Available online at www.sciencedirect.com
ScienceDirect
Journal of Computational and Applied Mathematics 216 (2008) 170–197
JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS
www.elsevier.com/locate/cam

Formulation and numerical solution of finite-level quantum optimal control problems☆

A. Borzì[a,*], J. Salomon[b], S. Volkwein[a]

[a] *Institut für Mathematik und Wissenschaftliches Rechnen, Karl-Franzens-Universität Graz, Heinrichstr. 36, A-8010 Graz, Austria*
[b] *CEREMADE, Universite' Paris-Dauphine, France*

Received 19 April 2006; received in revised form 27 April 2007

Abstract

Optimal control of finite-level quantum systems is investigated, and iterative solution schemes for the optimization of a control representing laser pulses are developed. The purpose of this external field is to channel the system's wavefunction between given states in its most efficient way. Physically motivated constraints, such as limited laser resources or population suppression of certain states, are accounted for through an appropriately chosen cost functional. First-order necessary optimality conditions and second-order sufficient optimality conditions are investigated. For solving the optimal control problems, a cascadic non-linear conjugate gradient scheme and a monotonic scheme are discussed. Results of numerical experiments with a representative finite-level quantum system demonstrate the effectiveness of the optimal control formulation and efficiency and robustness of the proposed approaches.

MSC: 35Q40; 49A10; 49K10; 49M05; 65M06; 81V80

Keywords: Quantum systems; Optimal control theory; Optimality conditions; Monotonic schemes; Cascadic acceleration; Non-linear conjugate gradient method

1. Introduction

Nowadays we witness a large growing interest in controlling quantum phenomena in a variety of application systems [9,15,29–32]. Present and perspective applications range from quantum optics and quantum chemistry to semiconductor nanostructures. In the last few years these research areas have received further impetus from the emerging fields of quantum computation and quantum communication [7], aiming at quantum devices where there is the need to manipulate wavefunctions with highest possible precision.

This high-fidelity quantum-state engineering can only be achieved putting together the most sophisticated experimental and theoretical techniques for control of quantum systems. However, within each field of application one has come up with its own strategies and it is only recently that a common consensus has arisen toward the use of optimal control theory [10,13,21].

☆ Supported in part by the Austrian Science Fund FWF project P18136-N13 "Quantum optimal control of semiconductor nanostructures".
* Corresponding author.
E-mail addresses: alfio.borzi@uni-graz.at (A. Borzì), julien.salomon@dauphine.fr (J. Salomon), stefan.volkwein@uni-graz.at (S. Volkwein).

0377-0427/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.cam.2007.04.029

In the optimal control framework, one starts by defining the optimality criteria in the form of a cost functional. For a desired quantum-state transition, this functional will depend on the final state, the need to suppress population of certain states during the control process, as well as other physically motivated constraints, e.g., limited laser resources. The strategy is then to minimize this cost functional while satisfying the constraints of the underlying dynamic equations governing the evolution of quantum states, e.g., the Schrödinger equation. The calculation of the necessary optimality conditions for this optimization problem results in a system of coupled equations to be solved.

While we focus on quantum optimal control problems we argue that many of the results of this paper can be extended to general time-dependent bilinear control problems. Bilinear systems [12,25] were introduced in the theory of automatic control in the 1960s for electrical engineering applications. They represent a class of non-linear control strategies with the aim to obtain better system response than possible with linear control. In general, the solution of most bilinear systems poses challenging theoretical and computational problems which are open or have been only partially addressed. This is in particular true for the control of the quantum mechanical systems discussed in this paper.

The purpose of this paper is to present a detailed formulation of a class of optimal control problems for finite-level quantum systems and to address their solution by iterative methods. We prove existence of solutions to the optimal control problems and investigate first-order necessary optimality conditions and second-order sufficient optimality conditions. We review modern monotonic iterative schemes and their convergence properties and use these schemes as benchmark for an alternative solution procedure that we propose in this paper. This procedure results from combining an appropriate extension of a newly proposed non-linear conjugate gradient (NCG) method with a cascadic acceleration scheme. Convergence of the proposed NCG method is proved, and its competitiveness in efficiency and robustness is demonstrated by results of numerical experiments.

In the following section, we introduce the class of finite-level quantum optimal control problems considered in this paper. Within an appropriate functional analytical setting, existence of at least one global optimal solution is proved. We discuss existence of Lagrange multiplier and first-order necessary conditions for a minimum. Second-order sufficient optimality conditions are also studied that allow to characterize local minima.

In Section 3, the proposed NCG scheme is formulated. Under appropriate less restrictive assumptions, we prove convergence of this scheme to a local minimizer. This scheme is embedded in a cascadic iteration to obtain almost optimal computational complexity.

To validate the computational performance of our cascadic non-linear conjugate gradient scheme (C-NCG), comparison with efficient monotonic schemes is presented showing that the former may outperform the latter. For completeness and for comparison of theoretical aspects, in Section 4 a detailed review of newly proposed monotonic schemes is given.

In Section 5, results of an extended set of numerical experiments are given. We show that the optimal solution is quite sensitive to the required tolerance of the norm of the gradient. We therefore discuss additional convergence criteria involving the order of accuracy of solutions. Further numerical experiments demonstrate efficiency of the C-NCG approach and its robustness with respect to change of values of the optimization parameters.

A section of conclusion completes the exposition of our work.

2. Quantum optimal control problems

This section is devoted to the formulation of a class of finite-level quantum optimal control problems, which are the subject of our investigations. Existence of solutions to the optimal control problems is proved. We investigate first-order necessary optimality conditions addressing existence and regularity issues of the Lagrange multipliers and, correspondingly, of the control functions. To characterize a solution to the first-order optimality conditions that corresponds to a local minimum, second-order sufficient optimality conditions are discussed.

To some extent the analysis presented in this section can be shortened utilizing results from the literature. However, we make use of specific properties of the underlying optimal control problem like the bilinear structure of the equality constraints and the fact that the variables (including the control function) are complex-valued. This is a situation which is not common in the optimal control literature and, therefore, we report all details of the proofs.

2.1. The minimization problem

We consider localized finite-level quantum systems modeled by a Schrödinger equation for an n-component wave function $\psi : [0, T] \to \mathbb{C}^n$ as follows:

$$\mathrm{i}\dot{\psi}(t) = H(\varepsilon(t))\psi(t) \quad \text{for } t \in (0, T] \quad \text{and} \quad \psi(0) = \psi_0, \tag{1}$$

where $T>0$ is a given terminal time, $H:\mathbb{C}\to\mathbb{C}^{n\times n}$ denotes the Hamiltonian matrix depending on the external control field $\varepsilon:[0,T]\to\mathbb{C}$ and $\psi_0\in\mathbb{C}^n$ is a fixed initial condition. The Hamiltonian $H=H_0+H_1(\varepsilon)$ has two constitutive components: the constant free Hamiltonian $H_0\in\mathbb{C}^{n\times n}$ describing the unperturbed (uncontrolled) system and $H_1:\mathbb{C}\to\mathbb{C}^{n\times n}$ modeling the coupling of the quantum state to an external control field ε.

The choice of T is a modeling issue motivated by physical considerations. Based on the quantum indeterminacy principle $\Delta E\Delta t\geqslant\hbar$ (where $\hbar$ is the Planck's constant that we set equal to one) we can state the following: for too small T a highly energetic optimal control results ($\Delta E\gg 1$) thus involving many energy levels. On the other hand, for too large T additional decoherence channels become important, which should be avoided; see, e.g., [15]. As a guideline, the choice of T should be related to the transition frequency.

Strictly speaking, the wavefunction description given in (1) is appropriate for an isolated quantum system and in that case the governing Hamiltonian H is hermitian. For a more realistic non-isolated system with environment couplings and subject to control, a more general density-matrix description would be required [35]. Alternatively, we follow the procedure outlined in [5] to construct a non-hermitian Hamiltonian H_0 accounting for environment losses.

We focus on localized quantum systems, where $H_1:\mathbb{C}\to\mathbb{C}^{n\times n}$ is hermitian and possesses the following structure:

$$H_1(z)=z_{\Re e}H_{1\Re e}+z_{\Im m}H_{1\Im m}\quad\text{for } z=z_{\Re e}+\mathrm{i}z_{\Im m}\in\mathbb{C},\tag{2}$$

where $H_{1\Re e}, H_{1\Im m}\in\mathbb{C}^{n\times n}$ are constants and $z_{\Re e}, z_{\Im m}\in\mathbb{R}$. Consequently,

$$\varrho(H_1(z))\leqslant K_0|z|\quad\text{for all } z\in\mathbb{C},\tag{3}$$

where $\varrho(\cdot)$ denotes the spectral norm of complex-valued $n\times n$ matrices, $|\cdot|$ stands for the absolute value of complex numbers and $K_0=2\max\{\varrho(H_{1\Re e}),\varrho(H_{1\Im m})\}\geqslant 0$. Moreover, by (3) we have

$$\varrho(H(z))\leqslant\varrho(H_0)+K_0|z|\quad\text{for all } z\in\mathbb{C}.$$

Remark 1. (a) A control Hamiltonian with the structure given in (2) and satisfying (3) is considered in the section of numerical experiments.

(b) Within quantum control theory, linear control Hamiltonians are of interest. They represent the first-order terms of multipole expansion of the charged particle–electric field interaction.

We call $\psi:[0,T]\to\mathbb{C}^n$ a *solution to* (1) if ψ belongs to $H^1(0,T;\mathbb{C}^n)$, $\psi(0)=\psi_0$ holds and ψ satisfies $\mathrm{i}\dot{\psi}=H(\varepsilon(\cdot))\psi$ in $[0,T]$ almost everywhere (a.e.). For the notion of Sobolev spaces we refer the reader to [1], for instance. Notice that (1) is a linear system of ordinary differential equations, which therefore admits a unique solution $\psi\in H^1(0,T;\mathbb{C}^n)$ for every $\varepsilon\in L^2(0,T;\mathbb{C})$. The Hilbert space $L^2(0,T;\mathbb{C}^n)$ is endowed with the canonical inner product

$$\langle\phi,\psi\rangle_{L^2(0,T;\mathbb{C}^n)}=\int_0^T\phi\cdot\psi^*\,\mathrm{d}t\quad\text{for }\phi,\psi\in L^2(0,T;\mathbb{C}^n),$$

where '*' means complex conjugate and the dot '·' denotes the usual vector–scalar product in $\mathbb{C}^n$. We also have $\|\cdot\|_{L^2(0,T;\mathbb{C}^n)}=\langle\cdot,\cdot\rangle^{1/2}$. Analogously, the inner product and the corresponding induced norm are defined on $H^1(0,T;\mathbb{C}^n)$.

To write (1) in a compact form, we define the Hilbert space

$$Y=L^2(0,T;\mathbb{C}^n)\times\mathbb{C}^n$$

endowed with the common product topology and introduce the non-linear operator $e=(e_1,e_2)$ by

$$e:H^1(0,T;\mathbb{C}^n)\times L^2(0,T;\mathbb{C})\to Y,\quad(\psi,\varepsilon)\mapsto\begin{pmatrix}\mathrm{i}\dot{\psi}-H(\varepsilon(\cdot))\psi\\ \psi(0)-\psi_0\end{pmatrix}.$$

Recall $H^1(0,T;\mathbb{C}^n)$ is continuously (even compactly) embedded in $C([0,T];\mathbb{C}^n)$, where $C([0,T];\mathbb{C}^n)$ denotes the Banach space of all continuous functions $\varphi:[0,T]\to\mathbb{C}^n$ that is endowed with the common norm, see [1]. Therefore, there exists an embedding constant $K_1>0$ satisfying

$$\|\psi\|_{C([0,T];\mathbb{C}^n)}\leqslant K_1\|\psi\|_{H^1(0,T;\mathbb{C}^n)}\quad\text{for all }\psi\in H^1(0,T;\mathbb{C}^n).\tag{4}$$

Let $(\psi, \varepsilon) \in H^1(0, T; \mathbb{C}^n) \times L^2(0, T; \mathbb{C})$. Then we infer from (3) that

$$|H(\varepsilon(t))\psi(t)|^2_{\mathbb{C}^n} \leqslant \varrho(H(\varepsilon(t)))^2 |\psi(t)|^2_{\mathbb{C}^n} \leqslant (2\varrho(H_0)^2 + 2K_0^2|\varepsilon(t)|^2)\|\psi\|^2_{C([0,T];\mathbb{C}^n)}$$

for almost all $t \in [0, T]$. Thus, $H(\varepsilon(\cdot))\psi$ belongs to $L^2(0, T; \mathbb{C}^n)$ and, therefore, e is well defined.

In the following we shall consider the problem of determining a control field $\varepsilon \in L^2(0, T; \mathbb{C})$, such that (1) is fulfilled and a number of optimality criteria are met. We require that the control sequence drives the system at time T close to the desired target state $\psi_d \in \mathbb{C}^n$. We account for limited laser resources and increased smoothness through a minimization of the control field strengths, and we require to suppress population of intermediate states which suffer strong environment losses, thus also enforcing our modeling choice of a non-hermitian free Hamiltonian. All these requirements are realized in the cost functional

$$J(\psi, \varepsilon) = \frac{1}{2}|\psi(T) - \psi_d|^2_{\mathbb{C}^n} + \frac{\gamma}{2}\|\varepsilon\|^2_{L^2(0,T;\mathbb{C})} + \frac{\mu}{2}\|\dot{\varepsilon}\|^2_{L^2(0,T;\mathbb{C})} + \frac{1}{2}\sum_{j\in I}\alpha_j\|\psi_j\|^2_{L^2(0,T,\mathbb{C})}, \tag{5}$$

where the constants $\gamma, \mu \geqslant 0$ are regularization parameters, which allow to vary the relative importance of the objectives represented by the various terms. We suppose that $\gamma + \mu > 0$. The goal of the first term of the cost functional is to track the state ψ close to a given terminal state at $t = T$. The second and third terms are for the regularization of the problem so that existence of at least one optimal control (see (**P**) below and proof of Theorem 3) is ensured. In the last term of (5), which penalizes the occupation of certain states ψ_j, the set $I \subset \{1, \ldots, n\}$ denotes a subset of possible state indices and $\alpha_j > 0$, $j \in I$, are weighting factors. Depending on the parameter μ, we define the Hilbert space X by

$$X = \begin{cases} H^1(0, T; \mathbb{C}^n) \times L^2(0, T; \mathbb{C}) & \text{if } \mu = 0, \\ H^1(0, T; \mathbb{C}^n) \times H_0^1(0, T; \mathbb{C}) & \text{if } \mu > 0 \end{cases}$$

and supply X with the natural product topology. Then, the optimal control problem can be written as the following abstract minimization problem

$$\min J(x) \quad \text{subject to } x = (\psi, \varepsilon) \in X \quad \text{and} \quad e(x) = 0 \text{ in } Y. \tag{P}$$

Remark 2. (1) As required in the quantum-mechanical framework, we choose $|\psi_0|_{\mathbb{C}^n} = 1$. In the case that H_0 is non-hermitian and dissipative, a target state with the normalization $|\psi_d|_{\mathbb{C}^n} = 1$ will not be attainable. This is typical of dissipative systems and motivates penalization of occupation of states representing environment losses.

