Computational techniques for a quantum control problem with H^1 -cost

G. von Winckel¹, A. Borzi²

Institut für Mathematik und Wissenschaftliches Rechnen, Karl-Franzens-Universität,
Heinrichstraße 36, 8010 Graz, Austria

E-mail: gregory.von-winckel@uni-graz.at, alfio.borzi@uni-graz.at

Abstract.

The accurate numerical computation of optimal controls of infinite-dimensional quantum control problems is a very difficult task that requires to take into account the features of the original infinite-dimensional problem. An important issue is the choice of the functional space where the minimization process is defined. A systematic comparison of L^2 - versus H^1 -based minimization shows that the choice of the appropriate functional space matters and has many consequences in the implementation of some optimization techniques.

1. Introduction

Optimal control theory provides a powerful framework for investigating quantum control problems of various complexity. Very often the aim of these problems is to design electromagnetic control fields that drive the system to a desired configuration corresponding to a final state represented by a wavefunction profile or to some property depending on given quantum expectation values. In this framework, the dynamic of the evolving systems is described by Schrödinger-type equations where the control fields are represented through potentials that enter in the equations as multiplicative functions of the state variable. Therefore, quantum control problems belong to the class of bilinear control systems that are attracting increasingly attention due to the fact that bilinear control is a nonlinear control strategy that may provide a better system response than possible with linear control.

To illustrate the challenge given by quantum control problems, we need to discuss their formulation in more detail. These problems consist of an evolutionary quantum system including the bilinear control mechanism and a functional modeling the purpose of the control and the cost of its action. Therefore the objective functional depends on the state configuration and features describing the desired target as well as the control fields. The resulting task is to minimize the objective subject to the constraint given by the dynamics equations. Therefore, we have an infinite dimensional optimization problem where the constraint is represented by a (nonlinear) time-dependent Schrödinger equation.

It turns out that quantum control problems possess very flat minima that are sensitive to problems' parameters and discretization parameters. Moreover, some of these local minima eventually disappear when sufficiently fine discretization is used. This fact renders the accurate computation of optimal controls very difficult and requires to take into account the features of the original infinite dimensional problem. In particular, we are interested to investigate the consequences of the choice of the functional space where the minimization process is defined. Our interest is motivated by the fact that usually quantum control problems (and corresponding gradients) are formulated and solved on L^2 spaces while the cost of the controls in the objective may be of H^1 type. We show that the choice of the appropriate functional space matters and has many consequences for the implementation of the optimization techniques considered in this paper.

As a benchmark for our discussion we consider the Bose–Einstein condensates control problem [24]. The objective of the control is quantified through a cost functional depending on the final state as well as the physically motivated constraint of smooth varying magnetic control fields. This functional is minimized subject to the condition that the time dynamics of the condensate is governed by the Gross–Pitaevskii equation [15]. This is a quantum model that describes nonlinear wavefunction evolution in complex Hilbert spaces.

We address the following issues that enter in the solution of quantum control

problems considering L^2 - and H^1 -based minimization: 1) Optimality systems in the L^2 and H^1 spaces. 2) Formulation of nonlinear conjugate gradient (NCG) and standard and cascadic BFGS schemes in these two spaces and consequences for the line search. 3) Numerical performance of L^2 - and H^1 -based optimization. We also discuss alternative minimization procedures that provide a link to a statistical approach to inverse problems.

In the next section, we discuss the formulation of the optimal control of the transport of Bose–Einstein condensates and consider two possible choices of the functional space where necessary first-order optimality conditions are formulated, the L^2 and H^1 spaces. For each of these two choices a coupled system of two nonlinear Schrödinger equations, with opposite time orientation, and an equation for the control results. These are the so-called state and adjoint equations and the optimality condition. Sequential solution of these equations provides the gradient in the appropriate space. The notion of the gradient is used in Section 3 where we define the Hager–Zhang NCG scheme and the BFGS scheme in a general functional space. The choice of line-search method is discussed. We also discuss interpolating the BFGS state history for a cascadic BFGS scheme. This method offers an increase in the rate at which the minimum is cases where obtaining the minimum is more difficult.

To solve these nonlinear optimality systems we proceed in Section 4 by introducing a time-splitting spectral discretization [4, 3] adapted to the resulting state and adjoint equations and use finite differences to discretize the optimality condition. Thus we obtain second-order accuracy in time-discretization and spectral accuracy in space.

In Section 5 results of numerical experiments are reported to compare the computational performance of the the solution strategies resulting from two distinct functional analytical settings. We obtain that minimization on L^2 vs. H^1 spaces have significant disparity in robustness with respect to regularization parameter and that CPU time to reaching the optimum is strongly affected.

In Section 6 we consider minimization with polynomial basis functions without the adjoint equation and explicit construction of the the gradient. A section of conclusion completes the exposition of our work.

2. Quantum optimal control problems

Recently, quantum optimal control of transport of Bose–Einstein condensates in magnetic microtraps was proposed [24] to solve the problem first considered in [20] for a trapped-atom interferometer setup where a dilute Bose–Einstein condensate should be split from a single to a double well groundstate. A Bose-Einstein condensate is a state of matter formed by bosons (e.g., helium-4, rubidium) cooled to temperatures very near to absolute zero. Under such conditions, the atoms with magnetic spin collapse into the lowest quantum state sharing the same wavefunction, and quantum effects become apparent on a macroscopic scale. Coherent manipulation of this wavefunction, as the splitting mentioned above, is one of the ultimate goals of modern atom optics, is a promising approach towards full control of matter waves on small scales; see

[16, 17, 20, 21, 41], and references therein.

The mean-field dynamics of a coherent Bose–Einstein condensate is described by the Gross–Pitaevskii equation [15]

$$i \frac{\partial}{\partial t} \psi(x, t) = \left(-\frac{1}{2} \nabla^2 + V(x, \lambda(t)) + g |\psi(x, t)|^2 \right) \psi(x, t), \quad (1)$$

where $x \in \Omega$ and $t \in [0, T]$, with g a coupling constant related to the scattering length of the atoms, density, and transversal confinement. Notice that, in principle, the evolution of the Bose–Einstein condensate is defined in unbounded spaces while the fact that V is a confinement potential results in wavefunctions ψ whose support is localized in a bounded region. Therefore, with Ω we represent a spatial domain that is large enough to represent the support of ψ during evolution.

