A Globalized Newton Method for the Optimal Control of Fermionic Systems

Gregory von Winckel

Abstract. A computational framework for determining optimal control fields for inducing energy state transitions in systems of several fermions in an infinite potential quantum well is presented. The full multiparticle system is numerically approximated using linear combinations of Slater determinants constructed from nodal trial functions, which leads to diagonalized matrix approximations of variable coefficient terms. First and second order optimality conditions are given for the control and a robust line search is described for computing a local minimizer.

1. Introduction

In recent years there has been growing interest in controlled quantum phenomena by means of external fields. The aim of quantum control is to effect change on a system whose dynamics are governed by the time-dependent Schrödinger equation such that the system reaches a particular configuration. Some applications for quantum control include quantum bits and logic devices, controlled chemical processes, and investigation of fundamental phenomena. Following the initial work of Peirce, Daleh, and Rabitz [12], the Lagrangian based optimal control strategy has become prevalent for determining the control field which drives a quantum system closest to a target state at a specified time. Optimal controls are typically computed using either monotonically convergent schemes [10], gradient based schemes such as nonlinear conjugate gradients, BFGS [15], or inexact Newton methods [17].

In practice, multiparticle systems are usually approximated using a many body approximation such as the Hubbard model, Born–Oppenheimer approximation, or Multiconfigurational Time-Dependent Hartree(–Fock) (MCTDH(F)) methods [11, 6], which tends to give a good approximation for large numbers of particles, but may be inaccurate when there are only a few particles. In this current work, we present an efficient discretization and optimal control technique for inducing state transitions in a one-dimensional system of noninteracting or interacting fermions. The basic approach for discretizing the multiple fermion system has been described in detail in [4] and the Krylov–Newton method for a single particle system has been presented in [16]. The current work, however, contains the first application of Newton's method to the multiparticle optimal control problem.

In most quantum control literature where the time-dependent Schrödinger equation (TDSE) is used as an equality constraint, one of the most common approaches is to replace the partial differential equation with a finite dimensional system of ordinary differential equations so there is a two or three level system [2, 1, 5, 13]. This method usually assumes the structure of the Hamiltonian in the eigenfunction basis and further assumes that higher level states play no role in the dynamics. It does, however, capture the basic bilinear structure of the full problem and is attractive since the small systems that result can be solved numerically very quickly.

Alternately, the TDSE may be discretized, for example, with the finite difference method [7], which also gives a finite dimensional, albeit considerably larger, system of equations. This approach is numerically more expensive, but makes direct use of the physical potential and allows for coupling into higher energy states. It may be the case, however, that the discretization may be superfluous in the sense that the method resolves states which do not have a significant occupation probability.

The approach in the current work it to combine both of these ideas. We discretize the multiparticle Hamiltonian directly and compute its eigenvectors which are used as a modal basis for the state. This would be the simplest version of the proper orthogonal decomposition (POD) applied to a symmetric quadratic problem, however, we do not use the term POD in this work as this approach of diagonalizing the Hamiltonian is completely standard practice in quantum mechanics. What is distinct here is that the diagonalization follows a spectral discretization of the full interaction problem and that the basis of eigenvectors is permuted for efficiency of solving the control problem. That is to say, the eigenvectors are ranked in importance to the problem by a heuristic method described below and the state is projected onto the first few selected vectors. The optimal control problem is then solved on this reduced basis. The state space is then augmented by adding the next few most important vectors and the optimization routine is restarted using the computed optimal control from the previous step. This process is repeated until augmenting the state space has no perceptible effect on the cost functional.

The organization of the paper is as follows: in section two, basic properties of fermionic systems in one dimension are presented. In section three, we give the space and time discretizations for the multiparticle system. In section four, the control problem is formulated and the state reduction method is described. Section five presents some computed optimal controls for the quantum well containing two, three, and four fermions, and section six contains the conclusion and discussion of future work.

2. Multiparticle systems in one dimension

To understand the time-dependent Schrödinger equation for multiple interacting fermions, it is advantageous first to consider the two particle case before its generalization to n particles. The TDSE for two fermions is

$$i\partial_t \psi(x_1, x_2, t) = \left\{ -(\partial_{x_1}^2 + \partial_{x_2}^2) + V(x_1, x_2, t) \right\} \psi(x_1, x_2, t). \qquad (2.1)$$

The wavefunction $\psi(x_1, x_2, t)$ contains information about both particles and in particular, moreover, following the Born rule, its modulus squared is understood to be a probability density function. The stationary states for the particle system are the solutions to the eigenvalue problem:

$$\left\{ -(\partial_{x_1}^2 + \partial_{x_2}^2) + V(x_1, x_2) \right\} \phi_j(x_1, x_2, t) = \lambda_j \phi(x_1, x_2), \qquad (2.2)$$

where the eigenvalue λ_j is the energy. Since these form a complete basis, the time dependent solution can be expanded as a linear combination of the eigenfunctions using time dependent coefficients:

$$\psi(x_1, x_2, t) = \sum_{k=1}^{\infty} c_k(t) \phi_k(x_1, x_2). \qquad (2.3)$$

Since the particles are indistinguishable, it is required that that this probability function be invariant under exchange of the particles, i.e.,

$$|\phi(x_1, x_2)|^2 = |\phi(x_2, x_1)|^2 \quad \Rightarrow \quad \phi(x_2, x_1) = \phi(x_1, x_2)e^{i\theta}. \qquad (2.4)$$

This means that $e^{i\theta}$ is the eigenvalue of a permutation operator P where $P\phi(x_1, x_2) = \phi(x_2, x_1) = e^{i\theta}\phi(x_1, x_2)$. Since $P^2 = I$, it follows that for two identical particles that $\phi(x_1, x_2) = \pm\phi(x_2, x_1)$. The Pauli exclusion principle for fermions stipulates that the wavefunction is antisymmetric. In the time independent case, if there is no interaction between the particles, then $V(x_1, x_2) = V_1(x_1) + V_2(x_2)$ and the system is said to be decomposable. The problem is separated into two uncoupled univariate problems by writing the Ansatz

$$\phi(x_1, x_2) = \phi_1(x_1)\phi_2(x_2) - \phi_1(x_2)\phi_1(x_2) = \begin{vmatrix} \phi_1(x_1) & \phi_2(x_1) \\ \phi_1(x_2) & \phi_2(x_2) \end{vmatrix}, \qquad (2.5)$$
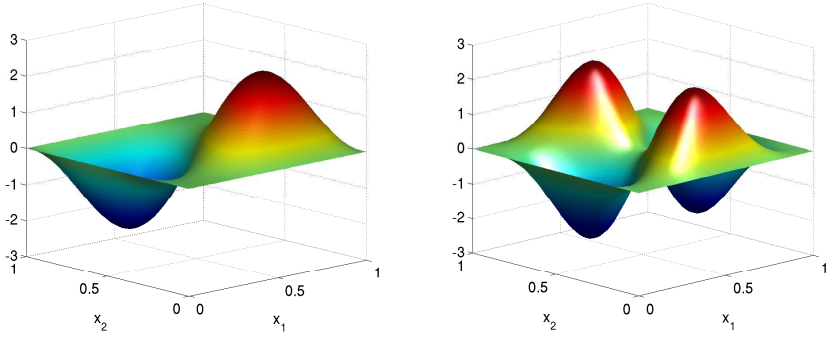
FIGURE 1. Two noninteracting fermions: state $|1,2\rangle$ and $|1,3\rangle$

where the determinant on the right hand side is called a Slater determinant. The eigenproblem can be separated into two uncoupled one-dimensional problems

$$\begin{aligned} \left\{-\partial_{x_1}^2 + V_1(x_1)\right\} \phi_{j_1}(x_1) &= \lambda_{j_1}\phi_{j_1}(x_1), \\ \left\{-\partial_{x_2}^2 + V_2(x_2)\right\} \phi_{j_2}(x_2) &= \lambda_{j_2}\phi_{j_2}(x_2). \end{aligned} \tag{2.6}$$