(2) Because of H^1 regularization ($\mu > 0$) we have a natural setting to impose zero (boundary) conditions on the control field outside of the control window $[0, T]$. This requirement is in agreement with the need of designing control pulses with compact support.

(3) Recall the Poincaré inequality

$$\|\varepsilon\|_{L^2(0,T;\mathbb{C})} \leqslant K_2\|\dot{\varepsilon}\|_{L^2(0,T;\mathbb{C})} \quad \text{for all } \varepsilon \in H_0^1(0, T; \mathbb{C}). \tag{6}$$

Therefore, if $\mu > 0$ the L^2-norm of ε can be bounded by the L^2-norm of its derivative. This implies that we can choose $\gamma = 0$ provided $\mu > 0$ holds.

The next theorem ensures that (**P**) has at least one global solution.

Theorem 3. *The optimal control problem* (**P**) *admits a solution* $(\psi, \varepsilon) \in X$.

Proof. We prove the claim by two steps.

Step 1: We need an a-priori estimate for the solution of (1). For given $\varepsilon \in L^2(0, T; \mathbb{C})$, Eq. (1) is a linear system of ordinary differential equations, which therefore has a unique solution $\psi \in H^1(0, T; \mathbb{C}^n)$. We now write (1) in integral form

$$\mathrm{i}\psi(t) = \mathrm{i}\psi_0 + \int_0^t (H_0 + H_1(\varepsilon(s)))\psi(s)\,\mathrm{d}s \quad \text{for } 0 \leqslant t \leqslant T. \tag{7}$$

Taking the Euclidean norm in $\mathbb{C}^n$ on both sides of (7) and using the triangle inequality on the right-hand side results in

$$|\psi(t)|_{\mathbb{C}^n} \leqslant |\psi_0|_{\mathbb{C}^n} + \int_0^t (\varrho(H_0) + \varrho(H_1(\varepsilon(s))))|\psi(s)|_{\mathbb{C}^n}\,\mathrm{d}s.$$

Now apply Gronwall's inequality [13] to obtain

$$|\psi(t)|_{\mathbb{C}^n} \leqslant |\psi_0|_{\mathbb{C}^n} \exp\left(\int_0^t (\varrho(H_0) + \varrho(H_1(\varepsilon(s))))\,\mathrm{d}s\right) \quad \text{for } 0 \leqslant t \leqslant T.$$

Using (3), squaring and integrating the above inequality over $[0, T]$ results in

$$\|\psi\|_{L^2(0,T;\mathbb{C}^n)} \leqslant \sqrt{T}\,|\psi_0|_{\mathbb{C}^n} \exp(c_1 + c_2\|\varepsilon\|_{L^2(0,T;\mathbb{C})}) \quad \text{for } 0 \leqslant t \leqslant T, \tag{8}$$

with $c_1 = T\varrho(H_0)$ and $c_2 = \sqrt{T}K_0$. Furthermore, using the state equation (1) yields that for every $\varepsilon \in L^2(0, T; \mathbb{C})$ with $\|\varepsilon\|_{L^2(0,T;\mathbb{C})} \leqslant c_3$ the corresponding states $\psi = \psi(\varepsilon)$ are bounded in $H^1(0, T; \mathbb{C}^n)$, i.e.,

$$\|\psi\|_{H^1(0,T;\mathbb{C}^n)} \leqslant c_4 \quad \text{for some } c_4 > 0.$$

Step 2: Let $\{\varepsilon_k\}_{k\geqslant 1}$ be a minimizing sequence for J, i.e.,

$$\lim_{k\to\infty} J(\psi_k, \varepsilon_k) = \inf\{J(x)|x = (\psi, \varepsilon) \in X \text{ and } e(x) = 0 \text{ in } Y\},$$

where we denote by $\psi_k = \psi(\varepsilon_k)$ the unique solution to (1) for $\varepsilon = \varepsilon_k$. Let $\gamma > 0$ and $\mu = 0$ hold. Hence,

$$J(\psi(\varepsilon), \varepsilon) \to \infty \quad \text{as } \|\varepsilon\|_{L^2(0,T;\mathbb{C})} \to \infty$$

so that the sequence $\{\varepsilon_k\}_{k\geqslant 1}$ is bounded in $L^2(0, T; \mathbb{C})$. Since the unit ball in a Hilbert space is weakly compact, there exists a weakly to an $\varepsilon \in L^2(0, T; \mathbb{C})$ convergent subsequence, which we again denote by $\{\varepsilon_k\}_{k\geqslant 1}$. Step 1 above ensures that the corresponding sequence $\{\psi_k\}_{k\geqslant 1}$ is bounded in $H^1(0, T; \mathbb{C}^n)$; thus, again by choosing a proper subsequence

$$\psi_k \rightharpoonup \psi \quad \text{in } H^1(0, T; \mathbb{C}^n),$$

it follows from the Sobolev embedding theorem [1] that

$$\psi_k \to \psi \quad \text{in } L^2(0, T; \mathbb{C}^n) \text{ and in } C([0, T]; \mathbb{C}^n).$$

We can now show that (ψ, ε) is a solution of the optimal control problem. Since ψ_k solves (1) for ε_k we have

$$\mathrm{i}\psi_k(t) = \mathrm{i}\psi_0 + \int_0^t (H_0 + H_1(\varepsilon_k(s)))\psi_k(s)\,\mathrm{d}s \quad \text{for } 0 \leqslant t \leqslant T. \tag{9}$$

Next, we consider the limit in $k \to \infty$ of (9). The weak convergence of ε_k to ε in $L^2(0, T; \mathbb{C})$ implies weak convergence also for the complex conjugates, i.e., $\varepsilon_k^* \rightharpoonup \varepsilon^*$ in $L^2(0, T; \mathbb{C})$. Strong convergence of ψ_k to ψ in $L^2(0, T; \mathbb{C}^n)$ allows to go to the limit as $k \to \infty$ on the right-hand side of (9). Thus, we find

$$\mathrm{i}\psi(t) = \mathrm{i}\psi_0 + \int_0^t (H_0 + H_1(\varepsilon(s)))\psi(s)\,\mathrm{d}s,$$

which shows that $\psi = \psi(\varepsilon)$, or equivalently that $e(\psi, \varepsilon) = 0$ in Y. We finally obtain

$$\begin{aligned} J(\psi, \varepsilon) &= \frac{1}{2}|\psi(T) - \psi_d|^2_{\mathbb{C}^n} + \frac{\gamma}{2}\|\varepsilon\|^2_{L^2(0,T;\mathbb{C})} + \frac{1}{2}\sum_{j\in I}\alpha_j\|\psi_j\|^2_{L^2(0,T;\mathbb{C}^n)} \\ &\leqslant \frac{1}{2}\lim_{k\to\infty}|\psi_k(T) - \psi_d|^2_{\mathbb{C}^n} + \frac{1}{2}\liminf_{k\to\infty}(\gamma\|\varepsilon_k\|^2_{L^2(0,T;\mathbb{C})}) \\ &\quad + \frac{1}{2}\sum_{j\in I}\alpha_j\lim_{k\to\infty}\|\psi_{k,j}\|^2_{L^2(0,T;\mathbb{C}^n)} \\ &= \inf\{J(x)|x = (\psi, \varepsilon) \in X \text{ and } e(x) = 0 \text{ in } Y\}, \end{aligned}$$

where we used the lower-semicontinuity of the L^2-norm. Thus, we have proved that (ψ, ε) is a solution to (**P**) for $\gamma > 0$ and $\mu = 0$. If $\mu > 0$ and $\gamma = 0$ hold the sequence $\{\varepsilon_k\}_{k \geqslant 1}$ is bounded in $L^2(0, T; \mathbb{C})$ by the Poincaré inequality (6). Thus, the proof follows by analogous arguments. □

2.2. First-order necessary optimality conditions

To solve the above optimal control problem, we use the method of Lagrange multipliers (see, e.g., [21]) to turn the constrained minimization problem (**P**) into an unconstrained optimization problem, and we focus on the necessary optimality conditions of first order.

We define the Lagrangian function $L : X \times Y \to \mathbb{R}$ by

$$\begin{aligned} L(\psi, \varepsilon, p, q) &= J(\psi, \varepsilon) + \Re e(\langle e(\psi, \varepsilon), (p, q)\rangle_Y) \\ &= J(\psi, \varepsilon) + \Re e(\langle \mathrm{i}\dot{\psi} - H(\varepsilon(\cdot))\psi, p\rangle_{L^2(0,T;\mathbb{C}^n)} + (\psi(0) - \psi_0) \cdot q^*), \end{aligned}$$

for $(\psi, \varepsilon) \in X$ and $(p, q) \in Y$. To derive first-order necessary optimality conditions we need the following constraint qualification.

Proposition 4. *The operator $e : X \to Y$ is Fréchet-differentiable and its linearization $\nabla e(x_\circ)$ is surjective for every $x_\circ = (\psi_\circ, \varepsilon_\circ) \in X$.*

Proof. We first prove the claim for the case $\mu > 0$. Let $x_\circ = (\psi_\circ, \varepsilon_\circ) \in X = H^1(0, T; \mathbb{C}^n) \times H^1(0, T; \mathbb{C})$ be arbitrary. Recall that $H_0^1(0, T; \mathbb{C})$ is continuously embedded into $C([0, T]; \mathbb{C})$ so that there exists an embedding constant $c_1 > 0$ satisfying

$$\|\varepsilon\|_{C([0,T];\mathbb{C})} \leqslant c_1 \|\varepsilon\|_{H^1(0,T;\mathbb{C})} \quad \text{for all } \varepsilon \in H_0^1(0, T; \mathbb{C}). \tag{10}$$

Recall that the Hamiltonian matrix H is of the form (2). We compute the directional derivative of the operator e. For any direction $x = (\psi, \varepsilon) \in X$, $\varepsilon = \varepsilon_{\Re e} + \mathrm{i}\varepsilon_{\Im m}$ with $\varepsilon_{\Re e}, \varepsilon_{\Im m} \in H_0^1(0, T; \mathbb{R})$, we infer from (2) that

$$\nabla e(x_\circ)x = \lim_{h \searrow 0} \frac{e(x_\circ + hx) - e(x_\circ)}{h} = \begin{pmatrix} \mathrm{i}\dot{\psi} - H(\varepsilon_\circ(\cdot))\psi - H_1(\varepsilon(\cdot))\psi_\circ \\ \psi(0) \end{pmatrix}.$$

Next we prove that the directional derivative is already the Fréchet-derivative. Note that

$$e(x_\circ + x) - e(x_\circ) - \nabla e(x_\circ)x = \begin{pmatrix} H_1(\varepsilon(\cdot))\psi \\ 0 \end{pmatrix}.$$

Thus, we derive from (10), $|\varepsilon_{\Re e}(t)| \leqslant |\varepsilon(t)|$ and $|\varepsilon_{\Im m}(t)| \leqslant |\varepsilon(t)|$:

$$\begin{aligned} &\|e(x_\circ + x) - e(x_\circ) - \nabla e(x_\circ)x\|_Y^2 \\ &\quad \leqslant K_0^2 \int_0^T |\varepsilon(t)|^2 |\psi(t)|_{\mathbb{C}^n}^2 \, \mathrm{d}t \leqslant K_0^2 \, \|\varepsilon\|_{C([0,T];\mathbb{C})}^2 \|\psi\|_{L^2(0,T;\mathbb{C}^n)}^2 \\ &\quad \leqslant c_2(\|\psi\|_{H^1(0,T;\mathbb{C}^n)}^2 + \|\varepsilon\|_{H^1(0,T;\mathbb{C})}^2)^2 = c_2\|x\|_X^4, \end{aligned}$$

where $c_2 = c_1^2 K_0^2/2$. Consequently,

$$0 \leqslant \lim_{\|x\|_X \searrow 0} \frac{\|e(x_\circ + x) - e(x_\circ) - \nabla e(x_\circ)x\|_Y}{\|x\|_X} \leqslant c_2 \lim_{\|x\|_X \searrow 0} \|x\|_X = 0$$

so that the directional derivative $\nabla e(x_\circ)$ is already the Fréchet-derivative.

Next we prove that the linear operator $\nabla e(x_\circ)$ is surjective for every $x_\circ = (\psi_\circ, \varepsilon_\circ) \in X$. Recall that $\psi_\circ \in C([0, T]; \mathbb{C}^n)$. Let $(f, f_0) \in Y$ be arbitrary. Then $\nabla e(x_\circ)(\psi, \varepsilon) = (f, f_0)$ is equivalent to

$$\mathrm{i}\dot{\psi} = H(\varepsilon_\circ(\cdot))\psi + H_1(\varepsilon(\cdot))\psi_\circ + f \quad \text{in } (0, T] \quad \text{and} \quad \psi(0) = f_0. \tag{11}$$

Applying (3) and (4), the right-hand side $g = H_1(\varepsilon(\cdot))\psi_\circ + f$ belongs to $L^2(0, T; \mathbb{C}^n)$ for every $\varepsilon \in L^2(0, T; \mathbb{C})$. Thus, (11) is a linear system of ordinary differential equations that admits a unique solution $\psi \in H^1(0, T; \mathbb{C}^n)$ for every $\varepsilon \in L^2(0, T; \mathbb{C})$. In particular, the operator $\nabla e(x_\circ)$ is surjective. □

Remark 5. It follows from the proof of Proposition 4 that the linear operator $e_\psi(x_\circ) : H^1(0, T; \mathbb{C}^n) \to Y$ is bijective, where $e_\psi(x_\circ)$ denotes the partial Fréchet-derivative at $x_\circ$ with respect to $\psi \in H^1(0, T; \mathbb{C}^n)$.

Notice that the quadratic cost functional $J : X \to [0, \infty)$ is twice continuously Fréchet-differentiable. Using Proposition 4, first-order necessary conditions for a minimum are obtained by equating to zero the Fréchet derivatives of L with respect to $(\psi, \varepsilon, p, q)$, where L_ψ, L_ε, L_p, and L_q denote the partial derivatives of L with respect to ψ, ε, p, and q, respectively. We have $\nabla L(\psi, \varepsilon, p, q) = (L_\psi, L_\varepsilon, L_p, L_q)$.