We set $\hbar = 1$, and measure mass in units of the atom mass and length in units of micrometers. We assume that the wavefunction $\psi(x, t)$ is normalized to one $\int_{\Omega} |\psi(x, t)|^2 dx = 1$, $t \geq 0$, and therefore g in (1) incorporates the number of atoms N_A . Alternatively as in, e.g., [15], we could choose to normalize the wavefunction to N_A in the condensates, then g represents the strength of interatomic interaction.

We consider confining potentials $V_{\lambda}(x, t) = V(x, \lambda(t))$ that are produced by magnetic microtraps whose variation is described by a control function $\lambda(t)$. We assume that $\lambda(t)$ is real and single-valued. We treat the case typical in experimental settings where the initial and final potential configuration are given, therefore we require that $\lambda(t)$ takes initial and final values of zero and one, respectively. These two extremal values correspond to the case where the potential $V_{\lambda}(x, t)$ is convex and to the case where it has a double-well structure. Furthermore we require that V_{λ} is spatially symmetric with respect to the origin of coordinates $V_{\lambda}(x, t) = V_{\lambda}(-x, t)$ and $V_{\lambda}(x, \cdot) \in C^2(\Omega)$, and assume that V_{λ} is twice continuously differentiable in λ .

Suppose that initially the system is in the ground state ψ_0 for the potential $V(x, \lambda)|_{\lambda=0}$. Upon varying $\lambda(t)$ in the time interval $t \in [0, T]$ from zero to one, the system will pass through a sequence of states and will end up in the final state $\psi(T)$. Our purpose is to determine an optimal time evolution of $\lambda(t)$ that allows to channel the system from the initial state ψ_0 at time zero to a desired state ψ_d at final time T . In accordance to [20], we assume ψ_d to be the ground state for the potential $V(x, \lambda)|_{\lambda=1}$ at time T . The ground state for a given potential $V(x, \lambda)$ is defined as the stationary state $\phi(x)$ with $\int_{\Omega} |\phi(x)|^2 dx = 1$ that minimizes the energy [2]

$$E_{\lambda}(\phi) = \int_{\Omega} \left(\frac{1}{2} |\nabla \phi(x)|^2 + V_{\lambda}(x) |\phi(x)|^2 + \frac{g}{2} |\phi(x)|^4 \right) dx.$$

For our purpose, in order to define a well-defined control problem, we introduce the following cost functional

$$J(\psi, \lambda) = \frac{1}{2} (1 - |\langle \psi_d, \psi(T) \rangle|^2) + \frac{\gamma}{2} \int_0^T \left(\dot{\lambda}(t) \right)^2 dt \quad (2)$$

with $\langle u, v \rangle = \int_{\Omega} u(x)^* v(x) dx$ the usual inner product in complex spaces, where the $*$ denotes complex conjugate, and $\|u\| = \langle u, u \rangle^{1/2}$.

Figure 1. Stationary solutions of the Gross-Pitaevskii equation for nonlinearity strength $g = 1$ and $g = 50$. The initial state is shown in solid and the target state is the dashed line.

The goal of the first term of the cost functional is to track the state ψ to a given terminal state ψ_d at $t = T$. Exact tracking would result in $|\langle \psi_d, \psi(T) \rangle|^2 = 1$. The second term in the cost functional represents H^1 costs and aims at penalizing fast varying confinement potentials that are more difficult to realize in real experiments. This regularization term ensures existence of at least one optimal control. The regularization parameter $\gamma > 0$ allows to vary the relative importance of the objectives represented by the tracking and the cost of control.

The control problem under consideration is therefore to minimize $J(\psi, \lambda)$ subject to the condition that $\psi(x, t)$ fulfills the Gross-Pitaevskii equation (1) with given initial conditions.

To solve this problem we introduce the Lagrange function

$$L(\psi, p, \lambda) = J(\psi, \lambda) + \Re \left(p \left| i\dot{\psi} - \left(-\frac{1}{2}\nabla^2 + V_\lambda + g|\psi|^2 \right) \psi \right. \right), \quad (3)$$

with $(u|v) = \int_0^T \int_\Omega u(x, t)^* v(x, t) dx dt$, and $p(x, t)$ is the Lagrange multiplier.

By formally equating to zero the Frechét derivatives of L with respect to the triple (ψ, p, λ) , we obtain the following optimality system characterizing the solution to our optimal control problem. We have

$$i \frac{\partial \psi}{\partial t} = \left(-\frac{1}{2}\nabla^2 + V_\lambda + g|\psi|^2 \right) \psi \quad (4)$$

$$i \frac{\partial p}{\partial t} = \left(-\frac{1}{2}\nabla^2 + V_\lambda + 2g|\psi|^2 \right) p + g\psi^2 p^* \quad (5)$$

$$\gamma \frac{d^2 \lambda}{dt^2} = -\Re \langle \psi, \frac{\partial V_\lambda}{\partial \lambda} p \rangle, \quad (6)$$

which has to be solved together with the initial and terminal conditions

$$\psi(0) = \psi_0 \quad (7)$$

$$ip(T) = -\langle \psi_d, \psi(T) \rangle \psi_d \quad (8)$$

$$\lambda(0) = 0, \quad \lambda(T) = 1. \quad (9)$$

Equation (8) follows from computing time integration by parts to determine the Frechét derivative with respect to the variable ψ . Notice that while the state equation (4) with initial condition $\psi(0) = \psi_0$ evolves forward in time, the adjoint equation (5) with terminal condition (8) is marching backwards. The control equation (6) provides the optimality condition that determines the optimal control. Notice that because of H^1 regularization we have a natural setting to impose the required Dirichlet boundary conditions on the control function, $\lambda(0) = 0$ and $\lambda(T) = 1$.