More generally for n non-interacting particles, the stationary states are still Slater determinants of size n:

$$\phi(x_1,\ldots,x_n) = \begin{vmatrix} \phi_1(x_1) & \phi_2(x_1) & \cdots & \phi_n(x_1) \\ \phi_1(x_2) & \phi_2(x_2) & \cdots & \phi_n(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(x_n) & \phi_2(x_n) & \cdots & \phi_n(x_n) \end{vmatrix}. \tag{2.7}$$

Consider the example of two noninteracting fermions in an infinite potential square quantum well with $x \in [0,1]$. The single particle eigenfunctions and eigenvalues are

$$\phi_j(x) = \sin(\pi j x), \quad \lambda_j = (j\pi)^2. \tag{2.8}$$

The first two eigenfunctions for the two particle problem are

$$\begin{aligned} |1,2\rangle &\equiv \phi_1(x_1,x_2) = \sin(\pi x_1)\sin(2\pi x_2) - \sin(2\pi x_1)\sin(\pi x_2), \\ |1,3\rangle &\equiv \phi_2(x_1,x_2) = \sin(\pi x_1)\sin(3\pi x_2) - \sin(3\pi x_1)\sin(\pi x_2), \end{aligned} \tag{2.9}$$

and the corresponding eigenvalues are $\lambda_1 = 5\pi^2$ and $\lambda_2 = 10\pi^2$. The first and second states are shown in figure 1.

If we add a third noninteracting fermion to the same well, then the first state $|1,2,3\rangle$ is the eigenfunction

$$\begin{aligned} \phi_1(x_1,x_2,x_3) = {}&\sin(\pi x_1)[\sin(2\pi x_2)\sin(3\pi x_3) - \sin(3\pi x_2)\sin(2\pi x_3)] \\ &+ \sin(2\pi x_1)[\sin(3\pi x_2)\sin(\pi x_3) - \sin(\pi x_2)\sin(3\pi x_3)] \\ &+ \sin(3\pi x_1)[\sin(\pi x_2)\sin(2\pi x_3) - \sin(2\pi x_2)\sin(\pi x_3)], \end{aligned} \tag{2.10}$$
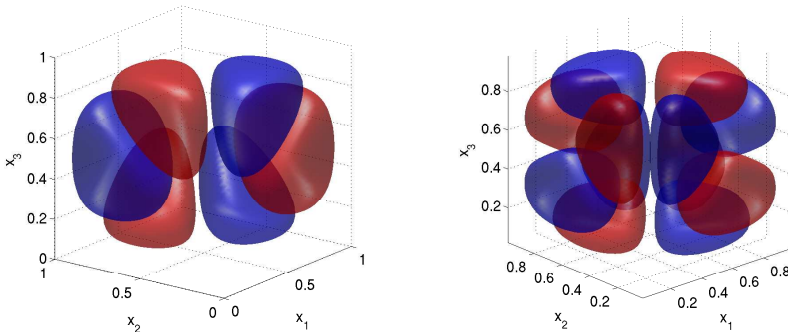
FIGURE 2. Three noninteracting fermions: state $|1, 2, 3\rangle$ and $|1, 2, 4\rangle$

and the first eigenvalue is $\lambda_1 = 14\pi$. The first two states for this system are depicted in Figure 2 where the red surfaces indicate the level sets where the eigenfunction is equal to half of its maximum value and the blue surfaces correspond to half the minimum value.

In stationary problems with nonzero interaction potential or time-dependent problems, it is no longer the case that the wave function is decomposable in this way and is not a single Slater determinant of one-dimensional functions. In the more general case, however, it is reasonable to write the wavefunction as a linear combination of Slater determinants.

The full multiparticle TDSE that we discretize and use as an equality constraint has the form

$$i\partial_t \psi(\mathbf{x}, t) = \left\{ -\Delta + \sum_{j=1}^{n} \left(V^c(x_j, t) + \sum_{k>j}^{n} V^i(x_j, x_k) \right) \right\} \psi_i(\mathbf{x}, t), \quad (2.11)$$

where $-1 \leq x_j \leq 1$ and the external potential experienced by the jth particle, $V^c(x_j, t) = u(t)x_j$ corresponds to a spatially-uniform electric field with time dependent amplitude. The interaction is modeled by a smoothed Coulomb potential

$$V^i(x_j, x_k) = \frac{q}{|x_j - x_k|} \approx \frac{q}{\sqrt{(x_j - x_k)^2 + \delta^2}} \quad (2.12)$$

with δ as a smoothing factor.

3. Numerical discretization

The wavefunction is discretized in each spatial dimension using the Legendre–Gauss numerical integration (G-NI) discretization, which is algebraically equivalent to the pseudospectral method, but leads to symmetric matrices. Consequently, all variable coefficient matrices will be diagonal, which is especially important for the interaction potential matrices which would otherwise be full in general.

For a single dimension, the approximation has the form

$$\psi(x) \approx \psi_p(x) = \sum_{k=1}^{p} \hat{\psi}_k \ell_k(x), \quad \ell_j(x) = \prod_{\substack{k=0 \\ k \neq j}}^{p+1} \frac{x - x_k}{x_j - x_k}, \tag{3.1}$$

where $j = 0, \dots, p + 1$. To enforce homogeneous Dirichlet conditions, we simply exclude $\ell_0(x)$ and $\ell_{p+1}(x)$. The Legendre–Gauss–Lobatto nodes are implicitly defined by

$$\{x_0, \dots, x_{p+1}\} = \{x | P'_{p+1}(x) = 0\} \cup \{\pm 1\}, \tag{3.2}$$

where $P_k(x)$ is the kth Legendre polynomial. From the nodes and Legendre polynomials, we also obtain the corresponding Lobatto weights

$$w_k = \frac{1}{(p+1)(p+2)} \frac{2}{[P_{p+1}(x_k)]^2}. \tag{3.3}$$

Starting with a one-dimensional eigenvalue problem such as in (2.6), expanding the wavefunction in the Lagrange trial basis, multiplying by a Lagrange test function and integrating by parts gives us the weak form of the eigenvalue problem

$$\sum_{k=1}^{p} [(\ell'_j, \ell'_k) + (\ell_j, V^c \ell_k)] \hat{\psi}_k = \lambda \sum_{k=1}^{p} (\ell_j, \ell_k) \hat{\psi}_k, \tag{3.4}$$

where the component matrices here are the stiffness or Laplacian matrix with elements, $\tilde{\mathbf{K}}_{jk} = (\ell'_j, \ell'_k)$, the confining potential matrix $\tilde{\mathbf{V}}^c_{jk} = (\ell_j, V^c \ell_k)$, and the mass matrix $\tilde{\mathbf{M}}_{jk} = (\ell_j, \ell_k)$. The inner products are then computed approximately using Legendre–Gauss–Lobatto quadrature:

$$\tilde{\mathbf{M}}_{jk} = \sum_{i=1}^{p} \ell_j(x_i) \ell_k(x_i) w_i, \quad \tilde{\mathbf{K}}_{jk} = \sum_{i=1}^{p} \ell'_j(x_i) \ell'_k(x_i) w_i. \tag{3.5}$$