Theorem 6. *Suppose that $x_\circ = (\psi_\circ, \varepsilon_\circ) \in X$ is a local solution to* (**P**). *Then there exist* (*unique*) *Lagrange multipliers $p_\circ \in H^1(0, T; \mathbb{C}^n)$ and $q_\circ \in \mathbb{C}^n$ satisfying*

$$\mathrm{i}\dot{\psi}_\circ = H(\varepsilon_\circ(\cdot))\psi_\circ \quad \text{in } (0, T], \tag{12a}$$

$$\psi_\circ(0) = \psi_0, \tag{12b}$$

$$\mathrm{i}(\dot{p}_\circ)_j = (H(\varepsilon_\circ(\cdot))^* p_\circ)_j - \alpha_j(\psi_\circ)_j \quad \text{in } (0, T], \quad j \in I, \tag{12c}$$

$$\mathrm{i}(\dot{p}_\circ)_j = (H(\varepsilon_\circ(\cdot))^* p_\circ)_j \quad \text{in } (0, T], \quad j \notin I, \tag{12d}$$

$$\mathrm{i}p_\circ(T) = \psi_\circ(T) - \psi_d, \tag{12e}$$

$$q_\circ = \mathrm{i}p_\circ(0), \tag{12f}$$

$$-\mu\ddot{\varepsilon}_\circ + \gamma\varepsilon_\circ = \Re e(H_{1\Re e}\psi_\circ \cdot p_\circ^*) + \mathrm{i}\Re e(H_{1\Im m}\psi_\circ \cdot p_\circ^*) \quad \text{in } (0, T], \tag{12g}$$

$$\varepsilon_\circ(T) = \varepsilon_\circ(0) = 0. \tag{12h}$$

in case of $\mu > 0$. Moreover, $\varepsilon_\circ \in C^2([0, T]; \mathbb{C}) \cap C([0, T]; \mathbb{C})$, i.e., $\varepsilon_\circ$ is a classical solution. If $\mu = 0$ holds, (12g) *and* (12h) *have to be replaced by*

$$\gamma\varepsilon_\circ = \Re e(H_{1\Re e}\psi_\circ \cdot p_\circ^*) + \mathrm{i}\Re e(H_{1\Im m}\psi_\circ \cdot p_\circ^*) \quad \text{in } (0, T). \tag{12g$'$}$$

Proof. It follows from Proposition 4 that there exist (unique) Lagrange multipliers $(p_\circ, q_\circ) \in Y$ satisfying

$$\nabla L(x_\circ, p_\circ, q_\circ) = 0 \quad \text{in } X \times Y. \tag{13}$$

The condition $L_p(x_\circ, p_\circ, q_\circ) = 0$ in $L^2(0, T; \mathbb{C}^n)$ implies (12a), whereas the equation $L_q(x_\circ, p_\circ, q_\circ) = 0$ in $\mathbb{C}^n$ yields (12b). Next we turn to the partial derivative of the Lagrangian with respect to ψ. Let $\phi \in H^1(0, T; \mathbb{C}^n)$ be arbitrary. Then we find

$$L_\psi(x_\circ, p_\circ, q_\circ)\phi = \Re e\left((\psi_\circ(T) - \psi_d) \cdot \phi(T)^* + \sum_{j\in I} \alpha_j \langle(\psi_\circ)_j, \phi_j\rangle_{L^2(0,T;\mathbb{C}^n)}\right)$$
$$+ \Re e(\langle \mathrm{i}\dot{\phi} - H(\varepsilon_\circ(\cdot))\phi, p_\circ\rangle_{L^2(0,T;\mathbb{C}^n)} + \phi(0) \cdot q_\circ^*). \tag{14}$$

Using integration by parts and (13) we obtain

$$\Re e(\langle \mathrm{i}\dot{\phi} - H(\varepsilon_\circ(\cdot))\phi, p_\circ\rangle_{L^2(0,T;\mathbb{C}^n)}) = \Re e(\langle \mathrm{i}\dot{p}_\circ - H(\varepsilon_\circ(\cdot))^* p_\circ, \phi\rangle_{H^1(0,T;\mathbb{C}^n)',H^1(0,T;\mathbb{C}^n)})$$
$$+ \Re e(\mathrm{i}p_\circ(T) \cdot \phi(T)^* - \mathrm{i}p_\circ(0) \cdot \phi(0)^*),$$

where $\langle\cdot, \cdot\rangle_{H^1(0,T;\mathbb{C}^n)',H^1(0,T;\mathbb{C}^n)}$ stands for the dual pairing of $H^1(0, T; \mathbb{C}^n)$ and its dual space $H^1(0, T; \mathbb{C}^n)'$. We infer from (14) that (12c)–(12d) are satisfied in $H^1(0, T; \mathbb{C}^n)'$. From (3) and the fact that $\psi_\circ, p_\circ \in L^2(0, T; \mathbb{C}^n)$, it

follows that the right-hand sides in (12c)–(12d) even belong to $L^2(0,T;\mathbb{C}^n)$. Hence, $p_\circ$ is an element in $H^1(0,T;\mathbb{C}^n)$ and (12c)–(12d) hold in $L^2(0,T;\mathbb{C}^n)$. Inserting (12c)–(12d) in (14) implies (12e)–(12f). In case of $\mu>0$, we conclude from $L_{\varepsilon_{\Re e}}(x_\circ,p_\circ,q_\circ)\varepsilon_{\Re e}=0$ for all $\varepsilon=\varepsilon_{\Re e}+\mathrm{i}\varepsilon_{\Im m}\in H^1(0,T;\mathbb{C})$ that

$$\gamma\langle\varepsilon_{\circ\Re e},\varepsilon_{\Re e}\rangle_{L^2(0,T;\mathbb{R})}+\mu\langle\dot\varepsilon_{\circ\Re e},\dot\varepsilon_{\Re e}\rangle_{L^2(0,T;\mathbb{R})}-\Re e\left(\int_0^T\varepsilon_{\Re e}H_{1\Re e}\psi_\circ\cdot p_\circ^*\right)=0$$

with $\varepsilon_\circ=\varepsilon_{\circ\Re e}+\mathrm{i}\varepsilon_{\circ\Im m}\in H_0^1(0,T;\mathbb{C})$ and $\varepsilon_{\circ\Re e},\varepsilon_{\circ\Im m}\in H_0^1(0,T;\mathbb{R})$. Applying integration by parts we derive

$$\langle\dot\varepsilon_{\circ\Re e},\dot\varepsilon_{\Re e}\rangle_{L^2(0,T;\mathbb{R})}=-\langle\ddot\varepsilon_{\circ\Re e},\varepsilon_{\Re e}\rangle_{H^1(0,T;\mathbb{R})',H^1(0,T;\mathbb{R})}.$$

Hence,

$$\gamma\varepsilon_{\circ\Re e}-\mu\ddot\varepsilon_{\circ\Re e}-\Re e(H_{1\Re e}\psi_\circ\cdot p_\circ^*)=0. \tag{15}$$

Analogously, $L_{\varepsilon_{\Im m}}(x_\circ,p_\circ,q_\circ)\varepsilon_{\Im m}=0$ for all $\varepsilon=\varepsilon_{\Re e}+\mathrm{i}\varepsilon_{\Im m}\in H^1(0,T;\mathbb{C})$ implies that

$$\gamma\varepsilon_{\circ\Im m}-\mu\ddot\varepsilon_{\circ\Im m}-\Re e(H_{1\Im m}\psi_\circ\cdot p_\circ^*)=0. \tag{16}$$

Multiplying (16) by the complex unit i and adding (15) we obtain (12g), whereas (12h) follows directly from $\varepsilon_\circ\in H_0^1(0,T;\mathbb{C})$. Since $\psi_\circ,p_\circ\in C([0,T];\mathbb{C}^n)$, the right-hand side in (12g) belongs to $C([0,T];\mathbb{C})$. Thus, $\varepsilon_\circ$ is a classical solution. In case of $\mu=0$ we have to replace (15)–(16) by

$$\gamma\varepsilon_{\circ\Im m}-\mu\ddot\gamma\varepsilon_{\circ\Re e}-\Re e(H_{1\Re e}\psi_\circ\cdot p_\circ^*)=0\quad\text{and}\quad\gamma\varepsilon_{\circ\Im m}-\Re e(H_{1\Im m}\psi_\circ\cdot p_\circ^*)=0, \tag{17}$$

respectively, so that we derive (12g′). □

Having computed a solution to the first-order conditions (12) it remains to verify that this solution corresponds to a local minimum. For this reason, second-order optimality conditions are considered in the next section.

Remark 7. Recall that (1) is uniquely solvable for every $\varepsilon\in L^2(0,T;\mathbb{C})$. Thus, it is meaningful to introduce the so-called reduced cost functional $\hat J:H^1(0,T;\mathbb{C})\to\mathbb{R}$ given by

$$\hat J(\varepsilon)=J(\psi(\varepsilon),\varepsilon), \tag{18}$$

where $\psi(\varepsilon)$ denotes the unique solution to (1) for ε. It follows from the proof of Theorem 6 that the gradient of $\hat J$ with respect to the real and imaginary parts of $\varepsilon=\varepsilon_{\Re e}+\mathrm{i}\varepsilon_{\Im m}$ is given by

$$\nabla\hat J(\varepsilon)=\begin{pmatrix}\gamma\varepsilon_{\Re e}-\mu\ddot\varepsilon_{\circ\Re e}-\Re e(H_{1\Re e}\psi\cdot p^*)\\ \gamma\varepsilon_{\Im m}-\mu\ddot\varepsilon_{\Im m}-\Re e(H_{1\Im m}\psi\cdot p^*)\end{pmatrix}, \tag{19}$$

where ψ and p solve (12a)–(12b) and (12c)–(12e), respectively, with ε instead of $\varepsilon_\circ$; compare (15)–(16). If $\mu=0$ holds, then $\hat J$ is defined on $L^2(0,T;\mathbb{C})$ and we have the representation

$$\nabla\hat J(\varepsilon)=\begin{pmatrix}\gamma\varepsilon_{\Re e}-\Re e(H_{1\Re e}\psi\cdot p^*)\\ \gamma\varepsilon_{\Im m}-\Re e(H_{1\Im m}\psi\cdot p^*)\end{pmatrix}; \tag{20}$$

see (17).

2.3. Second-order sufficient optimality conditions

Problem (**P**) is a non-convex programming problem so that different local minima will probably occur. Numerical methods will deliver a local minimum close to the given starting point. Suppose that $x_\circ=(\psi_\circ,\varepsilon_\circ)\in X$ is an optimal solution to (**P**) and that $(p_\circ,q_\circ)\in Y$ are the associated Lagrange multiplier satisfying (12c)–(12f). The second

Fréchet-derivative of the Lagrangian—denoted by L_{xx}—at $(x_\circ, p_\circ, q_\circ)$ with respect to x is given by

$$\begin{aligned} &L_{xx}(x_\circ, p_\circ, q_\circ)(x, x)\\ &\quad = |\psi(T)|^2_{\mathbb{C}^n} + \gamma\|\varepsilon\|^2_{L^2(0,T;\mathbb{C})} + \mu\|\dot{\varepsilon}\|^2_{L^2(0,T;\mathbb{C})} + \sum_{j\in I}\alpha_j\|\psi_j\|^2_{L^2(0,T;\mathbb{C}^n)}\\ &\qquad - 2\Re e\left(\int_0^T (H_1(\varepsilon(t))\psi(t))\cdot p_\circ(t)^*\,\mathrm{d}t\right), \end{aligned}$$

for every direction $x = (\psi, \varepsilon) \in X$.

In case of $\mu > 0$ the second-order sufficient optimality conditions for (**P**) are as follows (see, e.g., [24]): There exists a constant $\kappa > 0$ such that

$$L_{xx}(x_\circ, p_\circ, q_\circ)((\psi, \varepsilon), (\psi, \varepsilon)) \geqslant \kappa(\|\psi\|^2_{H^1(0,T;\mathbb{C}^n)} + \|\varepsilon\|^2_{H^1(0,T;\mathbb{C})}), \tag{21}$$

for all $(\psi, \varepsilon) \in X$ satisfying the linearized state equation

$$\mathrm{i}\dot{\psi} = H(\varepsilon_\circ(\cdot))\psi + H_1(\varepsilon(\cdot))\psi_\circ \quad \text{in } (0, T] \quad \text{and} \quad \psi(0) = 0. \tag{22}$$

It follows from Proposition 4 and Remark 5 that for every $\varepsilon \in L^2(0, T; \mathbb{C})$ the problem (22) possesses a unique solution.

In case of $\mu = 0$ we have to replace (21) by

$$L_{xx}(x_\circ, p_\circ, q_\circ)((\psi, \varepsilon), (\psi, \varepsilon)) \geqslant \kappa(\|\psi\|^2_{H^1(0,T;\mathbb{C}^n)} + \|\varepsilon\|^2_{L^2(0,T;\mathbb{C})}). \tag{23}$$

To prove the second-order sufficient optimality condition we need the following two lemmas.