We have that (4) is uniquely solvable for every $\lambda \in H^1(0, T; \mathbb{R})$ such that V_λ is a symmetric double well potential; see [34]. Thus, it is meaningful to introduce the so-called reduced cost functional $\hat{J} : H^1(0, T; \mathbb{R}) \rightarrow \mathbb{R}$ given by

$$\hat{J}(\lambda) = J(\psi(\lambda), \lambda), \quad (10)$$

where $\psi(\lambda)$ denotes the unique solution to (4) for given λ . One can show that the gradient of $\hat{J}$ with respect to λ is given by

$$\nabla \hat{J}(\lambda) = -\gamma \frac{d^2 \lambda}{dt^2} - \Re e \langle \psi, \frac{\partial V_\lambda}{\partial \lambda} p \rangle, \quad (11)$$

where ψ and p solve the state and the adjoint equations with given λ .

Now, the Taylor series of the reduced cost functional $\hat{J}(\lambda)$ in a Hilbert space is

$$\hat{J}(\lambda + \epsilon \varphi) = \hat{J}(\lambda) + \epsilon \left(\nabla \hat{J}(\lambda), \varphi \right) + \frac{\epsilon^2}{2} \left([\nabla^2 \hat{J}(\lambda)] \varphi, \varphi \right) + O(\epsilon^3) \quad (12)$$

The actual gradient will depend on the choice of which inner product space we use. If we choose the space $L^2(0, T)$, we have the inner product $(u, v) = \int_0^T u(t)v(t)dt$ and the gradient is

$$\nabla \hat{J}(\lambda) = -\gamma \frac{d^2 \lambda}{dt^2} - \Re e \langle \psi, \frac{\partial V_\lambda}{\partial \lambda} p \rangle \quad (13)$$

In the case of the H^1 formulation, the gradient must satisfies a one-dimensional Poisson equation with homogeneous Dirichlet conditions. That is,

$$\frac{d^2}{dt^2} [\nabla \hat{J}(\lambda)] = -\gamma \frac{d^2 \lambda}{dt^2} - \Re e \langle \psi, \frac{\partial V_\lambda}{\partial \lambda} p \rangle, \quad (14)$$

with $[\nabla \hat{J}(\lambda)](0) = 0$ and $[\nabla \hat{J}(\lambda)](T) = 0$. Two differences between the L^2 and H^1 formulations are immediately apparent. First, the L^2 gradient does not vanish at $t = T$ and second, the H^1 gradient possesses the same degree of smoothness as λ , while the L^2 gradient does not.

3. Optimization methods in L^2 and H^1 spaces

Because the control function $\lambda(t)$ is required to satisfy $\lambda(0) = 0$ and $\lambda(T) = 1$ and the update is $\lambda_{k+1} = \lambda_k + a_k \varphi_k$, it is necessary that $\varphi(0) = \varphi(T) = 0$. In the case of simple steepest decent, $\varphi = -\nabla J$, nonlinear conjugate gradients, and in the BFGS scheme, φ will always be in an element in the space spanned by $(u, v) = \int_0^T u'(t)v'(t)dt$

3.1. Function space HZ-NCG optimization

We discuss the minimization of the real-valued differentiable function $\hat{J}(\lambda)$, representing the reduced cost functional $J(\psi(\lambda), \lambda)$, where $\psi(\lambda)$ denotes the solution to the state equation (4). Denote with $g(\lambda) = \nabla \hat{J}(\lambda)$. In the following, we formulate the NCG scheme in a continuous setting in the L^2 space of real-valued functions. We have $(u, v) = (u, v)_{L^2(0, T; \mathbb{R})} = \int_0^T u(t)v(t)dt$ and $\|u\| = \sqrt{(u, u)}$.

In the common NCG variants, the basic idea is to avoid matrix operations and express the search directions recursively as

$$d_{k+1} = -g_{k+1} + \beta_k d_k, \quad (15)$$

for $k = 1, 2, \dots$, with $d_1 = -g_1$. In general, for convergence it is required that d_k is a descent direction for any k , i.e. $(g_k, d_k) < 0$ holds.

The iterates for a minimum point are given by

$$\lambda_{k+1} = \lambda_k + \tau_k d_k, \quad (16)$$

where $\tau_k > 0$ is a steplength. The parameter β_k is chosen so that (15)–(16) reduces to the linear CG scheme if $\hat{J}$ is a strictly convex quadratic function and τ_k is the exact one-dimensional minimizer of $\hat{J}$ along d_k .

We use on the NCG scheme of Hager and Zhang [19] based on the formula

$$\beta_k = \beta_k^{HZ} := \frac{(\sigma_k, g_{k+1})}{(d_k, y_k)}, \quad \sigma_k = y_k - 2d_k \frac{(y_k, y_k)}{(y_k, d_k)}, \quad (17)$$

where $y_k = g_{k+1} - g_k$. In our numerical experiments, the Hager-Zhang NCG formula covered to the minimum in fewer iterations than any other NCG formula the the numerical experiments.

In [14], convergence of the proposed NCG scheme is established requiring that the steplength τ_k , determined by backtracking, satisfies the Armijo condition of sufficient decrease of $\hat{J}$'s value given by

$$\hat{J}(\lambda_k + \tau_k d_k) \leq \hat{J}(\lambda_k) + \delta \tau_k (g_k, d_k) \quad (18)$$

together with the Wolfe condition

$$(g(\lambda_k + \tau_k d_k), d_k) > \sigma (g_k, d_k), \quad (19)$$

where $0 < \delta < \sigma < 1/2$; see [29]. The last condition means that the graph of $\hat{J}$ should not increase too fast beyond the minimum.

We implement the following NCG scheme

- Step 1. Given $k = 1$, λ_1 , $d_1 = -g_1$, if $\|g_1\| < \text{tol}$ then stop.
- Step 2. Compute $\tau_k > 0$ satisfying (18)–(19).
- Step 3. Let $\lambda_{k+1} = \lambda_k + \tau_k d_k$.
- Step 4. Compute $g_{k+1} = \nabla \hat{J}(\lambda_{k+1})$.
If $\|g_{k+1}\| < \text{tol}_{\text{abs}}$ or $\|g_{k+1}\| < \text{tol}_{\text{rel}} \|g_1\|$ or $k = k_{\text{max}}$ then stop.
- Step 5. Compute β_k by (17).
- Step 6. Let $d_{k+1} = -g_{k+1} + \beta_k d_k$.
- Step 7. Set $k = k + 1$, goto Step 2.