This choice of discretization diagonalizes the mass matrix so that it contains the quadrature weights along the diagonal $\tilde{\mathbf{M}}_{jk} = w_j \delta_{jk}$. Since the quadrature weights are all positive, a trivial Cholesky factorization can be employed:

$$\tilde{\mathbf{M}} = \mathbf{R}^\top \mathbf{R}, \quad \mathbf{R}_{jk} = \sqrt{w_j} \delta_{jk}. \tag{3.6}$$

Transforming the eigenbasis by $\mathbf{R}$ gives us the algebraically equivalent simple eigenvalue problem in contrast to what was a generalized eigenvalue problem

$$[\mathbf{K} + \mathbf{V}^c] \hat{\varphi} = \lambda \hat{\varphi}, \quad \mathbf{K} = \mathbf{R}^{-\top} \tilde{\mathbf{K}} \mathbf{R}^{-1}, \quad \mathbf{V}^c = \mathbf{R}^{-\top} \tilde{\mathbf{V}}^c \mathbf{R}^{-1}. \tag{3.7}$$

Although some integration accuracy is sacrificed to yield diagonal matrix approximations to the variable coefficients, it has been shown that this method is algebraically equivalent to the standard pseudospectral method on the Legendre–Gauss–Lobatto nodes [3], however, in this setting all of the matrices are symmetric.

Following the idea of the Slater determinant formula for the decomposable problem, the multiparticle wave function is discretized using linear combinations of Slater determinants of the one-dimensional Lagrange interpolants.
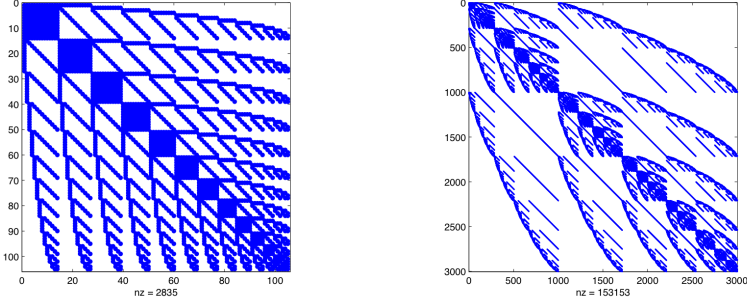
FIGURE 3. Left: $p = 15$ and $n = 2$, Right: $p = 15$ and $n = 5$

The n particle trial function φ is a Slater determinant of L^2-normalized Lagrange polynomials and to compute the Galerkin matrices, inner products must be computed involving trial and test functions, the latter being chosen from the same space for symmetry. Supposing we have two Slater determinants $A(x)$ and $B(x)$ such that

$$A(x) = \begin{vmatrix} a_1(x_1) & \cdots & a_n(x_1) \\ \vdots & \ddots & \vdots \\ a_1(x_n) & \cdots & a_n(x_n) \end{vmatrix}, \quad B(x) = \begin{vmatrix} b_1(x_1) & \cdots & b_n(x_1) \\ \vdots & \ddots & \vdots \\ b_1(x_n) & \cdots & b_n(x_n) \end{vmatrix}, \quad (3.8)$$

then the Löwden rule for Slater inner products [9] states that

$$\langle A(x), B(x) \rangle = \begin{vmatrix} \langle a_1, b_1 \rangle & \cdots & \langle a_1, b_n \rangle \\ \vdots & \ddots & \vdots \\ \langle a_n, b_1 \rangle & \cdots & \langle a_n, b_n \rangle \end{vmatrix}. \quad (3.9)$$

The total discretized Laplacian or stiffness matrix is $K = K^1 + K^2 + \cdots + K^n$ where $K^\nu_{jk} = \langle \partial_{x_\nu} \varphi_j, \partial_{x_\nu} \varphi_k \rangle$. We can write each of these components as

$$K^\nu_{jk} = \begin{vmatrix} \delta_{j_1,k_1} & \cdots & \delta_{j_1,k_{\nu-1}} & K_{j_1,k_\nu} & \delta_{j_1,k_{\nu+1}} & \cdots & \delta_{j_1,k_n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \delta_{j_n,k_1} & \cdots & \delta_{j_n,k_{\nu-1}} & K_{j_n,k_\nu} & \delta_{j_n,k_{\nu+1}} & \cdots & \delta_{j_n,k_n} \end{vmatrix}. \quad (3.10)$$

Ordinarily, discretizing an n-dimensional problem with p degrees of freedom per dimension would result in p^n grid points, however, exploiting the antisymmetry relations of the basis functions reduces the degrees of freedom to $N_p = \binom{p}{n}$. All variable coefficient matrices are diagonal, and the multiparticle Laplacian has a sparsity pattern that matches the adjacency matrix of the Johnson graph with the addition of a full diagonal band. The sparsity pattern for the Laplacian when $p = 15, n = 2$ and $p = 15, n = 5$ are displayed in Figure 3. An efficient method of computing the Laplacian, and variable coefficient matrices, which utilizes the combinatorial structure of the sparsity pattern arising from this discretization to achieve optimal run-time has been recently published [4].

After applying the spatial discretization, we obtain a semi-discrete equation of the form

$$i\psi_t = \{\mathbf{H}_0 + u(t)\mathbf{V^c}\}\psi, \quad \psi \in \mathbb{C}^{N_p}. \tag{3.11}$$

Typically N_p is quite large and it is unnecessary to solve the state equation using all degrees of freedom. Since we are mostly interested in transitions between low lying energy levels, the high level states will usually have extremely low occupancy probability and can be neglected. Instead, we use a reduced order approximation of the state by means of the eigenvalue decomposition. Compute first $N_s << N_p$ eigenpairs (Λ, Φ) of stationary Hamiltonian so that

$$\mathbf{H}_0\Phi = \Phi\Lambda, \quad \Phi \in \mathbb{R}^{N_p \times N_s}, \quad \Lambda \in \mathbb{R}^{N_s \times N_s}. \tag{3.12}$$

Projecting the state onto the subspace spanned by the computed eigenvectors gives the reduced state equation

$$iy_t = \{\Lambda + u(t)\mathbf{X}\}y, \quad y \in \mathbb{C}^{N_s}, \quad \mathbf{X} = \Phi^\top\mathbf{V^c}\Phi, \tag{3.13}$$

which can be more compactly written as $y_t = i\mathbf{A}(t)y$, where $\mathbf{A}(t) = \Lambda + u(t)\mathbf{X}$.