Lemma 8. *Let $(\tilde{\psi}_\circ, \tilde{\varepsilon}_\circ) \in L^2(0, T; \mathbb{C}^n) \times L^2(0, T; \mathbb{C})$ represent the linearization point and let $(\psi, \varepsilon) \in H^1(0, T; \mathbb{C}^n) \times L^2(0, T; \mathbb{C})$ satisfy* (22). *Then there exists a constant $C > 0$ depending on $\tilde{\psi}_\circ$ and $\tilde{\varepsilon}_\circ$ so that*

$$\|\psi\|_{L^\infty(0,T;\mathbb{C}^n)} + \|\psi\|_{L^2(0,T;\mathbb{C}^n)} \leqslant C\|\varepsilon\|_{L^2(0,T;\mathbb{C})} \quad \textit{for almost all } t \in (0, T]. \tag{24}$$

Proof. To prove the assertion we apply Gronwall's lemma. From (22) we have

$$\mathrm{i}\psi(t) = \mathrm{i}\psi(0) + \int_0^t \mathrm{i}\dot{\psi}(s)\,\mathrm{d}s = \int_0^t (H_0 + H_1(\tilde{\varepsilon}_\circ(s)))\psi(s) + H_1(\varepsilon(s))\tilde{\psi}_\circ(s)\,\mathrm{d}s$$

for almost all $t \in (0, T]$. Using (3) we find

$$\begin{aligned} |\psi(t)|_{\mathbb{C}^n} &\leqslant \int_0^t (\varrho(H_0) + K_0|\tilde{\varepsilon}_\circ(s)|)|\psi(s)|_{\mathbb{C}^n} + K_0|\varepsilon(s)||\tilde{\psi}_\circ(s)|_{\mathbb{C}^n}\,\mathrm{d}s\\ &\leqslant C_1(1 + \|\tilde{\varepsilon}_\circ\|^2_{L^2(0,T;\mathbb{C})})^{1/2}\left(\int_0^t |\psi(s)|^2_{\mathbb{C}^n}\,\mathrm{d}s\right)^{1/2} + C_2\|\varepsilon\|_{L^2(0,T;\mathbb{C})}, \end{aligned}$$

where $C_1 = \max\{2T\varrho(H_0)^2, 2K_0^2\}$ and $C_2 = K_0\|\tilde{\psi}_\circ\|_{L^2(0,T;\mathbb{C}^n)}$. Consequently,

$$|\psi(t)|^2_{\mathbb{C}^n} \leqslant C_3\left((1 + \|\tilde{\varepsilon}_\circ\|^2_{L^2(0,T;\mathbb{C})})\int_0^t |\psi(s)|^2_{\mathbb{C}^n}\,\mathrm{d}s + \|\varepsilon\|^2_{L^2(0,T;\mathbb{C})}\right),$$

with $C_3 = \max\{2C_1^2, 2C_2^2\}$. Applying Gronwall's lemma and using $\psi(0) = 0$ we obtain

$$|\psi(t)|^2_{\mathbb{C}^n} \leqslant C_3\|\tilde{\psi}_\circ\|^2_{L^2(0,T;\mathbb{C}^n)}\, e^{C_3 t(1+\|\tilde{\varepsilon}_\circ\|^2_{L^2(0,T;\mathbb{C})})}\|\varepsilon\|^2_{L^2(0,T;\mathbb{C})} \tag{25}$$

for almost all $t \in (0, T]$ and

$$\|\psi\|^2_{L^2(0,T;\mathbb{C}^n)} \leqslant TC_3\|\tilde{\psi}_\circ\|^2_{L^2(0,T;\mathbb{C}^n)}\, e^{C_3 T(1+\|\tilde{\varepsilon}_\circ\|^2_{L^2(0,T;\mathbb{C})})}\|\varepsilon\|^2_{L^2(0,T;\mathbb{C})}. \tag{26}$$

From (25) and (26) we infer (24). □

Remark 9. If in addition to the hypotheses of Lemma 8, we have $\tilde{\psi}_\circ \in L^\infty(0,T;\mathbb{C}^n)$ then we can give an estimate for ψ in the $H^1(0,T;\mathbb{C}^n)$-norm. Applying (3), it follows that

$$\begin{aligned}
\|\dot{\psi}\|^2_{L^2(0,T;\mathbb{C}^n)} &= \int_0^T |(H_0 + H_1(\tilde{\varepsilon}_\circ(t))\psi(t) + H_1(\varepsilon(s))\tilde{\psi}_\circ(t)|^2 \,\mathrm{d}t \\
&\leqslant \int_0^T |(\varrho(H_0) + K_0|\tilde{\varepsilon}_\circ(t)|)|\psi(t)|_{\mathbb{C}^n} + K_0|\varepsilon(t)||\tilde{\psi}_\circ(t)|_{\mathbb{C}^n}|^2 \,\mathrm{d}t \\
&\leqslant \int_0^T 2(\varrho(H_0) + K_0\,|\tilde{\varepsilon}_\circ(t)|)^2|\psi(t)|^2_{\mathbb{C}^n} + 2K_0^2|\varepsilon(t)|^2|\tilde{\psi}_\circ(t)|^2_{\mathbb{C}^n} \,\mathrm{d}t \\
&\leqslant 4\varrho(H_0)^2\,\|\psi\|^2_{L^2(0,T;\mathbb{C}^n)} + 4K_0^2\|\psi\|^2_{L^\infty(0,T;\mathbb{C}^n)}\|\tilde{\varepsilon}_\circ\|^2_{L^2(0,T;\mathbb{C})} \\
&\quad + 2K_0^2\|\tilde{\psi}_\circ\|^2_{L^\infty(0,T;\mathbb{C}^n)}\|\varepsilon\|^2_{L^2(0,T;\mathbb{C})} \\
&\leqslant C_4(\|\psi\|^2_{L^\infty(0,T;\mathbb{C}^n)} + \|\psi\|^2_{L^2(0,T;\mathbb{C}^n)} + \|\varepsilon\|^2_{L^2(0,T;\mathbb{C})}),
\end{aligned}$$

with

$$C_4 = \max\{4\varrho(H_0)^2,\ 4K_0^2\,\|\tilde{\varepsilon}_\circ\|^2_{L^2(0,T;\mathbb{C})},\ 2K_0^2\|\tilde{\psi}_\circ\|^2_{L^\infty(0,T;\mathbb{C}^n)}\}.$$

Thus, it follows from Lemma 8 that

$$\|\psi\|_{H^1(0,T;\mathbb{C}^n)} \leqslant C_5\|\varepsilon\|_{L^2(0,T;\mathbb{C})} \tag{27}$$

for a constant $C_5 > 0$ depending on $\|\psi_\circ\|_{L^\infty(0,T;\mathbb{C}^n)}$ and $\|\varepsilon_\circ\|_{L^2(0,T;\mathbb{C})}$.

Lemma 10. *Suppose that* $(\psi_\circ, \varepsilon_\circ)$ *is a local solution to* **(P)**. *Let* $\alpha = (\alpha_i) \in \mathbb{R}^n$ *be given by* $\alpha_i > 0$ *for* $i \in I$ *and* $\alpha_i = 0$ *otherwise. Then the solution* $p_\circ \in H^1(0,T;\mathbb{C}^n)$ *to* (12c)–(12e) *satisfies*

$$|p_\circ(t)|_{\mathbb{C}^n} \leqslant C(|\psi_\circ(T) - \psi_d|_{\mathbb{C}^n} + \|D_\alpha\psi_\circ\|_{L^1(0,T;\mathbb{C}^n)}) \quad \textit{for almost all } t \in [0,T],$$

where the constant $C > 0$ *depends on* $\varepsilon_\circ$ *and* D_α *denotes a diagonal matrix satisfying* $D_\alpha = \operatorname{diag}(\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^{n\times n}$.

Proof. From (12c)–(12d) we infer

$$(\mathrm{i}p_\circ(T) - \mathrm{i}p_\circ(t))_j = \left(\int_t^T (H_0^* + H_1(\varepsilon(s))^*)p_\circ(s) - \alpha_j(\psi_\circ(s))_j \,\mathrm{d}s\right)_j$$

for almost all $t \in [0,T)$ and $1 \leqslant j \leqslant n$. Using (3) and $i\ p_\circ(T) = \psi_\circ(T) - \psi_d$ it follows that

$$\begin{aligned}
|p_\circ(t)|_{\mathbb{C}^n} &\leqslant |\psi_\circ(T) - \psi_d|_{\mathbb{C}^n} + \int_t^T (\varrho(H_0) + K_0\,|\varepsilon(s)|)|p_\circ(s)|_{\mathbb{C}^n} + |D_\alpha\psi_\circ(s)|_{\mathbb{C}^n} \,\mathrm{d}s \\
&\leqslant |\psi_\circ(T) - \psi_d|_{\mathbb{C}^n} + \|D_\alpha\psi_\circ\|_{L^1(0,T;\mathbb{C}^n)} \\
&\quad + \tilde{C}_1(1 + \|\varepsilon_\circ\|^2_{L^2(0,T;\mathbb{C})})^{1/2}\left(\int_t^T |p_\circ(s)|^2_{\mathbb{C}^n} \,\mathrm{d}s\right)^{1/2},
\end{aligned}$$

where $\tilde{C}_1 = \max\{2T\,\varrho(H_0)^2, 2K_0^2\}$. Setting $\tilde{C}_2 = 4\max\{1, 2C_1^2\}$ we get

$$|p_\circ(t)|^2_{\mathbb{C}^n} \leqslant \tilde{C}_2\left(|\psi_\circ(T) - \psi_d|^2_{\mathbb{C}^n} + \|D_\alpha\psi_\circ\|^2_{L^1(0,T;\mathbb{C}^n)} + (1 + \|\varepsilon_\circ\|^2_{L^2(0,T;\mathbb{C})})\int_t^T |p_\circ(s)|^2_{\mathbb{C}^n} \,\mathrm{d}s\right).$$

Applying Gronwall's lemma it follows that

$$|p_\circ(t)|^2_{\mathbb{C}^n} \leqslant \tilde{C}_2 e^{\tilde{C}_2(T-t)(1+\|\varepsilon_\circ\|^2_{L^2(0,T;\mathbb{C})})}(|\psi_\circ(T)-\psi_d|^2_{\mathbb{C}^n} + \|D_\alpha\psi_\circ\|^2_{L^1(0,T;\mathbb{C}^n)})$$

for almost all $t \in [0, T)$, which gives the assertion. □

Now we turn to the second-order sufficient optimality conditions. Let $x_\circ = (\psi_\circ, \varepsilon_\circ) \in X$ be a local solution to (**P**) and $(p_\circ, q_\circ) \in X$ the associated pair of Lagrange multipliers. Suppose that $\mu > 0$ and $x = (\psi, \varepsilon) \in X$ satisfy (22). Due to Remark 9 and (6) there exists a constant $\tilde{C} > 0$ such that

$$\|\dot{\varepsilon}\|^2_{L^2(0,T;\mathbb{C})} \geqslant \frac{1}{\tilde{C}}\,\|\psi\|^2_{H^1(0,T;\mathbb{C}^n)}. \tag{28}$$

By Lemma 10 there exists a constant $\hat{C} > 0$ depending on $\varepsilon_\circ$ such that

$$|p_\circ(t)|_{\mathbb{C}^n} \leqslant \hat{C}\,\left(|\psi_\circ(T)-\psi_d|_{\mathbb{C}^n} + \|D_\alpha\psi_\circ\|_{L^1(0,T;\mathbb{C}^n)}\right) \quad \text{for all } t \in [0, T]. \tag{29}$$

From (3), (6), (28), and (29) we derive

$$\begin{aligned}
&L_{xx}(x_\circ, p_\circ, q_\circ)(x, x)\\
&\quad\geqslant \gamma\|\varepsilon\|^2_{L^2(0,T;\mathbb{C})} + \frac{\mu}{2}\|\dot{\varepsilon}\|^2_{L^2(0,T;\mathbb{C})} + \frac{\mu}{2\tilde{C}}\|\psi\|^2_{H^1(0,T;\mathbb{C}^n)}\\
&\qquad - 2K_0\int_0^T |\varepsilon(t)||\psi(t)|_{\mathbb{C}^n}|p_\circ(t)|_{\mathbb{C}^n}\,\mathrm{d}t\\
&\quad\geqslant \left(\gamma + \frac{\mu}{4K_2}\right)\|\varepsilon\|^2_{L^2(0,T;\mathbb{C})} + \frac{\mu}{4}\,\|\dot{\varepsilon}\|^2_{L^2(0,T;\mathbb{C})} + \frac{\mu}{2\tilde{C}}\,\|\psi\|^2_{H^1(0,T;\mathbb{C}^n)}\\
&\qquad - \hat{C}K_0(|\psi_\circ(T)-\psi_d|_{\mathbb{C}^n} + \|D_\alpha\psi_\circ\|_{L^1(0,T;\mathbb{C}^n)})(\|\varepsilon\|^2_{L^2(0,T;\mathbb{C})} + \|\psi\|^2_{H^1(0,T;\mathbb{C}^n)})\\
&\quad\geqslant \min\left\{\gamma + \frac{\mu}{4K_2} - \hat{C}K_0(|\psi_\circ(T)-\psi_d|_{\mathbb{C}^n} + \|D_\alpha\psi_\circ\|_{L^1(0,T;\mathbb{C}^n)}),\ \frac{\mu}{4}\right\}\|\varepsilon\|^2_{H^1(0,T;\mathbb{C})}\\
&\qquad + \left(\frac{\mu}{2\tilde{C}} - \hat{C}K_0(|\psi_\circ(T)-\psi_d|_{\mathbb{C}^n} + \|D_\alpha\psi_\circ\|_{L^1(0,T;\mathbb{C}^n)})\right)\|\psi\|^2_{H^1(0,T;\mathbb{C}^n)}.
\end{aligned}$$

If

$$|\psi_\circ(T)-\psi_d|_{\mathbb{C}^n} + \|D_\alpha\psi_\circ\|_{L^1(0,T;\mathbb{C}^n)} < \min\left\{\frac{4\gamma K_2+\mu}{4\hat{C}K_0K_2},\ \frac{\mu}{2\tilde{C}\hat{C}K_0}\right\} \tag{30}$$

then

$$\begin{aligned}
\kappa = \min\Bigg\{&\gamma + \frac{\mu}{4K_2} - \hat{C}K_0(|\psi_\circ(T)-\psi_d|_{\mathbb{C}^n} + \|D_\alpha\psi_\circ\|_{L^1(0,T;\mathbb{C}^n)}),\ \frac{\mu}{4},\\
&\frac{\mu}{2\tilde{C}} - \hat{C}K_0(|\psi_\circ(T)-\psi_d|_{\mathbb{C}^n} + \|D_\alpha\psi_\circ\|_{L^1(0,T;\mathbb{C}^n)})\Bigg\}
\end{aligned}$$

is positive and (21) holds. In case of $\mu = 0$ we replace (28) by (27). Then, the second-order sufficient condition can be shown analogously to the case $\mu > 0$. We summarize the results in the following theorem.

Theorem 11. *Suppose that $x_\circ = (\psi_\circ, \varepsilon_\circ) \in X$ is an optimal solution to* (**P**) *and that $(p_\circ, q_\circ) \in Y$ are the associated* (*unique*) *Lagrange multipliers satisfying* (12c)–(12f). *If* (30) *holds*, *then the second-order sufficient condition* (21) *is satisfied*.

Remark 12. The second-order sufficient optimality condition holds provided the terminal residuum $|\psi_\circ(T) - \psi_d|_{\mathbb{C}^n}$ as well as the term

$$\|D_\alpha \psi_\circ\|_{L^1(0,T;\mathbb{C}^n)} = \int_0^T \sum_{j\in I} \alpha_j |(\psi_\circ(t))_j| \, \mathrm{d}t$$

are sufficiently small. The term $\|D_\alpha \psi_\circ\|_{L^1(0,T;\mathbb{C}^n)}$ is a measure of the population lost occurring in unstable states modeled by ψ_j, $j \in I$. If the population lost turns out to be large, then the terminal state does not reach the desired state ψ_d. Thus, the residual $|\psi_\circ(T) - \psi_d|_{\mathbb{C}^n}$ is also large. On the other hand, if the population lost is small, we may also expect that the residual is small. In Section 5 we show that appropriate choices of optimization parameters result in small residual and reduced population lost, thus fulfilling (30).