3.2. Function space BFGS optimization

The BFGS algorithm is a quasi-Newton method which makes successive rank-two updates to a matrix B such that it serves as an approximation to the true Hessian. Typically, BFGS exhibits convergence rates superior to those of NCG schemes at the expense additional computational effort. The search direction at the k th step is $p_k = -B_k^{-1}g_k$. Just as with the HZ-NCG scheme, the gradient is denated $g_k = \nabla \tilde{J}(\lambda_k)$ and the difference between two successive λ as $s_k = \alpha_k p_k$ and two successive gradients is $y_k = g_{k+1} - g_k$. The matrix B can be formed explicitly via the well-known recurrence formula.

$$B_{k+1} = B_k - \frac{(B_k s_k)(B_k s_k)^\top}{s_k^\top B_k s_k} + \frac{y_k y_k^\top}{y_k^\top s_k}, \quad (20)$$

To compute the search direction, it is necessary to invert the matrix B . We denote its inverse as $H = B^{-1}$. Using the Sherman-Morrison-Woodbury formula, we can also establish a recurrence for H .

$$H_{k+1} = H_k + \frac{s_k^\top y_k + y_k^\top H_k y_k}{(s_k^\top y_k)^2} (s_k s_k^\top) - \frac{H_k y_k s_k^\top + s_k y_k^\top H_k}{s_k^\top y_k} \quad (21)$$

In the case where the control λ and the gradient of the objective function $\nabla \tilde{J}$ are elements in a function space, it is not immediately obvious how to directly use this formula since it requires forming outer products. To compute the search direction, however, we only need the action of H on a vector g and it is not necessary to construct the matrix. Supposing X is either $L^2(0, T)$ or $H^1(0, T)$ and that $x, y \in X$, then we can denote the function space analog of the outer product as a dyadic operator $x \otimes y : X \rightarrow X$. The action of this operator on a third element $z \in X$ can be expressed in terms of the inner product $(x \otimes y)z = (y, z)_X x$. From the recursion relation for H the following sum formula for the search direction is obtained.

$$p_k = -H_0 g_k - \sum_{j=0}^{k-1} c_j [d_j(s_j, g_k)_X s_j - (z_j, g_k)_X s_j - (s_j, g_k)_X z_j] \quad (22)$$

Where $c_j = (s_j, y_j)_X^{-1}$, $d_j = 1 + c_j(y_j, z_j)$ and we also have a similar sum formula for the term z_k .

$$z_k = H_0 y_k + \sum_{j=0}^{k-1} c_j \{ [d_j(s_j, y_k)_X - (z_j, y_k)_X] s_j - (s_j, y_k)_X z_j \} \quad (23)$$

In the numerical implementation, these functions are approximated on a uniform grid. The function space inner product is in either case really just approximated by a vector inner product with a weighting matrix. The can be written as $(u, v)_{L^2}$ and $(u, v)_{H^1} \approx u^\top K v$ where the stiffness matrix $K = D^\top M D$. The finite difference matrix D is toeplitz tridiagonal for taking centered differences except for the first and last two rows which use non-centered differenced to account for the boundary. The mass matrix M is diagonal and contains the quadrature weights for the composite Simpson's rule. If the the PDE itself were to be discretized with a higher order scheme, then it would be desirable to use correspondingly higher order quadrature and differentiation rules for the inner products. In all numerical tests, H_0 was taken to be the identity matrix.

With both the NCG and BFGS schemes, the new control λ_{k+1} is composed of a linear combination of the original control and the gradients at every step. Unlike the NCG formulas, the matrix-free BFGS formula requires progressively more computation for each optimization step, so it is important that the improved convergence properties at least compensate for the increased computational effort.

3.3. Cascadic BFGS

The idea of cascadic NCG schemes have been treated earlier in [35]. The basic approach of these methods is to begin with a course grid approximation to the optimal control problem, take a number of NCG steps and interpolate on to a finer grid before proceeding. This concept is extended here to the matrix-free BFGS algorithm. In the case of NCG, the computation of the next state depends only on the previous state and as such, one need only interpolate the current state and search direction to proceed to the next level of refinement. In the matrix-free BFGS algorithm, a history of vectors must be maintained. These are the s_j vectors which are the search steps themselves, the y_j which are the differences between successive gradients, and also the z_j vectors which are elements in the space spanned by $s_0, \dots, s_{j-1}$. The set of these vectors, which increases with successive BFGS steps, must each be interpolated on to the next finest grid in the cascadic scheme. In all of the numerical experiments the interpolation was performed using the MATLAB function `spline`. Cubic splines are used rather than linear interpolation because the latter is not suitable for use with the H^1 seminorm.

- Step 1. If on the finest grid, proceed with BFGS minization until unable to take another descent step otherwise go to Step 2.
- Step 2. Letting the current state be λ_k , back up one iteration to the last state from which a suitable search length was found.
- Step 3. Interpolate the λ_{k-1} and all of the history vectors s_j , y_j , and z_j for $j = 0 \dots k - 1$ on to the next finer grid.
- Step 4. Go to step 2, using the most recent s vector as a search direction.

3.4. Linesearch method

In every test case the same line search strategy was used. The method is nothing more than a bisection approach for the approximate derivative of the objective function with respect to the step length which locates the nearest local minimum along the search direction. Although this method is by no means optimal, it was found to be sufficiently robust in the numerical experiments. There are two key reasons for choosing this linesearch over the more conventional backtracking method. First, using a large step length α to begin a backtracking sweep can be dangerous as the PDE may be under resolved for a potential $V(x, \lambda(t) + \alpha p(t))$ as large potentials can lead to highly oscillatory wave functions and the cost functional and its gradient can cease to be physically meaningful. Second, our numerical experiments

In all cases the numerical derivative of the line search function $\phi(\alpha) = \tilde{J}(\lambda + \alpha p)$ was taken using centered differences with a differential step size of $\delta\alpha = 10^{-4}$. No restarting was used in either of the BFGS schemes, however, the HZ-NCG scheme was restarted whenever $(g_k, g_{k-1}) \geq 0.1 \|g_k\|^2$. The stopping condition was taken to be the first iteration in which the line search was unable to find a step length which reduces the objective function. The usual reason for failing to find a suitable step size that the numerical error in discretizing the optimality system leads to some small disparity between the objective function and its gradient.

```

Data:  $\alpha_L = 0, \alpha_R = 0$ 
while  $\phi(\alpha_R) < \phi(0)$  do
  |  $\alpha_R \leftarrow 1.1(\alpha_R + .01);$ 
end
 $\alpha_M = (\alpha_L + \alpha_R)/2$ 
while  $|\alpha_R - \alpha_L| > 10^{-6}$  do
  | if  $\phi'(\alpha_M)\phi'(\alpha_L) < 0$  then
  | |  $\alpha_R = \alpha_M;$ 
  | else
  | |  $\alpha_L = \alpha_M;$ 
  | end
  |  $\alpha_M = (\alpha_L + \alpha_R)/2$ 
end

```

Algorithm 1: The bisection line search method **Figure 2.** The cost functional along the search direction for successive descent steps.