The Crank–Nicolson method

$$\left(I - \frac{i\delta t}{2}[\mathbf{A}_k + \mathbf{A}_{k-1}]\right)y_k = \left(I + \frac{i\delta t}{2}[\mathbf{A}_k + \mathbf{A}_{k-1}]\right)y_{k-1} \tag{3.14}$$

is one of the more commonly used schemes to numerically integrate the TDSE. It is important to notice, however, that it is only symplectic when the potential is constant over a each time step. This is easily rectified by using the modified Crank–Nicolson method [14], where the control at the endpoints is replaced by the time averaged control over the time step. This approximation retains second order accuracy while making the scheme symplectic. Symplecticity is quite important in quantum control problems as the cost functional can be changed arbitrarily due to numerical loss or gain in the state equation solver otherwise.

4. Control problem formulation

Now that the state equation has been discretized in both state and time, the optimal control problem is finite dimensional. The goal is to find the control vector u which is defined on the grid so as to maximize the projection of the state onto the target at the final time. This is formulated as

$$\min_u J(y, \bar{y}, u) = 1 - \bar{y}_n^\top\mathbf{P}y_n + \frac{\gamma}{2}u^\top\mathbf{W}u, \tag{4.1}$$

where $\bar{y}$ is the complex conjugate of y, $\mathbf{P}$ is the orthogonal projector onto the target, $0 < \gamma \ll 1$ is a regularization parameter, and $\mathbf{W}$ is the symmetric positive matrix

$$\mathbf{W}_{jk} = \begin{cases} \frac{2\delta t}{3} + \frac{2\epsilon}{\delta t} & \text{if } j = k \\ \frac{\delta t}{6} - \frac{\epsilon}{\delta t} & \text{if } j = k \pm 1 \\ 0 & \text{otherwise} \end{cases} \tag{4.2}$$

such that $u^\top \mathbf{W} u$ is a second order approximation of an H^1 type of inner product such as $(u, u) + \epsilon(\dot{u}, \dot{u})$, where ϵ is a small positive parameter. Penalizing the derivative of the control enforces the condition that the control go continuously to zero at $t = 0$ and $t = T$. In the numerical experiments, the value $\epsilon = 10^{-3}$ was used.

The fully discretized Schrödinger equation (3.14) provides an equality constraint for every time step, namely that $e_k(y_k, y_{k-1}, u_k, u_{k-1}) = 0$ for $k = 1, \ldots, m$. A Lagrange multiplier is needed for each time step to enforce each equality constraint as well as its complex conjugate. The Lagrangian is

$$L(y, \bar{y}, u, \lambda, \bar{\lambda}) = J(y, \bar{y}, u) + \sum_{k=1}^{m} \lambda_k^\top e_k + \bar{\lambda}_k^\top \bar{e}_k. \qquad (4.3)$$

Taking variations with respect to each of the arguments and setting them to zero gives the first-order optimality conditions. For compact representation, let $B_k = I - \frac{i\delta t}{2}[A_k + A_{k-1}]$. Then we can write the optimality system as

$$\begin{cases} \mathbf{B}_k y_k = \mathbf{B}_k^* y_{k-1}, \quad y_0 \text{ given}, \\ \mathbf{B}_k \lambda_k = \mathbf{B}_{k+1}^* \lambda_{k+1}, \quad \lambda_N = \mathbf{P}\bar{y}_m, \\ \nabla \tilde{J}(u) = Wu - \frac{\delta t}{2}\text{Im}[\xi] = 0, \\ \xi_k = \lambda_k^\top \mathbf{X}(y_k + y_{k-1}) + \lambda_{k+1}^\top \mathbf{X}(y_{k+1} + y_k). \end{cases} \qquad (4.4)$$

We can formulate a reduced cost functional by using the fact that the state variable is an implicit function of the control $\tilde{J}(u) = J(y(u), \bar{y}(u), u)$ The control equation in (4.4) expresses the condition that the reduced gradient $\nabla \tilde{J}(u) = 0$.

4.1. Newton's method

The Lagrangian is not an analytic function of the state and adjoint variables since it also depends on their complex conjugates. Consequently, to compute the Hessian, we use the Wirtinger calculus [8]. In the Wirtinger calculus representation the Hessian is obtained by computing the Jacobian of the complex conjugate of the gradient, so that the Hessian is a complex-valued Hermitian matrix. In particular, L_{ab} really means $\partial_a(\partial_b L)^*$ which is equal to $(L_{ba})^*$. Taking second variations gives rise to the KKT system

$$\begin{pmatrix} L_{yy} & 0 & L_{yu} & 0 & L_{y\bar{\lambda}} \\ 0 & L_{\bar{y}\bar{y}} & L_{\bar{y}u} & L_{\bar{y}\lambda} & 0 \\ L_{uy} & L_{u\bar{y}} & L_{uu} & L_{u\lambda} & L_{u\bar{\lambda}} \\ 0 & L_{\lambda\bar{y}} & L_{\lambda u} & 0 & 0 \\ L_{\bar{\lambda}y} & 0 & L_{\bar{\lambda}u} & 0 & 0 \end{pmatrix} \begin{pmatrix} \delta y \\ \delta \bar{y} \\ \delta u \\ \delta \lambda \\ \delta \bar{\lambda} \end{pmatrix} = - \begin{pmatrix} 0 \\ 0 \\ L_u \\ 0 \\ 0 \end{pmatrix}. \qquad (4.5)$$

From the KKT system, we can formally write relationship between the differential change and state and adjoint variables due to a differential change in the control:

$$\begin{aligned} \delta y &= -L_{\bar{\lambda}y}^{-1} L_{\bar{y}u} \delta u, \\ \delta \lambda &= -L_{\bar{y}\lambda}^{-1} \left[L_{\bar{y}u} \delta u + L_{\bar{y}\bar{y}} \delta \bar{y} \right]. \end{aligned} \qquad (4.6)$$

From this we can write δy and $\delta \lambda$ as the solutions of forced difference equations similar to those for y and λ.

$$\mathbf{B}_k \delta y_k = \mathbf{B}_k^* \delta y_{k-1} + \frac{i\delta t}{2}(\delta u_k + \delta u_{k-1})\mathbf{X}(y_k + y_{k-1}),$$
$$\mathbf{B}_k \delta \lambda_k = \mathbf{B}_{k+1}^* \delta \lambda_{k-1} + \frac{i\delta t}{2}(\delta u_k + \delta u_{k+1})\mathbf{X}(\lambda_k + \lambda_{k+1}), \tag{4.7}$$

where the mth time step for $\delta \lambda$ will also contain an additional term $\mathbf{P}\delta y_m$ on the right hand side.

The action of the reduced Hessian on a test vector δu is

$$[\nabla^2 \tilde{J}(u)]\delta u = L_{uu}\delta u + 2\text{Re}[L_{uy}\delta y + L_{u\lambda}\delta \lambda]. \tag{4.8}$$

The Newton search direction can now be computed by iteratively solving the equation

$$[\nabla^2 \tilde{J}(u)]\delta u = -\nabla \tilde{J}(u). \tag{4.9}$$

Since the cost functional is nonconvex, typically even the reduced Hessian is indefinite and the standard conjugate gradient method will not converge. Instead, we use the symmetric LQ (SYMMLQ) method. The computed δu which approximately satisfies (4.9) may not be a descent direction. To handle this possibility, define our descent direction as

$$p = \begin{cases} \delta u & \text{if } \delta u^\top \nabla \tilde{J}(u) < 0, \\ -\delta u & \text{if } \delta u^\top \nabla \tilde{J}(u) > 0. \end{cases} \tag{4.10}$$

Of course, should the Newton direction be an ascent direction, it could be discarded in favor of the usual steepest descent direction, however, the scaling of directions produced by solving the Hessian equation tends to be much better. That is to say that the step lengths for sufficient decrease usually remain order 1 instead of order 10^4. Consequently, in our experience, using the sign-flipped Newton direction tends to expedite the line search.