3. Non-linear conjugate gradient approach

Our purpose is to give an appropriate formulation of the NCG method with a robust line search strategy and cascadic acceleration that results in a competitive scheme for quantum optimal control problems. Two features determine the success of the scheme proposed in this paper: (1) the use of a robust line search strategy together with an extension of a newly proposed formula [11] for determining conjugate search directions in the NCG method and (2) the use of cascadic acceleration. The latter is motivated by our computational experience and the results given in [4,36] regarding CG schemes with cascadic acceleration applied to the solution of elliptic problems. In [4,36] it is proved that cascadic CG schemes provide solvers with optimal computational complexity.

Gradient-type methods were the first to be used in the early day of quantum control computation [29], but appeared to be less competitive than monotonic schemes based on the Krotov method, see Section 4. However, some encouraging results were presented in [39] using an NCG scheme. The approach in [39] was a crude generalization of the NCG method with a special rule concerning the value of steplength [39, formula 2.32]). The resulting algorithm was not always robust and convergence slow-down could be observed.

In this section we illustrate our NCG approach that represents a considerable improvement with respect to the NCG scheme in [39]. Moreover, we investigate the convergence properties of our algorithm, thus giving justification of the use of the NCG scheme in the complex framework.

3.1. The non-linear conjugate gradient method

We start discussing the minimization by NCG methods of a differentiable function $f : \mathbb{R}^m \to \mathbb{R}$. We denote $g(x) = \nabla f(x)$, $x \in \mathbb{R}^m$.

NCG schemes represent extensions of linear conjugate gradient (CG) to non-quadratic problems; see, e.g., [14,37]. In the common variants, the basic idea is to avoid matrix operations and express the search directions recursively as

$$d_{k+1} = -g_{k+1} + \beta_k d_k, \tag{31}$$

for $k = 1, 2, \ldots$, with $d_1 = -g_1$. The iterates for a minimum point are given by

$$x_{k+1} = x_k + \tau_k d_k, \tag{32}$$

where $\tau_k > 0$ is a steplength. The parameter β_k is chosen so that (31)–(32) reduces to the linear CG scheme if f is a strictly convex quadratic function and τ_k is the exact one-dimensional minimizer of f along d_k. In this case the NCG scheme terminates in at most n steps in exact arithmetic. This case provides a lower bound to the computational complexity of NCG schemes.

We focus on the NCG scheme in [11] based on the formula

$$\beta_k = \beta_k^{DY} := \frac{\langle g_{k+1}, g_{k+1}\rangle_{\mathbb{R}^m}}{\langle d_k, y_k\rangle_{\mathbb{R}^m}},$$

where $y_k = g_{k+1} - g_k$. In [11], convergence of the proposed NCG scheme is established requiring that the steplength τ_k satisfies the standard Wolfe conditions given by

$$f(\varepsilon_k) - f(\varepsilon_k + \tau_k d_k) \geqslant -\delta\tau_k \langle g_k, d_k\rangle_{\mathbb{R}^m}$$

and

$$\langle g(\varepsilon_k + \tau_k d_k), d_k\rangle_{\mathbb{R}^m} > \sigma \langle g_k, d_k\rangle_{\mathbb{R}^m},$$

with $0 < \delta < \sigma < 1$. Replacing the second of these two conditions above with the following results in the strong Wolfe conditions:

$$|\langle g(\varepsilon_k + \tau_k d_k), d_k\rangle_{\mathbb{R}^m}| \leqslant -\sigma \langle g_k, d_k\rangle_{\mathbb{R}^m},$$

with $0 < \delta < \sigma < \frac{1}{2}$; see [14]. This condition means that the graph of f should not increase too fast (beyond the minimum). Note that convergence can only be proved provided d_k is a descent direction for any k, i.e., $\langle g_k, d_k\rangle < 0$ holds.

Next, we discuss the minimization of the real-valued differentiable function $\hat{J}(\varepsilon) = J(\psi(\varepsilon), \varepsilon)$ defined in (18); see Remark 7. Denote with $g(\varepsilon) = \nabla \hat{J}(\varepsilon)$. Note that $g(\varepsilon) \in H_0^1(0, T; \mathbb{C})$ for $\mu > 0$ and $g(\varepsilon) \in L^2(0, T; \mathbb{C})$ for $\mu = 0$. In the following we discuss the NCG scheme in a continuous setting on $L^2(0, T; \mathbb{C})$ with $\langle u, v\rangle = \langle u, v\rangle_{L^2(0,T;\mathbb{C})}$. In the discrete setting the following holds true with $\langle u, v\rangle = \delta t \sum_{\ell=0}^{N-1} u_\ell\, v_\ell^*$, $N\, \delta t = T$; see Section 4 for more details.

We define

$$\beta_k = \frac{\|g_{k+1}\|^2}{\Re e\, \langle d_k, y_k\rangle}. \tag{33}$$

We require that the steplength τ_k satisfies

$$\hat{J}(\varepsilon_k) - \hat{J}(\varepsilon_k + \tau_k d_k) \geqslant -\delta \tau_k \Re e\, \langle g_k, d_k\rangle, \tag{34}$$

$$\langle g(\varepsilon_k + \tau_k d_k), d_k\rangle > \sigma \Re e\, \langle g_k, d_k\rangle, \tag{35}$$

where the parameters are still chosen such that: $0 < \delta < \sigma < \frac{1}{2}$. The quality of line search in non-linear CG algorithms is crucial to preserve mutual conjugacy property of search directions and to ensure that each generated direction is one of descent. We use the Wolfe–Powell strategy to determine τ_k; see [26].

In the present framework, a sufficient descent condition is given by

$$\Re e\, \langle g_k, d_k\rangle \leqslant -c\, \|g_k\|^2,$$

for some $c > 0$ and for all $k \geqslant 1$ is guaranteed only replacing (35) with the stronger condition

$$|\langle g(\varepsilon_k + \tau_k d_k), d_k\rangle| \leqslant -\sigma \Re e\, \langle g_k, d_k\rangle$$

(and $c = 1/(1 + \sigma)$ results).

We consider the following NCG scheme:

Algorithm 13 (*NCG method*). *Step* 1: *Given* $k = 1$, ε_1, $d_1 = -g_1$, *if* $\|g_1\| < tol$ *then stop*.
Step 2: *Compute* $\tau_k > 0$ *satisfying* (34)–(35).
Step 3: *Let* $\varepsilon_{k+1} = \varepsilon_k + \tau_k d_k$.
Step 4: *Compute* $g_{k+1} = \nabla \hat{J}(\varepsilon_{k+1})$.
If $\|g_{k+1}\| < \mathrm{tol}_{\mathrm{abs}}$ *or* $\|g_{k+1}\| < tol_{rel}\|g_1\|$ *or* $k = k_{\max}$ *then stop*.
Step 5: *Compute* β_k *by* (33).
Step 6: *Let* $d_{k+1} = -g_{k+1} + \beta_k\, d_k$.
Step 7: *Set* $k = k + 1$, *goto Step* 2.

In the remaining of this section we prove convergence of the NCG scheme defined above that gives justification of the use of the NCG scheme in the complex framework.

We need the following assumption.

Assumption 14. (1) $\hat{J}$ is bounded from below and is continuously real differentiable in a neighborhood $\mathcal{N}$ of the level set $\mathcal{L} = \{\varepsilon \in E :\ \hat{J}(\varepsilon) \leqslant \hat{J}(\varepsilon_1)\}$.

(2) $\nabla \hat{J}$ is Lipschitz continuous in $\mathcal{N}$, i.e., there exists a constant $L > 0$ such that

$$\|\nabla \hat{J}(\varepsilon_1) - \nabla \hat{J}(\varepsilon_2)\| \leqslant L \|\varepsilon_1 - \varepsilon_2\| \quad \text{for all } \varepsilon_1, \varepsilon_2 \in \mathcal{N}.$$

Notice that in our setting Assumption 14-(1) is satisfied. Moreover, $\varepsilon \mapsto \hat{J}(\varepsilon)$ is twice continuously Fréchet-differentiable so that Assumption 14-(2) holds at least locally.

We have

$$\begin{aligned}\Re e\langle g_{k+1}, d_{k+1}\rangle &= \Re e\langle g_{k+1}, -g_{k+1} + \beta_k d_k\rangle = -\|g_{k+1}\|^2 + \beta_k \,\Re e\,\langle g_{k+1}, d_k\rangle \\ &= -\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2}{\Re e\,\langle d_k, y_k\rangle}\,\Re e\,\langle g_{k+1}, d_k\rangle \\ &= \frac{\|g_{k+1}\|^2}{\Re e\,\langle d_k, y_k\rangle}(-\Re e\,\langle d_k, y_k\rangle + \Re e\,\langle g_{k+1}, d_k\rangle) = \beta_k\,\Re e\,\langle g_k, d_k\rangle.\end{aligned}$$

Therefore, we have

$$\beta_k = \frac{\Re e\,\langle g_{k+1}, d_{k+1}\rangle}{\Re e\,\langle g_k, d_k\rangle}. \tag{36}$$

Now we need the following lemma which is an extension of Lemma 3.2 in [11].

Lemma 15. *Suppose that ε_1 is a starting point for which Assumption* 14 *is satisfied and consider any method of the form $\varepsilon_{k+1} = \varepsilon_k + \tau_k d_k$ where d_k is a descent direction and τ_k satisfies* (34)–(35). *Then the following Zoutendijk condition holds:*

$$\sum_{k\geqslant 1} \frac{\Re e\,\langle g_k, d_k\rangle^2}{\|d_k\|^2} < \infty.$$

Proof. From (35) it follows that

$$\Re e\,\langle y_k, d_k\rangle = \Re e\,\langle g_{k+1} - g_k, d_k\rangle \geqslant (\sigma - 1)\,\Re e\,\langle g_k, d_k\rangle.$$

The Lipschitz condition implies that

$$\Re e\,\langle y_k, d_k\rangle = \Re e\,\langle g_{k+1} - g_k, d_k\rangle \leqslant \tau_k L \|d_k\|^2.$$

Combining the two inequalities one obtains

$$\tau_k \geqslant \frac{\sigma - 1}{L}\,\frac{\Re e\,\langle g_k, d_k\rangle}{\|d_k\|^2}.$$

This result combined with (34) gives

$$\hat{J}(\varepsilon_k) - \hat{J}(\varepsilon_k + \tau_k d_k) \geqslant -\delta\tau_k \Re e\,\langle g_k, d_k\rangle \geqslant c\,\frac{\Re e\,\langle g_k, d_k\rangle^2}{\|d_k\|^2},$$

where $c = \delta(1 - \sigma)/L$ is a positive constant. Summing up for $k \geqslant 1$ and recalling that $\hat{J}$ is bounded below concludes the proof. □

We can now prove the following theorem representing an extension of Theorem 3.3 in [11].

Theorem 16. *Suppose that ε_1 is a starting point for which Assumption* 14 *holds. Let the sequence $\{\varepsilon_k\}_{k\geqslant 1}$ be generated by the NCG Algorithm* 13. *Then this algorithm either terminates at the stationary point or converges in the sense that*

$$\liminf_{k\to\infty} \|g_k\| = 0.$$

If, in addition, the sequence $\{\|d_k\|/\|g_k\|\}$ is bounded then $\lim_{k\to\infty} \|g_k\| = 0$.

Proof. First notice that if the NCG algorithm does not terminate after finite many iterations, we have

$$\|g_k\| > 0 \quad \text{for all } k.$$

Next we show that the search directions are descent, in the sense that

$$\Re e\,\langle g_k, d_k\rangle < 0. \tag{37}$$

This is obvious for $k = 1$. Now assume that it holds for k and recall the following two results:

$$\Re e\,\langle y_k, d_k\rangle \geqslant (\sigma - 1)\,\Re e\,\langle g_k, d_k\rangle \tag{38}$$

and

$$\Re e\,\langle g_{k+1}, d_{k+1}\rangle = \beta_k\,\Re e\,\langle g_k, d_k\rangle. \tag{39}$$

Notice that $(\sigma - 1)\,\langle g_k, d_k\rangle$ is positive. Therefore, we obtain

$$\Re e\,\langle g_{k+1}, d_{k+1}\rangle = \beta_k\,\Re e\,\langle g_k, d_k\rangle = \frac{\|g_{k+1}\|^2}{\Re e\,\langle d_k, y_k\rangle}\,\Re e\,\langle g_k, d_k\rangle \leqslant \frac{\|g_{k+1}\|^2}{(\sigma - 1)} < 0.$$

That is, (37) holds for all k (and hence $\beta_k > 0$ for all k).

Now, consider the scalar product of both sides of $d_{k+1} + g_{k+1} = \beta_k\, d_k$ with themselves. We have $\|d_{k+1}\|^2 = \beta_k^2 \|d_k\|^2 - 2\Re e\langle g_{k+1}, d_{k+1}\rangle - \|g_{k+1}\|^2$. Dividing with $\Re e\langle g_{k+1}, d_{k+1}\rangle^2$ and using (36), we obtain

$$\begin{aligned}\frac{\|d_{k+1}\|^2}{\Re e\,\langle g_{k+1}, d_{k+1}\rangle^2} &= \frac{\|d_k\|^2}{\Re e\,\langle g_k, d_k\rangle^2} - \left(\frac{1}{\|g_{k+1}\|} + \frac{\|g_{k+1}\|}{\Re e\,\langle g_{k+1}, d_{k+1}\rangle}\right)^2 + \frac{1}{\|g_{k+1}\|^2} \\ &\leqslant \frac{\|d_k\|^2}{\Re e\,\langle g_k, d_k\rangle^2} + \frac{1}{\|g_{k+1}\|^2},\end{aligned}$$

and for $k = 1$ we have $\|d_1\|^2/\Re e\langle g_1, d_1\rangle^2 = 1/\|g_1\|^2$. Therefore, we have

$$\frac{\|d_k\|^2}{\Re e\,\langle g_k, d_k\rangle^2} \leqslant \sum_{i=1}^{k} \frac{1}{\|g_i\|^2} \quad \text{for all } k \geqslant 1. \tag{40}$$

The proof of the theorem is concluded with a contradiction argument. If the theorem is not true, then there exists a constant $c > 0$ such that $\|g_k\| \geqslant c$ for all $k \geqslant 1$ and from (40) it follows that

$$\frac{\|d_k\|^2}{\Re e\,\langle g_k, d_k\rangle^2} \leqslant \frac{k}{c^2},$$

which implies that

$$\sum_{k\geqslant 1} \frac{\Re e\,\langle g_k, d_k\rangle^2}{\|d_k\|^2} \geqslant c^2 \sum_{k\geqslant 1} \frac{1}{k} = \infty,$$

thus contradicting Lemma 15. □

3.2. Cascadic acceleration

The cascadic approach results from combining nested iteration techniques with iterative schemes.