4. Discretization method

In this section, we discuss the discretization of (4) and (5) using a unconditionally stable explicit second-order norm-preserving time-splitting spectral scheme (TSSP) [4, 3]. These properties make the TSSP scheme particularly suitable for computation on a hierarchy of grids. To introduce the time-splitting technique and the corresponding notation we discuss the discretization of the forward equation. For the backward adjoint equation additional work is required to implement the time-splitting method appropriately. This is due to the presence of the term $g\psi^2 p^*$ in the adjoint equation. This problem is discussed in detail below.

For ease of notation we take $\Omega = (-L/2, L/2) \subset \mathbb{R}$ where L is large enough so that the support of the state and adjoint variables are well within the domain. Assume the interval $(-L/2, L/2)$ is divided in N subintervals of size $h = L/N$. Subinterval

end-points are denoted by $x_j = (j-1)h - L/2$, $j = 1, \dots, N$, where we take N to be an integer power of 2. The point x_{N+1} corresponds to $x = L/2$. Further we assume that the time interval $(0, T)$ is divided in M subintervals thus the time step-size is given by $\delta t = T/M$. The approximation to $\psi(x, t)$ at x_j for the time $t_m = m\delta t$ is denoted with ψ_j^m . We set $\mu_k = \frac{2\pi}{L}k$. We assume that the function ψ is periodic in $(-L/2, L/2)$ in the sense that $\psi(-L/2^+) = \psi(L/2^-)$.

For a given continuous periodic function ψ , consider the polynomial

$$I_N\psi(x) = \sum_{k=-(\frac{N}{2}-1)}^{\frac{N}{2}} \tilde{\psi}_k e^{i\mu_k x}, \quad (24)$$

where

$$\tilde{\psi}_k = \frac{1}{N} \sum_{j=1}^N \psi(x_j) e^{-i\mu_k x_j} \quad \text{with} \quad x_j = (j-1)h - L/2. \quad (25)$$

The function $I_N\psi(x)$ is the $\frac{N}{2}$ -degree trigonometric interpolant of ψ at the nodes x_j , i.e.

$$I_N\psi(x_j) = \psi(x_j) \quad j = 1, \dots, N. \quad (26)$$

This polynomial is the discrete Fourier series of ψ .

The Fourier pseudospectral derivative of ψ is defined by $D_N\psi = (I_N\psi)'$. That is,

$$D_N\psi(x) = \sum_{k=-(\frac{N}{2}-1)}^{\frac{N}{2}} \tilde{\psi}'_k e^{i\mu_k x} \quad \text{where} \quad \tilde{\psi}'_k = i\mu_k \tilde{\psi}_k.$$

Further, the Laplacian is given by

$$D_N^2\psi(x) = \sum_{k=-(\frac{N}{2}-1)}^{\frac{N}{2}} \tilde{\psi}''_k e^{i\mu_k x} \quad \text{where} \quad \tilde{\psi}''_k = -\mu_k^2 \tilde{\psi}_k.$$

Now let $H = H_0 + V$, where $H_0 = -\frac{1}{2}\nabla^2$ is the free Hamiltonian and $V = V_\lambda + g|\psi|^2$ represents the effective potential including the magnetic confinement potential. Next we illustrate the time-splitting spectral scheme for the Gross–Pitaevskii equation written as follows; see [3] for more details.

$$i \frac{\partial \psi}{\partial t} = (H_0 + V) \psi.$$

The time-splitting scheme for this equation can formally be written in the following form

$$\psi^{m+1} = e^{-i\frac{\delta t}{2}V^{m+1}} e^{-i\delta t H_0} e^{-i\frac{\delta t}{2}V^m} \psi^m. \quad (27)$$

From time t_m to time t_{m+1} we have three steps. For a $\delta t/2$ time step we first solve

$$i \frac{\partial \psi}{\partial t} = V \psi.$$

For $j = 1, \dots, N$, we obtain

$$\psi_j^+ = \exp \left(-i(V_\lambda(x_j, t_m) + g|\psi_j^m|^2) \delta t/2 \right) \psi_j^m. \quad (28)$$

With this value, we compute a full time step for the equation $i \frac{\partial \psi}{\partial t} = H_0 \psi$ obtaining

$$\psi_j^{++} = \sum_{k=-(\frac{N}{2}-1)}^{\frac{N}{2}} \tilde{\psi}_k^+ e^{(-i\mu_k^2 \delta t/2)} e^{i\mu_k x_j}, \quad j = 1, \dots, N, \quad (29)$$

where

$$\tilde{\psi}_k^+ = \frac{1}{N} \sum_{j=1}^N \psi_j^+ e^{-i\mu_k x_j}.$$

The final step consists in another $\delta t/2$ time step of the evolution governed by V with ψ^{++} as initial condition. Hence we have

$$\psi_j^{m+1} = \exp \left(-i(V_\lambda(x_j, t_{m+1}) + g|\psi_j^{++}|^2) \delta t/2 \right) \psi_j^{++}, \quad j = 1, \dots, N. \quad (30)$$

Additional work is required to implement the time-splitting scheme for the adjoint equation (5) because of the presence of the term $g \psi^2 p^*$ that prevents a straightforward application of an exponential solution formula. We proceed as follows. The adjoint equation can be written in the following form

$$i \frac{\partial}{\partial t} (p_r + i p_i) = \left(-\frac{1}{2} \nabla^2 + A \right) (p_r + i p_i) + (a_r + i a_i) (p_r - i p_i),$$

where $p = p_r + i p_i$, $A = V_\lambda + 2g|\psi|^2$, and $g \psi^2 = a_r + i a_i$. We focus on the first and third step of the time-splitting formula and therefore we consider

$$\frac{d}{dt} (p_r + i p_i) = -iA (p_r + i p_i) - i(a_r + i a_i) (p_r - i p_i).$$

This equation is equivalent to the following system of differential equations

$$\dot{p}_r = A p_i + (a_i p_r - a_r p_i) \quad (31)$$

$$\dot{p}_i = -A p_r - (a_r p_r + a_i p_i). \quad (32)$$

We now introduce the Pauli matrices to write this system in a form suitable for exponential representation. The Pauli matrices are

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \text{and} \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

and we write $\bar{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$.