Since the cost functional is nonconvex, a line search strategy is needed to globalize the Newton method. Here we make the observation that the cost functional contains two terms: the physical tracking term $1 - \bar{y}_m^\top \mathbf{P} y_m$ which is uniformly bounded between 0 and 1 and the regularization term which is a pure quadratic, consequently once a descent direction p is computed, the reduced cost functional $\tilde{J}(u + \alpha p)$ is an asymptotically quadratic function. This means that the cost functional along the search direction can be bounded from below by the quadratic polynomial

$$\tilde{J}(u + \alpha p) \geq d_2 \alpha^2 + d_1 \alpha + d_0, \quad \begin{aligned} d_0 &= \tfrac{\gamma}{2} u^\top \mathbf{W} u - \tilde{J}(u) \leq 0, \\ d_1 &= \gamma u^\top \mathbf{W} p, \\ d_2 &= \tfrac{\gamma}{2} p^\top \mathbf{W} p. \end{aligned} \tag{4.11}$$

Since $d_0 \leq 0$, the quadratic equation $d_2\alpha^2 + d_1\alpha + d_0 = 0$ has real roots and we can establish an upper bound on the largest feasible step length α that

can still possibly reduce the cost:

$$\alpha_{\max} = \frac{\sqrt{d_1^2 - 4d_0 d_2} - d_1}{2d_2}. \qquad (4.12)$$

The local minimizer is now guaranteed to satisfy $\alpha^* \in [0, \alpha_{\max}]$.

The strong Wolfe conditions that the step length α must satisfy to give a sufficient decrease in the cost and magnitude of the directional derivative are

$$\tilde{J}(u + \alpha p) \leq \tilde{J}(u) + c_1 \alpha p^\top \nabla \tilde{J}(u), \quad 0 < c_1 \ll 1,$$
$$|p^\top \nabla \tilde{J}(u + \alpha p)| \leq c_2 |p^\top \nabla \tilde{J}(u)|, \quad c_1 < c_2 < 1. \qquad (4.13)$$

In the numerical experiments, the values $c_1 = 10^{-4}$ and $c_2 = 0.5$ were used.

Data: Given a descent direction p and the function
$\qquad f(\alpha) = \tilde{J}(u + \alpha p)$ and $f'(\alpha) = p^\top \nabla \tilde{J}(u + \alpha p)$
Compute $\alpha_{\max}$ based on (4.12)
if $\alpha_{max} > 2$ **then**
$\quad$ Evaluate $f(1)$ and $f'(1)$;
$\quad$ **if** $\alpha = 1$ *satisfies* (4.13) **then**
$\quad\quad$ $\alpha^* \leftarrow 1$;
$\quad$ **else**
$\quad\quad$ Construct cubic model on $[0, 1]$ and compute its minimum α_m;
$\quad\quad$ Evaluate $f(\alpha_m)$ and $f'(\alpha_m)$
$\quad\quad$ **if** $\alpha = \alpha_m$ *satisfies SWC* **then**
$\quad\quad\quad$ $\alpha^* \leftarrow \alpha_m$;
$\quad\quad$ **else**
$\quad\quad\quad$ **if** $[0, \alpha_m]$ *brackets a minimum* **then**
$\quad\quad\quad\quad$ $\alpha_r \leftarrow \alpha_m$;
$\quad\quad\quad$ **else if** $[0, 1]$ *brackets a minimum* **then**
$\quad\quad\quad\quad$ $\alpha_r \leftarrow 1$;
$\quad\quad\quad$ **else**
$\quad\quad\quad\quad$ $\alpha_r \leftarrow \alpha_{max}$;
$\quad\quad\quad$ **end**
$\quad\quad\quad$ $a^* \leftarrow \text{bisect}(0, a_r)$.
$\quad\quad$ **end**
$\quad$ **end**
else
$\quad$ $\alpha^* \leftarrow \text{bisect}(0, \alpha_{max})$
end

Algorithm 1: Line search algorithm

Data: α_l and α_r which bracket a minimum point. $L = \alpha_r - \alpha_l$.
$\quad\quad f(\alpha) = \tilde{J}(u + \alpha p)$ and $f'(\alpha) = p^\top \nabla \tilde{J}(u + \alpha p)$
while $L > tol$ **do**
$\quad$ Compute the midpoint $\alpha_m = \frac{1}{2}(\alpha_l + \alpha_r)$ and evaluate $f(\alpha_m)$
$\quad$ and $f'(\alpha_m)$
$\quad$ **if** α_m *satisfies* (4.13) **then**
$\quad\quad \mid \quad \alpha^* \leftarrow \alpha_m$;
$\quad$ **end**
$\quad$ **if** $f'(\alpha_l) < 0$ *and either* $f'(\alpha_r) > 0$ *or* $f(\alpha_r) > f(\alpha_l)$ **then**
$\quad\quad \mid \quad \alpha_r \leftarrow \alpha_m$;
$\quad$ **else if** $f'(\alpha_l) > 0$ *and* $f'(\alpha_r) < 0$ *or* $f(\alpha_r) < f(\alpha_l)$ **then**
$\quad\quad \mid \quad \alpha_r \leftarrow \alpha_m$;
$\quad$ **else**
$\quad\quad \mid \quad \alpha_l \leftarrow \alpha_m$;
$\quad$ **end**
$\quad$ $L \leftarrow (\alpha_r - \alpha_l)$
end

Algorithm 2: Bisection minimizer

4.2. State model reduction

Although as a general rule of thumb, the occupancy probability of the lowest energy states are going to be greatest, we can estimate how strong the control potential couples any two states by writing down the state equation in the interaction picture. Given stationary states $\phi_1, \phi_2, \ldots$ and corresponding eigenvalues $\lambda_1, \lambda_2, \ldots$, we can introduce a time dependent change of basis

$$y(t) = \exp(-i\Lambda t)z(t). \tag{4.14}$$

Using this as an Ansatz in our state equation gives the new equation for the transformed state $z(t)$ as

$$\dot{z}(t) = -iu(t)\exp(i\Lambda t)\mathbf{X}\exp(-i\Lambda t)z(t), \tag{4.15}$$

where we can think of the time-dependent similarity transformed matrix $\mathbf{X}$ as being an interaction matrix

$$\tilde{\mathbf{X}}(t) = \exp(i\Lambda t)\mathbf{X}\exp(-i\Lambda t), \tag{4.16}$$

and the specific elements of this matrix are $\tilde{\mathbf{X}}_{jk}\exp(i\omega_{jk}t)$ where $\omega_{jk} = \lambda_j - \lambda_k$. In integral form, the transformed solution at a time t is

$$z(t) = z(0) - i\int_0^t u(\tau)\tilde{\mathbf{X}}(\tau)z(\tau)\,d\tau. \tag{4.17}$$

Of course, $u(t)$ and $z(t)$ are not known in advance; however, $\tilde{\mathbf{X}}(t)$ is known and integration acts as a lowpass filter. In particular, integrating the interaction

matrix gives the elements

$$\hat{\mathbf{X}}_{jk}(t) = \int\limits_{0}^{t} \tilde{\mathbf{X}}_{jk}(\tau) \, d\tau = \tilde{\mathbf{X}}_{jk}(t) \frac{\exp(i\omega_{jk}t) - 1}{i\omega_{jk}}. \qquad (4.18)$$

The magnitude of the $\hat{\mathbf{X}}_{jk}$ gives a rough sense of how strongly the interaction couples state $|j\rangle$ to state $|k\rangle$. Namely, the larger this element is, the more readily we can expect a particle that starts in state $|j\rangle$ to transfer into $|k\rangle$ at some point in time. It stands to reason then, that when considering which basis functions play the most significant role in the dynamics of wavefunction, we should consider not just the difference in eigenvalues, but also the interaction strength.