Consider a hierarchy of nested grids with index $\text{lev} = \text{lev}_0, \ldots, \text{lev}_f$. The idea is to start from a coarse grid, with index lev_0, where the size of the problem is small and therefore the problem can be solved by the iterative scheme with a reasonable computational effort. Let us denote with x_{lev_0} the solution obtained by this process with, e.g., 'zero'

initialization $x^*_{\mathrm{lev}_0}$. The step that follows is to interpolate this solution to the next finer grid, using an interpolation operator $I^{\mathrm{lev}+1}_{\mathrm{lev}}$. Therefore, we obtain an initialization of the iterative process on the finer grid that is given by

$$x^*_{\mathrm{lev}+1} = I^{\mathrm{lev}+1}_{\mathrm{lev}} x_{\mathrm{lev}},$$

where $\mathrm{lev} = \mathrm{lev}_0$. With this initialization, after a sufficient number of iterations we obtain the solution $x_{\mathrm{lev}+1}$. This process is repeated until the finest grid is reached and the desired solution is obtained.

An algorithm of this method is given by the following. Denote with $x_{\mathrm{lev}} = S_{\mathrm{lev}}(x^*_{\mathrm{lev}})$ the result of the iteration, with x^*_{lev} as initialization, that is applied until a given stopping criterion is satisfied. We have:

Algorithm 17 (*Cascadic method*). *Step* 1: *Given* $\mathrm{lev} = \mathrm{lev}_0$, $x^*_{\mathrm{lev}_0}$.
Step 2: *Compute* $x_{\mathrm{lev}} = S_{\mathrm{lev}}(x^*_{\mathrm{lev}})$.
Step 3: *If* $\mathrm{lev} = \mathrm{lev}_f$ *then stop*.
Step 4: *Else if* $\mathrm{lev} < \mathrm{lev}_f$ *then interpolate* $x^*_{\mathrm{lev}+1} = I^{\mathrm{lev}+1}_{\mathrm{lev}} x_{\mathrm{lev}}$.
Step 5: *Set* $\mathrm{lev} = \mathrm{lev} + 1$, *goto Step* 2.

The original motivation for using the cascadic approach comes from our previous computational experience and the results given in [4,36] where a cascadic conjugate gradient method is discussed and optimal computational complexity for elliptic problems is proved.

While we are not able to extend the convergence theory in [4,36] to the case of non-convex optimization problems using the NCG scheme, we are able to obtain considerable computational improvement combining our NCG scheme with the cascadic method. We denote the resulting scheme by C-NCG. For results of numerical experiments, see Section 5.

4. Monotonic schemes for quantum control

Monotonic schemes have been initially introduced in a general framework in [17–19]. Following this approach, Tannor et al. [38] and then Zhu and Rabitz [41] have proposed two procedures for quantum control computation. These algorithms have a common basis, as it appears in [23], where a unified framework is presented. Other extensions have then been designed to obtain bounded and bang–bang controls [40], stochastic monotonic schemes [34] or to optimize more general cost functionals [27] and systems involving dissipative states [28]. Recently, a relationship between these algorithms and local trajectory tracking procedures has been established, providing a new interpretation of these schemes [34]. At the theoretical level, some proofs of the convergence of the monotonic schemes have been presented using either compactness and semi-group theory [20] or the Łojasiewicz inequality and its extensions [2]. On the other hand, usual time discretization of the monotonic schemes often leads to instabilities that prevent to reach numerical convergence. This problem is studied in [22,33], where an appropriate time discretization is proposed which avoids instability. In order to tackle the control problem of the finite-level system presented in the section of numerical experiments, we present here a time-discretized monotonic scheme based on a Crank–Nicholson propagator. The resulting algorithm is unconditionally stable and allows us to work with a large range of time steps.

For simplicity, throughout this section we will consider that $\mu = 0$ and require the following assumption that characterizes dissipation in the system.

Assumption 18. We assume that

$$\left\langle \frac{H_0 - H_0^*}{i}\psi, \psi \right\rangle_{L^2(0,T;\mathbb{C}^n)} \leqslant 0 \quad \text{for all } \psi \in H^1(0, T; \mathbb{C}^n).$$

This assumption corresponds to the dissipative character of the system, i.e., the Euclidian norm of the wavefunction decreases with respect to time. For reason of simplicity, we introduce an auxiliary cost functional

$$\widetilde{J}(\varepsilon) := \Re e(\psi_d^* \cdot \psi(T)) - \frac{\gamma}{2}\|\varepsilon\|^2_{L^2(0,T;\mathbb{C})} + \frac{1}{2}\langle \psi, \Lambda\psi\rangle_{L^2(0,T;\mathbb{C}^n)},$$

where ψ and ε are linked by (1) and Λ is defined by

$$(\Lambda\psi)_j = -\alpha_j\psi_j - \left(\frac{H_0 - H_0^*}{i}\psi\right)_j.$$

We consider the optimal control problem corresponding to the maximization of $\widetilde{J}(\varepsilon)$ under the constraint (1). Notice that this problem is equivalent to the optimal control problem (**P**) since

$$\begin{aligned}\hat{J}(\varepsilon) &= |\psi_0|^2_{\mathbb{C}^n} - \Re e(\psi_d^* \cdot \psi(T)) - \frac{1}{2}\langle \psi, \Lambda\psi\rangle_{L^2(0,T;\mathbb{C}^n)} + \frac{\gamma}{2}\|\varepsilon\|^2_{L^2(0,T;\mathbb{C})} \\ &= \frac{|\psi_0|^2_{\mathbb{C}^n} + |\psi_d|^2_{\mathbb{C}^n}}{2} - \widetilde{J}(\varepsilon).\end{aligned}$$

We assume that the following holds.

Assumption 19. The parameters α_j fulfill the condition:

$$-\sum_{j\in I}\alpha_j |\psi_j|^2 - \left\langle \frac{H_0 - H_0^*}{i}\psi, \psi\right\rangle_{L^2(0,T;\mathbb{C}^n)} \geqslant 0 \quad \text{for all } \psi \in \mathbb{C}^n.$$

This assumption means that the parameters α_j have to be calibrated in conjunction with the dissipative character of the system to make the operator Λ positive. This condition is necessary to guarantee the monotonicity of our algorithm (see (43) and Lemma 21).

Before presenting a monotonic scheme corresponding to the maximization of $\widetilde{J}(\varepsilon) = \widetilde{J}(\psi(\varepsilon), \varepsilon)$, we need to compute the variation of this cost functional between two control fields. Let us first define the Lagrange multiplier q by

$$\begin{aligned}&\mathrm{i}\dot{q} = (H_0^* + H_1(\varepsilon(\cdot)))\, q + \Lambda\psi, \\ &\mathrm{i}q(T) = -\psi_d.\end{aligned} \tag{41}$$

Remark 20. Given $\varepsilon \in L^2(0,T;\mathbb{C})$ and ψ the solution to (1), the Lagrange multipliers q and p are linked by the relation

$$\mathrm{i}q = \mathrm{i}p - \psi.$$

Eq. (12g), often used to update ε in optimization procedures, holds when replacing p by q. However, this equivalence is not true if ψ and q are computed with different fields, which is the case during the iterations of the monotonic schemes (see, e.g., Lemma 21).

Consider two fields ε and ε', and the corresponding wavefunctions ψ and ψ', and Lagrange multipliers q and q', respectively. We have

$$\begin{aligned}\widetilde{J}(\varepsilon') - \widetilde{J}(\varepsilon) = {} & \Re e(\psi_d^* \cdot (\psi'(T) - \psi(T))) + \Re e\langle \Lambda\psi, \psi' - \psi\rangle_{L^2(0,T;\mathbb{C}^n)} \\ & + \frac{1}{2}\langle \psi' - \psi, \Lambda(\psi' - \psi)\rangle_{L^2(0,T;\mathbb{C}^n)} \\ & - \frac{\gamma}{2}(\|\varepsilon'\|^2_{L^2(0,T;\mathbb{C})} - \|\varepsilon\|^2_{L^2(0,T;\mathbb{C})}).\end{aligned} \tag{42}$$

Focusing on the first two terms of this sum, we obtain

$$\begin{aligned}&\Re e(\psi_d^* \cdot (\psi'(T) - \psi(T))) + \Re e\,\langle \Lambda\psi, \psi' - \psi\rangle_{L^2(0,T;\mathbb{C}^n)} \\ &\quad = \Re e(\mathrm{i}q(T)^* \cdot (\psi'(T) - \psi(T))) + \Re e\,\langle \Lambda\psi, \psi' - \psi\rangle_{L^2(0,T;\mathbb{C}^n)} \\ &\quad = \Re e\left\langle -\mathrm{i}q, \frac{H_0 + H_1(\varepsilon'(\cdot))}{\mathrm{i}}\psi' - \frac{H_0 + H_1(\varepsilon(\cdot))}{\mathrm{i}}\psi\right\rangle_{L^2(0,T;\mathbb{C}^n)} \\ &\qquad + \Re e\,\langle -\mathrm{i}\dot{q}, \psi' - \psi\rangle_{L^2(0,T;\mathbb{C}^n)} + \Re e\,\langle \Lambda\psi, \psi' - \psi\rangle_{L^2(0,T;\mathbb{C}^n)} \\ &\quad = \Re e\,\langle q, (H_1(\varepsilon'(\cdot)) - H_1(\varepsilon(\cdot)))\psi'\rangle_{L^2(0,T;\mathbb{C}^n)}.\end{aligned}$$

Thus, the variation in $\widetilde{J}$ reads as follows:

$$\widetilde{J}(\varepsilon') - \widetilde{J}(\varepsilon) = \int_0^T \Re e(q(s)^* \cdot (H_1(\varepsilon'(s)) - H_1(\varepsilon(s)))\psi'(s)) - \frac{\gamma}{2}(|\varepsilon'(s)|^2 - |\varepsilon(s)|^2)\,\mathrm{d}s + \frac{1}{2}\langle \psi' - \psi, \Lambda(\psi' - \psi)\rangle_{L^2(0,T;\mathbb{C}^n)}. \tag{43}$$

This formula is the starting point for the design of the monotonic algorithm. For example, we have the following result.

Lemma 21. *Given $\varepsilon \in L^2(0, T; \mathbb{C})$, suppose there exists ε' such that*

$$\varepsilon'_{\Re e}(t) = \frac{1}{\gamma}\,\Re e(q(t)^* \cdot H_{1\Re e}\psi'(t)),$$

$$\varepsilon'_{\Im m}(t) = \frac{1}{\gamma}\,\Re e(q(t)^* \cdot H_{1\Im m}\psi'(t)).$$

Then $\widetilde{J}(\varepsilon') \geqslant \widetilde{J}(\varepsilon)$.

4.1. Time-discretized algorithm

Due to their sequential feature, monotonic schemes require a particular time discretization in order to keep their monotonicity at the discrete level. This discretization is discussed in detail in this section.

For any given integer N, let us introduce the discretization parameter δt defined by $N\,\delta t = T$ and $\varepsilon_\ell = \varepsilon_{\Re e,\ell} + \mathrm{i}\varepsilon_{\Im m,\ell}$, ψ_ℓ, and q_ℓ represents approximations to $\varepsilon(\ell\delta t)$, $\psi(\ell\delta t)$, $q(\ell\delta t)$.

Remark 22. Note that all what follows can be rewritten with (small enough—see Theorem 27) variable time steps. We do not use such a discretization in order to make it simple.

Moreover, we denote by H_ℓ the approximation of the Hamiltonian $H_0 + H_1(\varepsilon(\ell\delta t))$.

Given an initial state ψ_0, we solve numerically (1) and (41) by a Crank–Nicholson scheme. This discretization gives rise to the following iteration:

$$\psi_{\ell+1} = \left(I_d - \frac{\delta t\, H_\ell}{2\mathrm{i}}\right)^{-1}\left(I_d + \frac{\delta t\, H_\ell}{2\mathrm{i}}\right)\psi_\ell \tag{44}$$

and

$$q_{\ell+1} = \left(I_d + \frac{\delta t\, H_\ell^*}{2\mathrm{i}}\right)\left(I_d - \frac{\delta t\, H_\ell^*}{2\mathrm{i}}\right)^{-1} q_\ell - \mathrm{i}\delta t\,\Lambda\psi_{\ell+1},$$
$$\mathrm{i}q_N = -\psi_d, \tag{45}$$

where I_d is the identity matrix. Since $H_1(\varepsilon)$ is hermitian and because of Assumption 18, the following uniform bounds can be obtained:

$$|\psi_\ell|_{\mathbb{C}^n} \leqslant |\psi_0|_{\mathbb{C}^n}, \quad |q_\ell|_{\mathbb{C}^n} \leqslant T\,\rho(\Lambda)\,|\psi_0|_{\mathbb{C}^n} + |\psi_d|_{\mathbb{C}^n} \quad \text{for all } \ell = 0, \ldots, N. \tag{46}$$

We also introduce the time-discretized cost functional

$$\widetilde{J}_{\delta t}(\varepsilon) = \Re e(\psi_d^* \cdot \psi_N) - \frac{\gamma\delta t}{2}\sum_{\ell=0}^{N-1}|\varepsilon_\ell|^2 + \frac{\delta t}{2}\sum_{\ell=0}^{N-1}(\Lambda\psi_{\ell+1}^* \cdot \psi_{\ell+1}).$$

Remark 23. Here, we use a first-order approximation for the integrals appearing in the cost functional for reason of simplicity. Yet, our approach can be applied with higher order approximations.