Therefore the system (31)–(32) can be written as follows

$$\frac{d}{dt} \begin{pmatrix} p_r \\ p_i \end{pmatrix} = i(\bar{u} \cdot \bar{\sigma}) \begin{pmatrix} p_r \\ p_i \end{pmatrix},$$

where $\bar{u} = (ia_r, A, -ia_i)$. We can now use the exponential representation to obtain

$$\begin{pmatrix} p_r \\ p_i \end{pmatrix} (t + \delta t) = \exp(i \bar{u} \cdot \bar{\sigma} \delta t) \begin{pmatrix} p_r \\ p_i \end{pmatrix} (t). \quad (33)$$

Further, we use the fact that

$$\begin{aligned} \exp(i \bar{u} \cdot \bar{\sigma} \delta t) &= \cos(|\bar{u}| \delta t) I + i \sin(|\bar{u}| \delta t) \frac{\bar{u}}{|\bar{u}|} \cdot \bar{\sigma} \\ &= \cos(|\bar{u}| \delta t) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \sin(|\bar{u}| \delta t) \frac{1}{|\bar{u}|} \begin{pmatrix} a_i & A - a_r \\ -A - a_r & -a_i \end{pmatrix}, \end{aligned}$$

where $|\bar{u}| = \sqrt{a_r^2 + a_i^2 + A^2}$. We can therefore use this formula in (33) and replace δt with $-\delta t/2$ to obtain the (backward) time evolution of the adjoint variable for the first and third step of the time-splitting procedure. Notice that in this case because of the terms A and a_r , a_i , also the values of the wavefunction ψ at intermediate half time-steps are required. For this purpose, the ψ function is also computed in the backward evolution using (28)–(30) with δt replaced by $-\delta t$. This is possible since the time-splitting scheme is time-reversible.

Evaluation of the gradient of the reduced cost functional is given by the following

$$\nabla \hat{J}(\lambda)^m = -\gamma \frac{\lambda^{m+1} - 2\lambda^m + \lambda^{m-1}}{\delta t^2} - \Re \sum_{j=1}^N h (p_j^m)^* \frac{\partial V_\lambda}{\partial \lambda} \Big|_{\lambda=\lambda^m} \psi_j^m.$$

The accuracy of the discretization method can be observed by examining the error term

$$E(\epsilon, h) = \left| \frac{\tilde{J}(\lambda + \epsilon\varphi) - \tilde{J}(\lambda - \epsilon\varphi)}{2\epsilon} - \left(\varphi, \nabla \tilde{J}(\lambda) \right)_X \right| \quad (34)$$

Where φ is a test function which vanishes at the boundaries. When ϵ is sufficiently small that $E(\epsilon, h)$ no longer changes with respect to ϵ , we can observe that this error scales quadratically with the time step, h . This is consistent with the order of the split operator time-stepping scheme. The initial state $\psi(x, 0)$ and the target state $\psi_d(x)$ are obtained by solving the stationary Gross-Pitaevskii equation $-\psi_{xx} + 2V\psi + 2g|\psi|^2\psi = 0$ with λ equal to zero or one respectively. The nonlinear problem is solved using Newton's method with the ground state eigenfunction of the linear $g = 0$ problem as an initial guess.

5. Numerical investigation

Although the HZ-NCG method exhibits superior convergence properties for an NCG scheme, it does not minimize the cost functional as quickly as the BFGS methods do.

When the nonlinearity becomes large or the simulation interval becomes shorter, the optimal control function become more highly oscillatory and further away from the initial linear control function.

The regularization parameter, γ plays an important role in the L^2 scheme as it forces the solution to have a finite H^1 seminorm. The limit as $\gamma \rightarrow 0$ small gives a control function, $\lambda(t)$, which is discontinuous at $t = T$. Since the H^1 scheme produces $\lambda(t) \in H^1$ at every optimization step, the regularization term is not required for a smooth control. Rather the γ term only serves to penalize highly oscillatory solutions.

Figure 3. The optimal control function in the L^2 and H^1 settings for decreasing transition time.

- A plot showing (left) the difficulty of the and (right) discrepancy between computed gradient and finite difference gradient.
- Discussion of increasing difficulty as T becomes smaller. As Fig. 3 (left) for L^2 (right).
- Table 4 and Table ?? with T small; CPU times. Similar results with NCG.

Figure 4. Dependence of the optimal control function on the regularization parameter γ for the L^2 and H^1 settings.

Figure 5. Effect of the regularization parameter, γ on the convergence in the L^2 and H^1 settings.

Figure 6. Reduction of cost function over time for HZ-NCG, BFGS, and cascadic BFGS. The cascadic scheme gives superior performance for shorter time horizons

Figure shows the optimal control functions obtained using the Hagar-Zhang Nonlinear Conjugate Gradient method for the cases where the gradient is defined in L^2 and H^1 . The very large slope of $\lambda(t \rightarrow T)$ is typical of our observations using the L^2 gradient. Computing the optimal solution in L^2 took 54 steps and yielded $\tilde{J}(\lambda_{opt}^{L^2}) = 3.5248 \times 10^{-3}$ and $\|\nabla \tilde{J}(\lambda_{opt}^{L^2})\|_{L^2} = 1.2331 \times 10^{-1}$, whereas in H^1 , 69 search steps were taken resulting in $\tilde{J}(\lambda_{opt}^{H^1}) = 1.3624 \times 10^{-3}$ and $\|\nabla \tilde{J}(\lambda_{opt}^{H^1})\|_{L^2} = 1.4399 \times 10^{-2}$.