Once the eigenfunctions are known, the basic idea of the reduced model method is sort the eigenfunctions in decreasing importance to the dynamics, start with only the first few, and then compute the optimal control given that state basis. Once the optimal control is known, the state basis is then enlarged by adding the next few most important states and repeating this process until enlarging the state space no longer has a perceptible effect on the cost functional.

Consider as an example, the problem of two-particle with the control term $u(t)x$, after discretizing and computing the eigenfunctions, we obtain a matrix $\mathbf{X}$ as in (3.13). The sparsity pattern for this matrix is shown in the left side of Figure 4. In general, this matrix will be full and this sparsity pattern is a consequence of special parity properties of this problem, however, it is easier to draw the connectivity graph when some states are not directly connected to others. The connectivity graphs shows that, while it is possible to go from any state to any other state, state 1 is not directly connected to state 3, so a particle must go through an intermediate state, such as 2, 5, or 7 first.

The states most strongly directly coupled to the ith state are indicated by the largest elements of the vector

$$r_1 = |e_i + \hat{\mathbf{X}}e_i|, \qquad (4.19)$$

where e_i is the ith canonical vector. By extension, the states most strongly coupled to state i in two steps will be the largest elements of the vector

$$r_2 = |e_i + \hat{\mathbf{X}}e_i + \hat{\mathbf{X}}^2 e_i|. \qquad (4.20)$$

Assuming that any number of intermediate states are allowed, we have the ranking vector

$$r_\infty = \left| \sum_{j=0}^{\infty} \hat{\mathbf{X}}^j e_i \right| = |(I - \hat{\mathbf{X}})^{-1} e_i|, \qquad (4.21)$$

where we are guaranteed that $(I - \hat{\mathbf{X}})^{-1}$ exists since $\hat{\mathbf{X}}$ is skew Hermitian. Recall that the interaction matrix in (4.18) contained an arbitrary phase shift term of $e^{i\omega_{jk}t}$. Although intuitively, this term should be neglected, we did consider the ranking the state importance both with and without this term
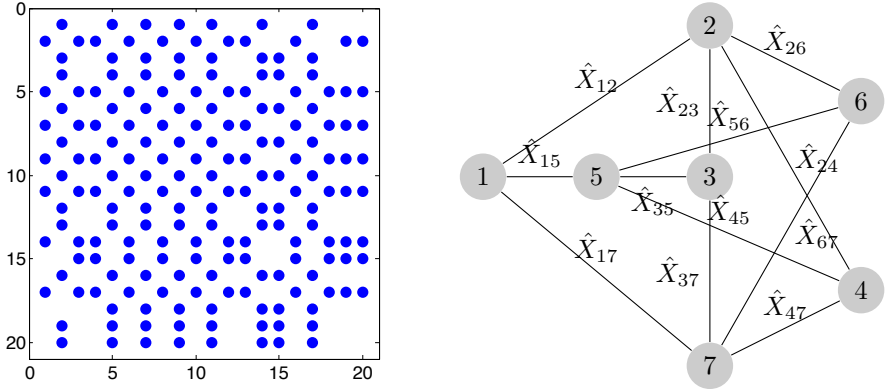
FIGURE 4. Left: Sparsity pattern for **X** for two particles in
a quantum well. Right: The connectivity graph for the first
seven states.

and found that the reduced model method converges more rapidly when it is
neglected.

5. Numerical results

In the test cases, a weak interaction potential (2.12) with unit charge $q = 1$
and smoothing factor $\delta = 0.1$ was used. For both the two and four particle
case, 400 uniform time steps were used and the final time was taken to be
$T = 1$ for the two particle case and $T = 2$ for the four particle case. For two
interacting particles, the order of the eigenstates as ranked by the interaction
criteria, excluding the complex exponential term, from strongest to weakest
coupling is

$$\{1, 2, 7, 5, 3, 14, 11, 8, 4, 10, 9, 16, 12, 20, 15, 13, 6, 17, 18, 19\}$$

when $N_s = 20$ in (3.12).

To compute the control, start with the first five eigenvectors ($N_s = 5$) with indices $\{1, 2, 7, 5, 4\}$ and compute a minimizer by conducting line
searches in the Newton search directions. When a local minimizer is ob-
tained, the state space is augmented to include the first ten modes ($N_s = 10$)
with indices $\{1, 2, 7, 5, 3, 14, 11, 8, 4, 10\}$. The computed control is no longer a
minimizer for the state constraint as there are now allowed transitions into
higher states, which increases the objective function. Using the previously
computed optimal control from the five dimensional state space as an initial
guess, we again compute a sequence of line searches in the Newton directions
until we have a new minimizer. This process of augmenting the state space
and minimizing until augmenting the space does not noticeably increase the
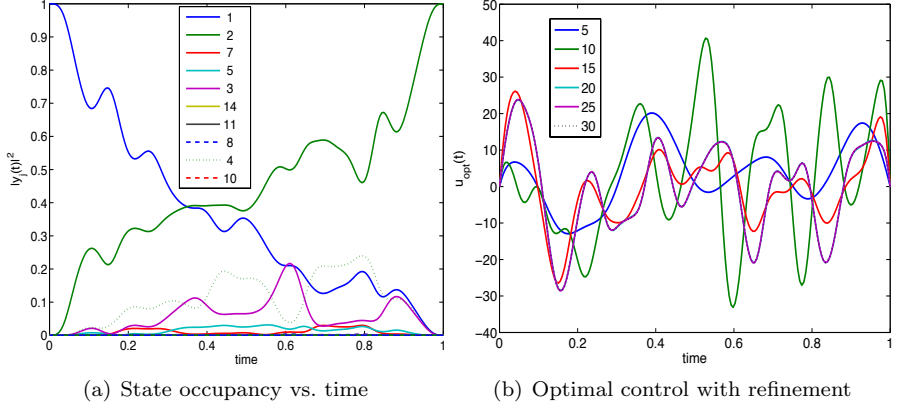objective function.