Consider two control fields ε and ε'. Repeating the computations of the previous section at the discrete level, we obtain the following equivalent to (43):

$$\widetilde{J}_{\delta t}(\varepsilon') - \widetilde{J}_{\delta t}(\varepsilon) = \sum_{\ell=0}^{N-1} \Re e(q_\ell^* \cdot \mathscr{D}H_\ell \psi'_\ell) - \frac{\gamma \delta t}{2}(|\varepsilon'_\ell|^2 - |\varepsilon_\ell|^2) + \frac{\delta t}{2} \sum_{\ell=0}^{N-1} ((\psi'_\ell - \psi_\ell)^* \cdot \Lambda(\psi'_\ell - \psi_\ell)), \tag{47}$$

where

$$\mathscr{D}H_\ell = \mathrm{i}\left(\left(I_d + \frac{\delta t H_\ell}{2\mathrm{i}}\right)^{-1}\left(I_d - \frac{\delta t H_\ell}{2\mathrm{i}}\right)\left(I_d - \frac{\delta t H'_\ell}{2\mathrm{i}}\right)^{-1}\left(I_d + \frac{\delta t H'_\ell}{2\mathrm{i}}\right) - I_d\right).$$

We now present a monotonic scheme to optimize $\widetilde{J}_{\delta t}$. First, notice that given $\varepsilon = (\varepsilon_\ell)_{0 \leqslant \ell \leqslant N-1}$ and ψ'_ℓ, the term $\Re e(q_\ell^* \cdot \mathscr{D}H_\ell \psi'_\ell) - \delta t \frac{\gamma}{2}(|\varepsilon'_\ell|^2 - |\varepsilon_\ell|^2)$ in (47) depends only on ε'_ℓ. Starting from this remark, the algorithm we propose consists in optimizing recursively each term of the first sum in (47) with respect to ε'_ℓ via one iteration of a Newton method.

Let us compute a Taylor expansion with respect to $\delta\varepsilon_\ell = \varepsilon'_\ell - \varepsilon_\ell = \delta\varepsilon_{\Re e,\ell} + \mathrm{i}\delta\varepsilon_{\Im m,\ell}$. Defining $\Delta H_\ell = H_1(\varepsilon') - H_1(\varepsilon)$, we have

$$(q_\ell^* \cdot \mathscr{D}H_\ell \psi'_\ell) = \frac{\delta t}{2}(\widetilde{q}_\ell^* \cdot \Delta H_\ell \check{\psi}'_\ell) - \mathrm{i}\,\frac{\delta t^2}{4}\left[\widetilde{q}_\ell^* \cdot \Delta H_\ell \left(I_d - \frac{\delta t H_\ell}{2\mathrm{i}}\right)^{-1} \Delta H_\ell \check{\psi}'_\ell\right] + \mathrm{o}((\Delta H_\ell)^2),$$

where

$$\widetilde{q}_\ell = \left(I_d + \left(I_d + \frac{\delta t H_\ell^*}{2\mathrm{i}}\right)\left(I_d - \frac{\delta t H_\ell^*}{2\mathrm{i}}\right)^{-1}\right) q_\ell = q_\ell + q_{\ell+1} + \mathrm{i}\delta t \Lambda \psi_{\ell+1},$$

$$\check{\psi}'_\ell = \left(I_d - \frac{\delta t H_\ell}{2\mathrm{i}}\right)^{-1} \psi'_\ell.$$

This expression can be rewritten in terms of the quantity $\delta\varepsilon_\ell$:

$$(q_\ell^* \cdot \mathscr{D}H_\ell \psi'_\ell) = \frac{\delta t}{2} A_\ell^\dagger \begin{pmatrix} \delta\varepsilon_{\Re e,\ell} \\ \delta\varepsilon_{\Im m,\ell} \end{pmatrix} - \frac{\mathrm{i}\delta t^2}{4} \begin{pmatrix} \delta\varepsilon_{\Re e,\ell} \\ \delta\varepsilon_{\Im m,\ell} \end{pmatrix}^\dagger B_\ell \begin{pmatrix} \delta\varepsilon_{\Re e,\ell} \\ \delta\varepsilon_{\Im m,\ell} \end{pmatrix} + \mathrm{o}((\delta\varepsilon_\ell)^2),$$

where † stands for transpose, and the matrices A_ℓ and B_ℓ are given by

$$A_\ell = \begin{pmatrix} (\widetilde{q}_\ell^* \cdot H_{1\Re e} \check{\psi}'_\ell) \\ (\widetilde{q}_\ell^* \cdot H_{1\Im m} \check{\psi}'_\ell) \end{pmatrix}, \quad B_\ell = \begin{pmatrix} B_{\ell,1,1} & B_{\ell,1,2} \\ B_{\ell,2,1} & B_{\ell,2,2} \end{pmatrix}, \tag{48}$$

with

$$B_{\ell,1,1} = \left(\widetilde{q}_\ell^* \cdot H_{1\Re e}\left(I_d - \frac{\delta t H_\ell}{2\mathrm{i}}\right)^{-1} H_{1\Re e} \check{\psi}'_\ell\right)$$

$$B_{\ell,2,2} = \left(\widetilde{q}_\ell^* \cdot H_{1\Im m}\left(I_d - \frac{\delta t H_\ell}{2\mathrm{i}}\right)^{-1} H_{1\Im m} \check{\psi}'_\ell\right)$$

$$B_{\ell,1,2} = B_{\ell,2,1} = \frac{1}{2}\left[\widetilde{q}_\ell^* \cdot \left\{ H_{1\Im m}\left(I_d - \frac{\delta t\, H_\ell}{2\mathrm{i}}\right)^{-1} H_{1\Re e}\check{\psi}'_\ell + H_{1\Re e}\left(I_d - \frac{\delta t\, H_\ell}{2\mathrm{i}}\right)^{-1} H_{1\Im m}\check{\psi}'_\ell \right\}\right].$$

On the other hand, one has

$$|\varepsilon'_\ell|^2 - |\varepsilon_\ell|^2 = 2\begin{pmatrix}\varepsilon_{\Re e,\ell}\\ \varepsilon_{\Im m,\ell}\end{pmatrix}^\dagger \begin{pmatrix}\delta\varepsilon_{\Re e,\ell}\\ \delta\varepsilon_{\Im m,\ell}\end{pmatrix} + \begin{pmatrix}\delta\varepsilon_{\Re e,\ell}\\ \delta\varepsilon_{\Im m,\ell}\end{pmatrix}^\dagger \begin{pmatrix}\delta\varepsilon_{\Re e,\ell}\\ \delta\varepsilon_{\Im m,\ell}\end{pmatrix},$$

and the corresponding variation in the cost functional reads

$$\begin{aligned}\widetilde{J}_{\delta t}(\varepsilon') - \widetilde{J}_{\delta t}(\varepsilon) = \delta t \sum_{\ell=0}^{N-1} &\left(\frac{1}{2}\,\Re e A_\ell - \gamma \begin{pmatrix}\varepsilon_{\Re e,\ell}\\ \varepsilon_{\Im m,\ell}\end{pmatrix}\right)^\dagger \begin{pmatrix}\delta\varepsilon_{\Re e,\ell}\\ \delta\varepsilon_{\Im m,\ell}\end{pmatrix}\\ &- \begin{pmatrix}\delta\varepsilon_{\Re e,\ell}\\ \delta\varepsilon_{\Im m,\ell}\end{pmatrix}^T \left(-\frac{\delta t}{4}\,\Im m B_\ell + \frac{\gamma}{2} I_2\right)\begin{pmatrix}\delta\varepsilon_{\Re e,\ell}\\ \delta\varepsilon_{\Im m,\ell}\end{pmatrix} + \mathrm{o}((\delta\varepsilon_\ell)^2)\\ &+ \frac{\delta t}{2}\sum_{\ell=0}^{N-1}((\psi'_\ell - \psi_\ell)^* \cdot \Lambda(\psi'_\ell - \psi_\ell)),\end{aligned}$$

where I_2 denotes the identity matrix of $\mathbb{R}^2$.

Algorithm 24 (*Crank–Nicholson monotonic scheme (CNMS)*). *Let us denote by* $\|.\|_d$ *the Euclidian norm on* $\mathbb{R}^2$. *Given an initial control amplitude* ε^0, *its associated state* ψ^0 *and Lagrange multiplier* q^0 *and a tolerance* $\mathrm{tol} > 0$, *suppose that* ψ^k, q^k, ε^k, *have already been computed. The derivation of* ψ^{k+1}, q^{k+1}, ε^{k+1}, *is done as follows*:

Step 1. *If* $\sum_{\ell=0}^{N-1}\left\|\Re e\begin{pmatrix}((\widetilde{q}_\ell^k)^* \cdot H_{1\Re e}\check{\psi}_\ell^k)\\ ((\widetilde{q}_\ell^k)^* \cdot H_{1\Im m}\check{\psi}_\ell^k)\end{pmatrix} - \gamma\begin{pmatrix}\varepsilon_{\Re e,\ell}^k\\ \varepsilon_{\Im m,\ell}^k\end{pmatrix}\right\|_d \leqslant \mathrm{tol}$, *then stop*.

Step 2. *Forward propagation*: *Given* $\psi_0^{k+1} = \psi_0$, *compute recursively* $\psi_{\ell+1}^{k+1}$ *from* ψ_ℓ^{k+1} *by*

Step 2.1. (*Newton iteration*) *Compute* ε_ℓ^{k+1} *by*

$$\begin{pmatrix}\varepsilon_{\Re e,\ell}^{k+1}\\ \varepsilon_{\Im m,\ell}^{k+1}\end{pmatrix} = \begin{pmatrix}\varepsilon_{\Re e,\ell}^{k}\\ \varepsilon_{\Im m,\ell}^{k}\end{pmatrix} + \frac{1}{2}\left(-\frac{\delta t}{4}\,\Im m B_\ell^k + \frac{\gamma}{2} I_2\right)^{-1}\left(\frac{1}{2}\,\Re e A_\ell^{k+1} - \gamma\begin{pmatrix}\varepsilon_{\Re e,\ell}^{k}\\ \varepsilon_{\Im m,\ell}^{k}\end{pmatrix}\right), \tag{49}$$

where A^{k+1} *is defined by* (48), *with* $\widetilde{q} = \widetilde{q}^k$ *and* $\check{\psi}' = \check{\psi}^{k+1}$.

Step 2.2. *Compute* $\psi_{\ell+1}^{k+1}$ *by* (44).

Step 3. *Backward propagation*: *Given* $q_N^{k+1} = \mathrm{i}\psi_d$, *compute recursively* q_ℓ^{k+1} *from* $q_{\ell+1}^{k+1}$ *by* (45).

Step 4. *Go back to Step* 1.

4.2. Convergence of the algorithm

We present some results concerning the convergence of the CNMS. An important property of this scheme is that the sequence $(\varepsilon^k)_{k\in\mathbb{N}}$ is bounded, as claimed in the next lemma.

Lemma 25. *For small enough* δt, *there exists* M, *such that*

$$\forall k \in \mathbb{N}, \quad \forall \ell = 0 \ldots N-1, \quad |\varepsilon_{\Re e,\ell}^k| \leqslant M, \quad |\varepsilon_{\Im m,\ell}^k| \leqslant M.$$

Proof. Consider an initial control field ε^0 and define m and M by

$$m = 2K_0|\psi_0|_{\mathbb{C}^n}(|\psi_d|_{\mathbb{C}^n} + \rho(\Lambda)(T + \delta t|\psi_0|_{\mathbb{C}^n})),$$

$$\lambda = \frac{2m}{\gamma} + \max_{\ell=0\ldots N-1} |\varepsilon_\ell^0|,$$

$$M = 3\lambda.$$

Let us denote by $A_{\Re e,\ell}$ and $A_{\Im m,\ell}$ the components of A_ℓ. Given ε such that

$$\max_{\ell=0\ldots N-1} |\varepsilon_{\Re e,\ell}| \leqslant M, \qquad \max_{\ell=0\ldots N-1} |\varepsilon_{\Im m,\ell}| \leqslant M, \tag{50}$$

the estimates (46) and the definition (48) give

$$\max(|A_{\Re e,\ell}|, |A_{\Im m,\ell}|) \leqslant \frac{m}{1 - \delta t((\rho(H_0) + K_0\sqrt{2}M)/2)}.$$

A similar estimate can be obtained for the coefficients of B_ℓ, when ε satisfies (50). Let us denote by b the bound obtained in this case.

The iteration (49) reads

$$\varepsilon_\ell^{k+1} = (I_2 - \delta_\ell^k)\varepsilon_\ell^k + \frac{\delta_\ell^k}{2\gamma}\,\Re e\, A_\ell^k,$$

where

$$\delta_\ell^k = \left(-\frac{\delta t}{2\gamma}\,\Im m\, B_\ell^k + I_2\right)^{-1}.$$

Now suppose that, for given $k, \ell \in \mathbb{N}$, $\max(|\varepsilon^k_{\Re e,\ell}|, |\varepsilon^k_{\Im m,\ell}|) \leqslant M$ and that δt is such that:

$$\frac{m}{1 - \delta t((\rho(H_0) + K_0\sqrt{2}M)/2)} \leqslant \gamma\lambda.$$

Since the coefficients of B_ℓ^k are bounded by b, it can be supposed that the diagonal coefficients of δ_ℓ^k belong to $[\frac{1}{2}, \frac{3}{2}]$ and the others to $[-\frac{1}{8}, \frac{1}{8}]$. Then

$$\max(|A_{\Re e,\ell}|, |A_{\Im m,\ell}|) \leqslant \gamma\lambda.$$

Consequently:

$$\begin{aligned}
|\varepsilon^{k+1}_{\Re e,\ell}| &\leqslant |1 - \delta^k_{\ell,1,1}||\varepsilon^k_{\Re e,\ell}| + |\delta^k_{\ell,1,2}||\varepsilon^k_{\Im m,\ell}| + \frac{|\delta^k_{\ell,1,1}|}{2\gamma}|\Re e A_{\Re e,\ell}| + \frac{|\delta^k_{\ell,1,2}|}{2\gamma}|\Re e A_{\Im m,\ell}| \\
&\leqslant |1 - \delta^k_{\ell,1,1}||\varepsilon^k_{\Re e,\ell}| + |\delta^k_{\ell,1,2}||\varepsilon^k_{\Im m,\ell}| + \frac{|\delta^k_{\ell,1,1}|}{2}\lambda + \frac{|\delta^k_{\ell,1,2}|}{2}\lambda \\
&\leqslant \left(|1 - \delta^k_{\ell,1,1}| + |\delta^k_{\ell,1,2}| + \frac{|\delta^k_{\ell,1,1}| + |\delta^k_{\ell,1,2}|}{6}\right) M \\
&\leqslant M,
\end{aligned}$$

where $\delta^k_{\ell,i,j}$ denotes the components of δ^k_ℓ. The same result holds for $|\varepsilon^{k+1}_{\Im m,\ell}|$. The lemma follows. □

The next lemma gives a result about the monotonicity of CNMS.