6. Alternative approaches

In the above work, only minimization schemes which use the optimality conditions have been considered. One might observe that the optimal control functions $\lambda_{opt}(t)$ frequently resemble low-order polynomials.

Figure 7. Reduction of cost function over time for three different optimization algorithms.

grid points	$\tilde{J}_{min}$	$\ \nabla \tilde{J}_{min}\ _{L^2}$	iterations	CPU time (sec)
400	2.9676×10^{-3}	4.4182×10^{-2}	81	1.6271×10^2
800	1.6492×10^{-3}	2.113×10^{-2}	110	4.1789×10^2
1600	1.8103×10^{-3}	1.1400×10^{-2}	155	1.1932×10^3
3200	3.4755×10^{-3}	1.13815×10^{-1}	76t	1.1985×10^3

Table 1. Results with L^2 BFGS minimization with $g=10$, $T=10$.

6.1. Minimization with basis functions

To Motivate the discussion below, spectral analysis of resulting λ for different T , γ , and g .

This spectral analysis should show that H^1 -based computation provides smoother λ .

As an alternate to the gradient-based optimization methods using the adjoint equation, we could reduce the dimensionality of the problem by writing the control function $\lambda(t)$ in terms of a suitable series of basis functions.

$$\lambda(t) = \frac{t}{T} + \sum_{k=1}^{N_k} a_k \phi_k \left(1 - \frac{2t}{T} \right), \quad \phi_k(\cdot) = T_{k-1}(\cdot) - T_{k+1}(\cdot) \quad (35)$$

where $T_k(\cdot)$ is the k th Chebyshev polynomial of the first kind. We can then attempt to minimize the functional $J(\mathbf{a})$ with respect to the coefficients using a black-box type algorithm such as `fminunc` which is included in the MATLAB optimization toolbox.

grid points	$\tilde{J}_{min}$	$\ \nabla \tilde{J}_{min}\ _{L^2}$	iterations	CPU time (sec)
400	3.0295×10^{-2}	2.7397×10^{-1}	24	3.3911×10^1
800	2.2888×10^{-3}	2.9977×10^{-2}	217	7.9421×10^2
1600	1.1995×10^{-3}	1.8310×10^{-2}	153	1.1986×10^3
3200	2.4388×10^{-3}	4.5315×10^{-3}	74	1.1883×10^3

Table 2. Results with H^1 BFGS minimization with $g=10$, $T=10$.

grid points	$\tilde{J}_{min}$	$\ \nabla \tilde{J}_{min}\ _{L^2}$	iterations	CPU time (sec)
400	1.8922×10^{-3}	5.8348×10^{-2}	25	6.1990×10^1
800	1.2472×10^{-3}	9.7978×10^{-2}	34	1.7211×10^2
1600	3.5128×10^{-4}	1.3702×10^{-2}	105	9.6031×10^2
3200	1.1394×10^{-4}	5.8215×10^{-3}	500	8.1292×10^3

Table 3. Results with L^2 HZ-NCG minimization with $g=10$, $T=10$.

grid points	$\tilde{J}_{min}$	$\ \nabla \tilde{J}_{min}\ _{L^2}$	iterations	CPU time (sec)
400	1.6605×10^{-2}	1.4288×10^{-1}	15	3.8407×10^1
800	5.5963×10^{-4}	4.5284×10^{-2}	62	2.8107×10^2
1600	2.9634×10^{-4}	1.0733×10^{-2}	30	3.6334×10^2
3200	1.0562×10^{-4}	3.6378×10^{-3}	37	9.6153×10^2

Table 4. Results with H^1 HZ-NCG minimization with $g=10$, $T=10$.

7. Conclusions

Acknowledgement

Supported in part by the Austrian Science Fund FWF project P18136-N13 “Quantum optimal control of semiconductor nanostructures”.

References

- [1] A.D. Bandrauk, M.C. Delfour, C. Le Bris, (Eds.), *Quantum control: Mathematical and Numerical Challenges*, CRM proceedings & lecture notes series, Vol. 33, American Mathematical Society, 2004.
- [2] W. Bao and Q. Du, *Computing the ground state solution of Bose–Einstein condensates by a normalized gradient flow*, SIAM J. Sci. Comput., 25 (2004), pp. 1674–1697.
- [3] W. Bao, D. Jaksch, and P.A. Markowich, *Numerical solution of the Gross-Pitaevskii Equation for Bose–Einstein condensation*, J. Comput. Phys., 187 (2003), pp. 318–342.
- [4] W. Bao, S. Jin, and P.A. Markowich, *On time-splitting spectral approximation for the Schrödinger equation in the semiclassical regime*, J. Comput. Phys., 175 (2002), pp. 487–524.
- [5] F.A. Bornemann and P. Deuffhard, *The cascadic multigrid method for elliptic problems*, Numer. Math., 75 (1996), pp. 135–152.
- [6] A. Borzì, *On the convergence of the MG/OPT method*, PAMM, 5(1) (2005), pp. 735–736.
- [7] A. Borzì and E. Decker, *Analysis of a leap-frog pseudospectral scheme for the Schroedinger equation*, J. Comp. Appl. Math., 193 (2006), pp. 65–88.

