekels. Providence, RI: American Mathematical Society, 2010, pp. xvi+399.

[Tur07] G. Turinici. "Beyond bilinear controllability: applications to quantum control." In: *Control of coupled partial differential equations*. Vol. 155. Internat. Ser. Numer. Math. Birkhäuser, Basel, 2007, pp. 293–309.

[Ulb11] M. Ulbrich. *Semismooth Newton methods for variational inequalities and constrained optimization problems in function spaces*. Vol. 11. MOS-SIAM Series on Optimization. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA; Mathematical Optimization Society, Philadelphia, PA, 2011, pp. xiv+308.

[WB08] G. von Winckel and A. Borzì. "Computational techniques for a quantum control problem with H^1-cost." In: *Inverse Probl.* 24.3 (2008), p. 034007.

[WBV10] G. von Winckel, A. Borzì, and S. Volkwein. "A globalized Newton method for the accurate solution of a dipole quantum control problem." In: *SIAM J. Sci. Comput.* 31.6 (2009/10), pp. 4176–4203.

[WG07] J. Werschnik and E. K. U. Gross. "Quantum optimal control theory." In: *J. Phys. B At. Mol. Opt. Phys.* 40.18 (2007), R175–R211.

[Zin57] I. Zinger. "Linear functionals on the space of continuous mappings of a compact Hausdorff space into a Banach space." In: *Rev. Math. Pures Appl.* 2 (1957), pp. 301–315.