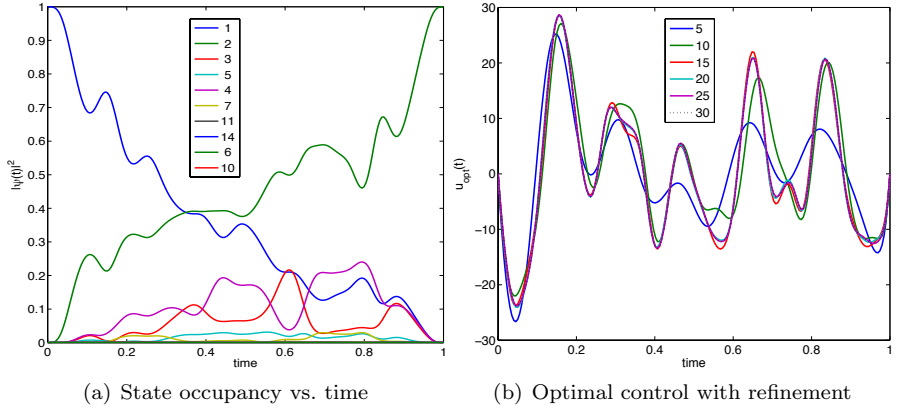(a) State occupancy vs. time   (b) Optimal control with refinement

FIGURE 5. *Left:* State transition $|1, 2\rangle \rightarrow |1, 3\rangle$ for two interacting particles. *Right:* The computed optimal control for $5, 10, 15, 20, 25, 30$ eigenvectors using the ranking which includes the complex exponential term.

(a) State occupancy vs. time   (b) Optimal control with refinement

FIGURE 6. *Left:* State transition $|1, 2\rangle \rightarrow |1, 3\rangle$ for two interacting particles. *Right:* The computed optimal control for $N_s = 5, 10, 15, 20, 25, 30$ modes (eigenvectors) without the complex exponential term. The controls for $Ns = 20$ and $Ns = 30$ appear to be identical.

When the complex phase factor in (4.18) is included in determining the ranking, we see that the optimal control in Figure 5(b) changes significantly as the number of modes N_s increases. When this phase term is excluded, as is the case in Figure 10(b), the optimal control as a function of N_s stabilizes rapidly. In fact, after $N_s = 25$ there is no significant effect on the cost by further augmenting the space and moreover, the optimal control for $N_s = 30$
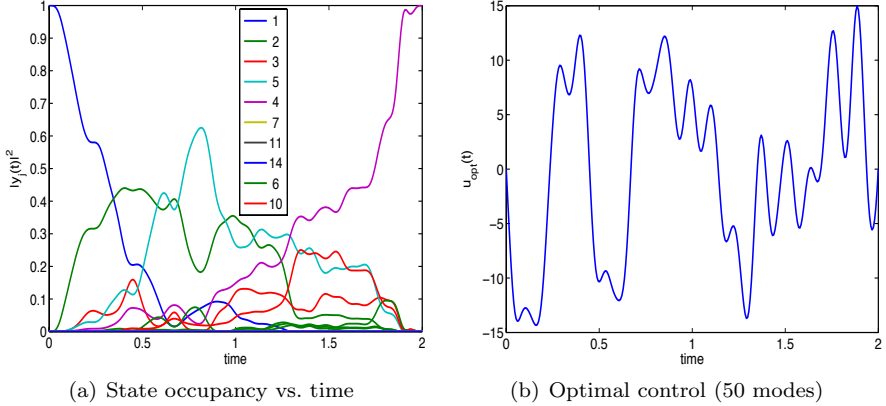
(a) State occupancy vs. time

(b) Optimal control (50 modes)

FIGURE 7. *Left:* State transition $|1, 2\rangle \to |1, 4\rangle$ for two interacting particles. *Right:* The computed optimal control for $N_s = 30$ modes.

is not visually distinct from the optimal control for $N_s = 25$. This shows that this ordering strategy is more effective at determining which eigenstates play the most important role in the dynamics. The occupancy of the states shown in Figures 5(a) and 10(a), is practically identical. It is important to note that the results become effectively discretization independent after twenty or thirty modes are used, which is a significant savings over the full space degrees of freedom $N_p = \binom{15}{2} = 105$.

Thirty modes is not sufficient to resolve the transition from the first state $|1, 2\rangle$ to the third state $|1, 4\rangle$ as see in Figure 7(a) and 7(b) due to the significantly higher energy in the control needed to make the transition, as evidenced by the higher frequency terms in the optimal control. Instead fifty modes were needed to adequately resolve the dynamics. More of the states in 7(a) have a significant occupation probability than was the case with 10(a).

The problem becomes more challenging as additional particles are added, since the energy spacing of the eigenstates increases considerably. In the four particle system, twenty Lagrange basis functions per particle were needed for the eigenstates to be resolved. This means that before reduction the state has dimension $N_p = \binom{20}{4} = 4845$. However, to compute optimal controls for the transitions between the first state $|1, 2, 3, 4\rangle$ and the second $|1, 2, 3, 5\rangle$ taking the first fifty modes was sufficient. This transition is shown in Figure 8. Exciting the system to the third state $|1, 2, 3, 6\rangle$ proved quite difficult as at the final time there was only a 97% probability of finding the particles in the desired state. We also see that the H^1 norm of the control is becoming quite large and to resolve this problem finer grids will be needed.

In Figure 10, the reduction of the cost functional and of the L^2 norm of the gradient is shown for the problem of two interacting fermions making the transition from the state $|1, 2\rangle$ to the state $|1, 3\rangle$. The globalized Newton

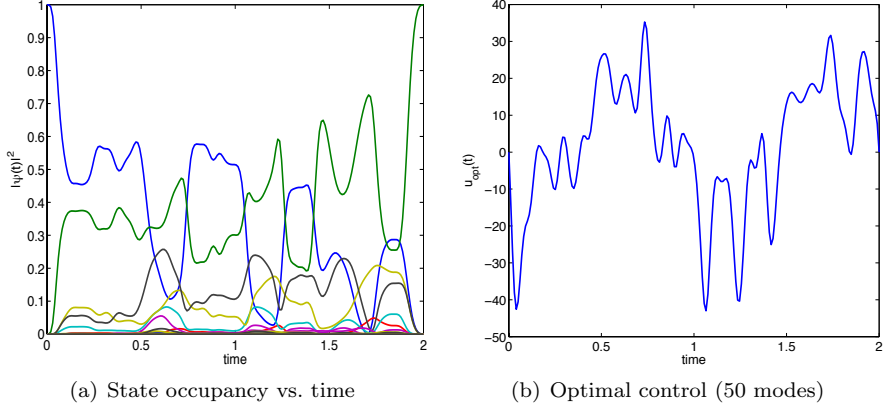(a) State occupancy vs. time   (b) Optimal control (50 modes)

FIGURE 8. *Left:* State transition $|1, 2, 3, 4\rangle \rightarrow |1, 2, 3, 5\rangle$ for four interacting particles. *Right:* The computed optimal control for $N_s = 50$ modes.