- [8] A. Borzi, J. Salomon, and S. Volkwein, *Formulation and numerical solution of finite-level quantum optimal control problems*, to appear in JCAM (<http://math.uni-graz.at/reports/archive-2006/IMA13-06.pdf>).
- [9] A. Borzi, G. Stadler, and U. Hohenester, *Optimal quantum control in nanostructures: Theory and application to a generic three-level system*, Phys. Rev. A 66, (2002) 053811.
- [10] A. Brandt, *Multi-level adaptive solutions to boundary-value problems*, Mathematics of Computation, 31 (1977), pp. 333–390.
- [11] P.W. Brumer and M. Shapiro, *Principles of the Quantum Control of Molecular Processes*, Wiley-VCH, Berlin, 2003.
- [12] A.G. Butkovskiy and Yu.I. Samoilenko, *Control of quantum-mechanical processes and systems*, Kluwer, 1990.
- [13] E. Charron, M. Cirone, A. Negretti, J. Schmiedmayer, and T. Calarco, *Theoretical analysis of the implementation of a quantum phase gate with neutral atoms on atom chips*, Phys. Rev. A, 74 (2006), 012308.
- [14] Y.H. Dai and Y. Yuan, *A non-linear conjugate gradient with a strong global convergence property*, SIAM J. Opt. 10 (1999), pp. 177–182.
- [15] F. Dalfovo, G. Stefano, L.P. Pitaevskii, S. Stringari, *Theory of Bose–Einstein condensation in trapped gases*, Rev. Mod. Phys., 71 (1999), pp. 463–512.
- [16] R. Folman, P. Krüger, D. Cassettari, B. Hessmo, T. Maier, and J. Schmiedmayer, *Controlling Cold Atoms using Nanofabricated Surfaces: Atom Chips*, Phys. Rev. Lett., 84 (2000), 4749.
- [17] R. Folman, P. Krüger, D. Cassettari, J. Schmiedmayer, J. Denschlag, and C. Henkel, *Microscopic atom optics: from wires to an atom chip*, Adv. in Atom. Mol. and Opt. Phys., 48 (2002), 263.
- [18] J.C. Gilbert and J. Nocedal, *Global convergence properties of conjugate gradient methods for optimization*, SIAM J. Opt. 2 (1992), pp. 21–42.
- [19] W.W. Hager and H. Zhang, *A New Conjugate Gradient Method with Guaranteed Descent and an Efficient Line Search*, SIAM Journal on Optimization, 16 (2005), pp. 170–192.
- [20] W. Hänsel, J. Reichel, P. Hommelhoff, and T.W. Hänsch, *Trapped-atom interferometer in a magnetic microtrap*, Phys. Rev. A, 64(2001), 063607.
- [21] W. Hänsel, P. Hommelhoff, T.W. Hänsch, and J. Reichel, Nature, 413 (2001), pp. 498–501.
- [22] W. Hackbusch and A. Reusken, Analysis of a damped nonlinear multilevel method, Numer. Math., 55 (1989), pp. 225–246.
- [23] U. Hohenester, *Optical properties of semiconductor nanostructures: Decoherence versus quantum control*, to appear in M. Rieth and W. Schommers, (Eds.), *Handbook of Theoretical and Computational Nanotechnology* (American Scientific Publishers, 2005); cond-mat/0406346.
- [24] U. Hohenester, P.-K. Rekdal, A. Borzi, J. Schmiedmayer, *Optimal quantum control of Bose Einstein condensates in magnetic microtraps*, Phys. Rev. A, 75 (2007), 023602.
- [25] I. Lesanovsky, T. Schumm, S. Hofferberth, L.M. Andersson, P. Krüger, and J. Schmiedmayer, *Adiabatic radio-frequency potentials for the coherent manipulation of matter waves*, Phys. Rev. A, 73 (2006), 033619.
- [26] R.M. Lewis and S.G. Nash, Model problems for the multigrid optimization of systems governed by differential equations, SIAM J. Sci. Comp., 26 (2005), pp. 1811–1837.
- [27] Y. Maday, J. Salomon, G. Turinici, *Monotonic time-discretized schemes in quantum control*, Num. Math., 103 (2006), pp. 323–338.
- [28] S.G. Nash, A multigrid approach to discretized optimization problems, Optimization Methods and Software, 14 (2000), pp. 99–116.
- [29] J. Nocedal and S.J. Wright, *Numerical Optimization*, Springer, New York, 1999.
- [30] J.P. Palao and R. Kosloff, *Quantum Computing by an Optimal Control Algorithm for Unitary Transformations*, Phys. Rev. Lett. 89, 188301 (2002).
- [31] T. Paul and K. Richter and P. Schlagheck, *Nonlinear resonant transport of Bose–Einstein condensates*, Phys. Rev. Lett., 94 (2005), 020404.
- [32] A.P. Peirce, M.A. Dahleh, and H. Rabitz, *Optimal control of quantum-mechanical systems:*

- Existence, numerical approximation, and applications*, Phys. Rev. A, 37 (1988), pp. 4950–4964.
- [33] H. Rabitz, G. Turinici, and E. Brown, *Control of Molecular Motion: Concepts, Procedures, and Future Prospects*, Ch. 14 in Handbook of Numerical Analysis, Volume X, P. Ciarlet and J. Lions, Eds., Elsevier, Amsterdam, 2003.
- [34] A. Sacchetti, *Nonlinear time-dependent one-dimensional Schrodinger equation with double-well potential*, SIAM J. Math. Anal., 35 (2004), pp. 1160–1176.
- [35] V.V. Shaidurov, *Some estimates of the rate of convergence of the cascadic conjugate gradient method*, Computers Math. Applic. 31 (1996), pp. 161–171.
- [36] D.F. Shanno, *Conjugate gradient methods with inexact searches*, Math. Oper. Res. 3 (1978), pp. 244–256.
- [37] T. Schumm and S. Hofferberth, L.M. Andersson, S. Wildermuth, S. Groth, I. Bar-Joseph, J. Schmiedmayer, and P. Krüger, *Matter-wave interferometry in a double well on an atom chip*, Nature Phys., 1 (2005), pp. 57–62.
- [38] S.H. Tersigni, P. Gaspard, and S.A. Rice, *On using shaped light pulses to control the selectivity of product formation in a chemical reaction: An application to a multiple level system*, J. Chem. Phys., 93 (1990), pp. 1670–1680.
- [39] U. Trottenberg, C. Oosterlee, and A. Schüller, *Multigrid*, Academic Press, London, 2001.
- [40] S. Wildermuth, S. Hofferberth, I. Leshchinsky, E. Haller, L. Mauritz-Andersson, S. Groth, I. Bar-Joseph, P. Krüger, and J. Schmiedmayer, *Microscopic magnetic-field imaging*, Nature (London), 435 (2005), 440.
- [41] J. Fortagh and C. Zimmermann, *Magnetic microtraps for ultracold atoms*, Reviews of Modern Physics, 79 (2007), pp. 235–289.