(a) State occupancy vs. time   (b) Optimal control (50 modes)
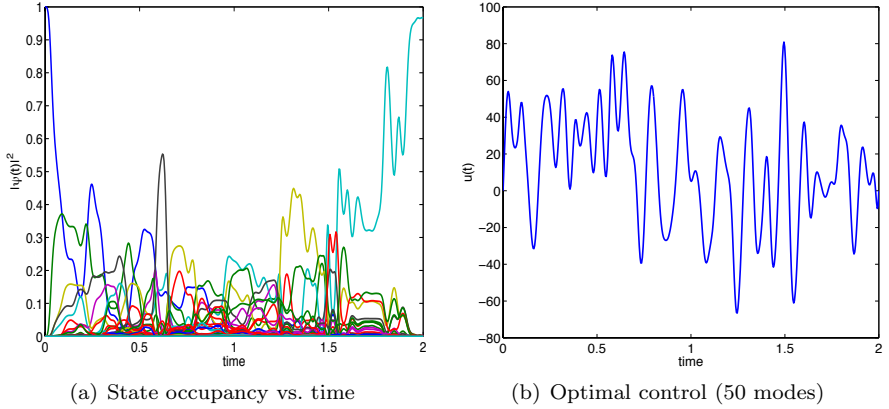
FIGURE 9. *Left:* State transition $|1, 2, 3, 4\rangle \rightarrow |1, 2, 3, 6\rangle$ for four interacting particles. *Right:* The computed optimal control for $N_s = 50$ modes.

method is started with a state space of dimension $N_s = 5$, and after 100 iterations, the state space is repeatedly augmented by more sorted eigenfunctions until further augmentation does not appreciably change the cost functional.

After numerous tests with differing interaction terms, more particles, initial and final states, state space dimension, and time scale, there does not appear to be any completely typical pattern with respect to how often the unit step length of the Newton method yields sufficient decrease and when a line search is needed. It can be said that generally, parameters which make the
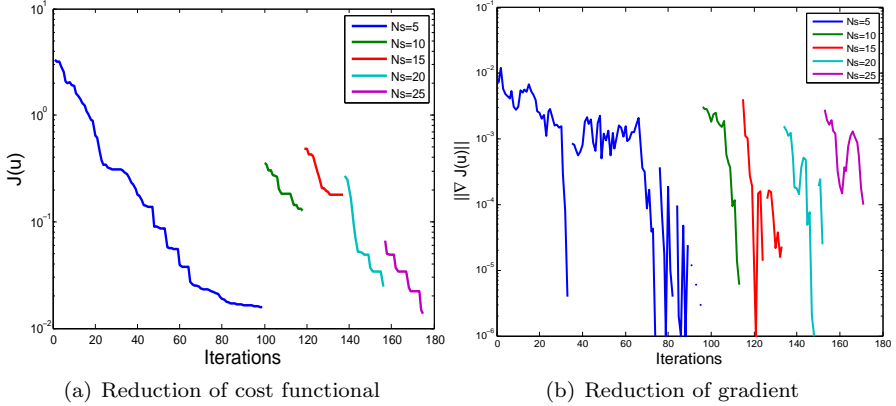
(a) Reduction of cost functional   (b) Reduction of gradient

FIGURE 10. *Left:* Cost functional value reduction for the state transition $|1,2\rangle \rightarrow |1,3\rangle$ for two interacting particles. *Right:* Reduction of L^2-norm of the gradient.

minimization problem harder are those which increase the energy separation between initial and final states, and this tends to require more line searches.

6. Conclusion

We have presented an efficient discretization method and optimization method for controlling energy state transitions for multiple fermions. It was observed that the computational effort of the problem can be reduced by projecting the state onto a suitable reduced basis, which is then augmented as needed to adequately resolve the dynamics. Since optimal control and corresponding state are smooth functions, in future work, higher order symplectic methods will be used to discretize in time so that the number of degrees of freedom in the optimization problem of computing the optimal control can be reduced.

References

1. Alfio Borzì, Georg Stadler, and Ulrich Hohenester, *Optimal quantum control in nanostructures: Theory and application to a generic three-level system*, Phys. Rev. A **66** (2002), 053811.

2. Ugo Boscain, Gregoire Charlot, Jean-Paul Gauthier, Stephane Guerin, and Hans-Rudolf Jauslin, *Optimal control in laser-induced population transfer for two- and three-level quantum systems*, Journal of Mathematical Physics **43** (2002), no. 5, 2107–2132.

3. Claudio G Canuto, M. Yousuff Hussaini, Alfio Quarteroni, and Thomas A. Zang, *Spectral Methods*, Springer, Heidelberg, 2007.

4. Christian Clason and Gregory von Winckel, *A general spectral method for the numerical simulation of one-dimensional interacting fermions*, Computer Physics Communications **183** (2012), no. 2, 405 – 417.

5. Matthew Grace, Constantin Brif, Herschel Rabitz, Ian A Walmsley, Robert L Kosut, and Daniel A Lidar, *Optimal control of quantum gates and suppression of decoherence in a system of interacting two-level particles*, Journal of Physics B: Atomic, Molecular and Optical Physics **40** (2007), no. 9, S103.

6. Julian Grond, Gregory von Winckel, Jörg Schmiedmayer, and Ulrich Hohenester, *Optimal control of number squeezing in trapped Bose-Einstein condensates*, Phys. Rev. A **80** (2009), 053625.

7. Katherine Kime, *Finite difference approximation of control via the potential in a 1-d Schrödinger equation*, Electronic Journal of Differential Equations **2000** (2000), no. 26, 1–10.

8. Ken Kreutz-Delgado, *The complex gradient operator and the CR-calculus*, 09064835 (2009), 1–74.

9. Per-Olov Löwdin, *Quantum theory of many-particle systems. I. Physical interpretations by means of density matrices, natural spin-orbitals, and convergence problems in the method of configurational interaction*, Phys. Rev. **97** (1955), no. 6, 1474–1489.

10. Yvon Maday and Gabriel Turinici, *New formulations of monotonically convergent quantum control algorithms*, The Journal of Chemical Physics **118** (2003), no. 18, 8191–8196.

11. Michael Mundt and David J Tannor, *Optimal control of interacting particles: a multi-configuration time-dependent Hartree–Fock approach*, New Journal of Physics **11** (2009), no. 10, 105038.

12. Anthony P. Peirce, Mohammed A. Dahleh, and Herschel Rabitz, *Optimal control of quantum-mechanical systems: Existence, numerical approximation, and applications*, Phys. Rev. A **37** (1988), 4950–4964.

13. D. Sugny, C. Kontz, and H. R. Jauslin, *Time-optimal control of a two-level dissipative quantum system*, Phys. Rev. A **76** (2007), no. 2, 023419.

14. Vidar Thomée, *Galerkin Finite Element Methods for Parabolic Problems*, Springer Series in Computational Mathematics, Springer-Verlag, Secaucus, NJ, USA, 2006.

15. Gregory von Winckel and Alfio Borzì, *Computational techniques for a quantum control problem with H^1-cost*, Inverse Problems **24** (2008), no. 3, 034007.

16. Gregory von Winckel and Alfio Borzì, *Qucon: A fast Krylov–Newton code for dipole quantum control problems*, Computer Physics Communications **181** (2010), no. 12, 2158 – 2163.

17. Gregory von Winckel, Alfio Borzì, and Stefan Volkwein, *A globalized Newton method for the accurate solution of a dipole quantum control problem*, SIAM J. Sci. Comput. **31** (2009), no. 6, 4176–4203.

Gregory von Winckel
Previous address:
Institut für Mathematik und Wissenschaftliches Rechnen
Karl-Franzens-Universität Graz
36 Heinrichstraße
A-8010 Graz
Austria

Current address:
Center for High Technology Materials
University of New Mexico
1313 Goddard Street SE
Albuquerque, New Mexico 87106
United States
e-mail: `gregvw@chtm.unm.edu.at`