Lemma 26. *For small enough δt, the CNMS converges monotonically, in the sense that*

$$\exists \eta > 0, \quad \widetilde{J}_{\delta t}(\varepsilon^{k+1}) - \widetilde{J}_{\delta t}(\varepsilon^{k}) \geqslant \eta \|\varepsilon^{k+1} - \varepsilon^{k}\|^2_{\mathbb{C}^N}.$$

Proof. The variation in $\widetilde{J}_{\delta t}$ between two iterations reads

$$\widetilde{J}_{\delta t}(\varepsilon^{k+1}) - \widetilde{J}_{\delta t}(\varepsilon^{k}) = \frac{\delta t \gamma}{2} \sum_{\ell=0}^{N-1} |\delta\varepsilon_\ell^k|^2 + \frac{\delta t}{2} \sum_{\ell=0}^{N-1} ((\psi'_\ell - \psi_\ell)^* \cdot \Lambda(\psi'_\ell - \psi_\ell)) + \sum_{\ell=0}^{N-1} \Re e\, R_\ell^k,$$

where

$$R_\ell^k = (q_\ell^{k\,*} \cdot \mathscr{D} H_\ell \psi_\ell^{k+1}) - \frac{\delta t}{2} (A_\ell^k)^\dagger \begin{pmatrix} \delta\varepsilon^k_{\Re e,\ell} \\ \delta\varepsilon^k_{\Im m,\ell} \end{pmatrix} + \mathrm{i}\, \frac{\delta t^2}{4} \begin{pmatrix} \delta\varepsilon^k_{\Re e,\ell} \\ \delta\varepsilon^k_{\Im m,\ell} \end{pmatrix}^\dagger B_\ell^k \begin{pmatrix} \delta\varepsilon^k_{\Re e,\ell} \\ \delta\varepsilon^k_{\Im m,\ell} \end{pmatrix}.$$

Using the Taylor–Lagrange formula and Lemma 25, we obtain

$$R_\ell^k = |\varepsilon_\ell^k|^2\, \mathrm{o}(\delta t),$$

and the result follows. □

Further analysis shows that there exists $\upsilon > 0$ such that

$$\|\nabla \widetilde{J}_{\delta t}(\varepsilon^{k+1})\|_{\mathbb{C}^N} \leqslant \upsilon \|\varepsilon^{k+1} - \varepsilon^{k}\|_{\mathbb{C}^N},$$

where $\nabla \widetilde{J}_{\delta t}$ denotes the gradient of $\widetilde{J}_{\delta t}$ with respect to ε. This fact combined with Lemma 26 enables us to claim the following convergence result.

Theorem 27. *For all initial value ε^0, the sequence $(\varepsilon^k)_{k\in\mathbb{N}}$ converges toward a critical point of $\widetilde{J}_{\delta t}$. Denoting by ε^∞ this limit, there exists $c > 0$ and $\chi > 0$ such that*:

$$\|\varepsilon^k - \varepsilon^\infty\|_{\mathbb{C}^N} \leqslant c k^{-\chi}.$$

Note that, for large values of γ, it can be proved that the convergence rate is indeed linear. We refer to [33] for the details of the proof.

5. Numerical experiments

We present results of numerical experiments considering the control of a representative three-level quantum system. The main purpose of this section is to show that the proposed cascadic NCG scheme is efficient and robust with respect to the choice of the optimization parameters. In fact, at least with the model considered in this section, the cascadic NCG scheme results more efficient and robust than the monotonic scheme. With the improved computational performance obtained with the cascadic NCG scheme we are able to solve the quantum optimal control problem to very small residual and investigate the tracking performance of our optimal control formulation also in the case of very small regularization (optimization) parameters.

Consider a three-level quantum system whose configuration is represented by $\psi = (\psi_1, \psi_2, \psi_3) \in H^1(0, T; \mathbb{C}^3)$ which consists of two long-lived states ψ_1 and ψ_2, which are energetically separated by some amount δ, and a state ψ_3, which has a finite lifetime because of environment coupling (wiggled line); see Fig. 1. Such Λ-type configurations have a long-standing history in quantum optics and have been demonstrated successful in the explanation of many coherence-phenomena in atomic systems [3]; more recently, similar configurations have received increasing attention also in semiconductor quantum dots [6,16].

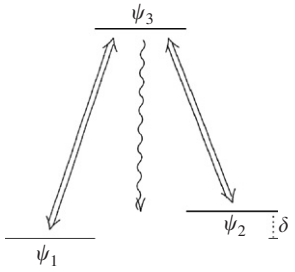

Fig. 1. Prototypical Λ-type three-level scheme: ψ_1 and ψ_2 are long-lived states whereas ψ_3 is a short-lived state which is optically coupled to both ψ_1 and ψ_2; wiggled line indicates relaxation and dephasing of state ψ_3.

Time evolution of this finite-level quantum system is governed by the following Hamiltonian [3,16]:

$$H_0 = \frac{1}{2}\begin{pmatrix} -\delta & 0 & 0 \\ 0 & \delta & 0 \\ 0 & 0 & -\mathrm{i}\gamma_0 \end{pmatrix}, \tag{51}$$

where the term $-\mathrm{i}\gamma_0$ accounts for environment losses (e.g., spontaneous photon emissions). The coupling to the external field $\varepsilon = \varepsilon_{\Re e} + \mathrm{i}\varepsilon_{\Im m}$, reads

$$H_1(\varepsilon) = -\frac{1}{2}\begin{pmatrix} 0 & 0 & \mu_1\varepsilon \\ 0 & 0 & \mu_2\varepsilon \\ \mu_1\varepsilon^* & \mu_2\varepsilon^* & 0 \end{pmatrix}, \tag{52}$$

where μ_1 and μ_2 describe the coupling strengths of states ψ_1 and ψ_2 to the inter-connecting state ψ_3 (e.g., optical dipole matrix elements).

Typical initial and final states are given by

$$\psi_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \psi_d = \begin{pmatrix} 0 \\ \mathrm{e}^{-\mathrm{i}\delta t} \\ 0 \end{pmatrix},$$

respectively.

Concerning the optimality condition (12g) we have

$$H_{1\Re e} = -\frac{1}{2}\begin{pmatrix} 0 & 0 & \mu_1 \\ 0 & 0 & \mu_2 \\ \mu_1 & \mu_2 & 0 \end{pmatrix} \quad \text{and} \quad H_{1\Im m} = -\frac{1}{2}\begin{pmatrix} 0 & 0 & \mathrm{i}\mu_1 \\ 0 & 0 & \mathrm{i}\mu_2 \\ -\mathrm{i}\mu_1 & -\mathrm{i}\mu_2 & 0 \end{pmatrix}.$$

Assuming that the system is initially prepared in state ψ_0, we use the optimal control approach to determine the most efficient way to bring the system close to ψ_d. The form of $H_1(\varepsilon)$ is such that direct optical transition between ψ_1 and ψ_2 is forbidden. The presence of the third auxiliary state ψ_3 allows this transition through intermediate population transfer while introducing losses because of environment coupling. Therefore, we require to find a sequence of laser pulses that minimizes the population of level ψ_3 along evolution. Therefore, we have $I = \{3\}$ in the cost functional (5) and we denote $\alpha_3 = \alpha$.

To determine the evolution of state and adjoint variables we consider an implicit second-order Crank–Nicholson scheme. Given the solution at time step ℓ, the value of the wave function at the next time step $\ell + 1$ is given by

$$\mathrm{i}\frac{\psi_{\ell+1} - \psi_\ell}{\delta t} = \frac{1}{2}H_{\ell+1}\psi_{\ell+1} + \frac{1}{2}H_\ell\psi_\ell.$$

Thus $\psi_{\ell+1}$ is given by

$$\psi_{\ell+1} = (I_d - \frac{\delta t}{2\mathrm{i}}H_{\ell+1})^{-1}\left(I_d + \frac{\delta t}{2\mathrm{i}}H_\ell\right)\psi_\ell, \quad \ell = 0, \ldots, N-1.$$

Table 1
Optimization results for different values of tolerance

tol	$\|\text{res}_\varepsilon\|$	$\|\psi(T) - \psi_d\|_{\mathbb{C}^3}$	J	CPU
10^{-4}	$2.93 \cdot 10^{-3}$	$5.46 \cdot 10^{-3}$	$7.69 \cdot 10^{-2}$	1.09
10^{-6}	$1.49 \cdot 10^{-3}$	$2.99 \cdot 10^{-3}$	$6.77 \cdot 10^{-3}$	3.04
10^{-8}	$2.39 \cdot 10^{-5}$	$3.13 \cdot 10^{-3}$	$6.77 \cdot 10^{-3}$	5.43
10^{-10}	$5.23 \cdot 10^{-6}$	$3.15 \cdot 10^{-3}$	$6.77 \cdot 10^{-3}$	8.18
10^{-12}	$8.84 \cdot 10^{-7}$	$3.15 \cdot 10^{-3}$	$6.77 \cdot 10^{-3}$	28.29

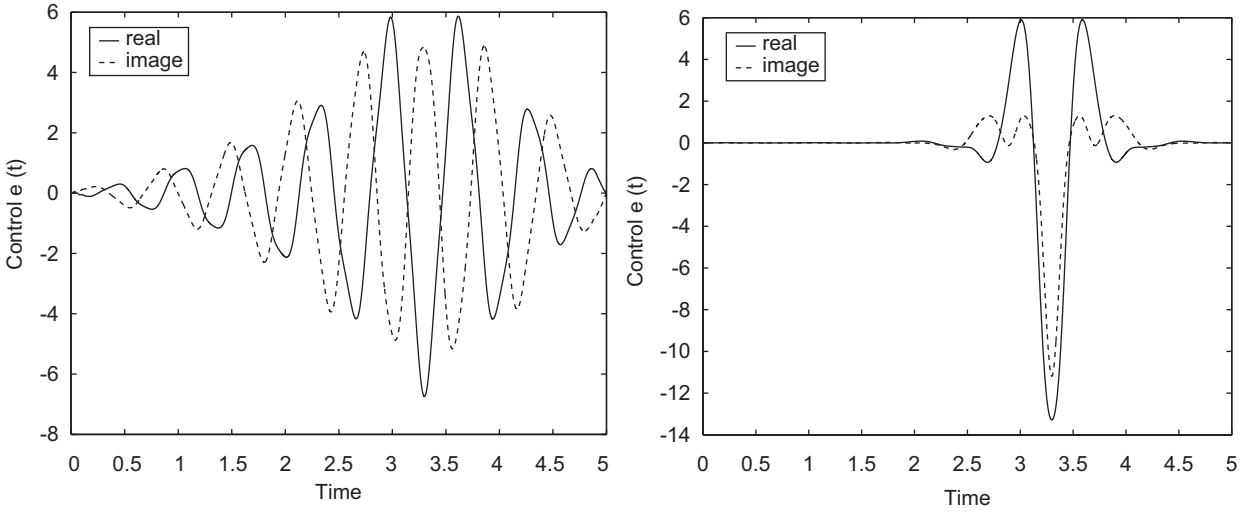

Fig. 2. Control fields computed with different choices of tolerance tol $= \{10^{-4}, 10^{-8}\}$ (from left to right).

In our case of a three-level quantum system, the operator $(I_d + \mathrm{i}(\delta t/2)H_{\ell+1})$ is a 3×3 complex matrix which is easily invertible. The formula above holds for the adjoint equation marching backwards by inverting the time direction. Notice that this scheme is slightly different from the one presented in Section 4.1, this latter being relevant to the sequential feature of the CNMS scheme. However, in both the cases we expect second-order accuracy of solutions. This is demonstrated by the results of numerical experiments given below.

Some of problems' parameters are held fixed. We choose $\delta = 20$, $\gamma_0 = 0.01$, $\mu_1 = 1$, $\mu_2 = 1$, $\alpha_1 = 0$, $\alpha_2 = 0$, and $T = 5$. N is the number of time steps and $\delta t = T/N$. The figures are computed with $N = 4096$.

First we focus on the problem of assessing convergence of iterative solvers to quantum optimal control solutions. We give evidence that optimal solutions are quite sensitive to the order of tolerance required in computations. For this purpose consider the results reported in Table 1. These results have been obtained using the convergence criterion

$$\frac{|J^{k+1} - J^k|}{J^k} \leqslant \text{tol},$$

where J^k is the value of the reduced cost functional after k iterations. This criterion is commonly used in the scientific computing community and typical values for the tolerance are tol $\in (10^{-6}, 10^{-4})$. One should notice that in the case of minimization problems with flat minima basins, such convergence criteria may be misleading. This fact can be partly seen in Table 1 considering the values of $|\psi(T) - \psi_d|_{\mathbb{C}^3}$ and of J. In Fig. 2, a more dramatic picture is given of how sensitive the optimal solution is with respect to the chosen tolerance.

On the other hand, resulting values of the discrete $L^2(0, T; \mathbb{C})$-norm (denoted by $\|\cdot\|$) of the residual of the optimality conditions (12g), $\|\text{res}_\varepsilon\|$, suggest to use this value in order to have a robust convergence criteria. In fact, we use the criteria as given in NCG Algorithm 13, that is,

$$\|g_{k+1}\| < \text{tol}_{\text{abs}} \quad \text{and} \quad \|g_{k+1}\| < \text{tol}_{\text{rel}}\, \|g_1\|,$$

where we take $\text{tol}_{\text{rel}} = 10\,\text{tol}_{\text{abs}}$.

Table 2
Approximation results for different meshes; $N = 2^{\text{lev}}$

	N	$\|e_{\psi_1,\text{lev}-1}\|$	$\|e_{\psi_2,\text{lev}-1}\|$	$\|e_{\psi_3,\text{lev}-1}\|$	$\|e_{\varepsilon,\text{lev}-1}\|$
$\text{tol}_{\text{abs}} = 10^{-7}$	512	$1.4 \cdot 10^{-2}$	$5.4 \cdot 10^{-3}$	$4.2 \cdot 10^{-3}$	$6.1 \cdot 10^{-13}$
	1024	$3.5 \cdot 10^{-3}$	$1.3 \cdot 10^{-3}$	$1.0 \cdot 10^{-3}$	$6.8 \cdot 10^{-3}$
	2048	$8.9 \cdot 10^{-4}$	$3.3 \cdot 10^{-4}$	$2.6 \cdot 10^{-4}$	$1.7 \cdot 10^{-3}$
	4096	$2.2 \cdot 10^{-4}$	$8.7 \cdot 10^{-5}$	$7.0 \cdot 10^{-5}$	$3.6 \cdot 10^{-4}$
$\text{tol}_{\text{abs}} = 10^{-5}$					
	512	$1.4 \cdot 10^{-2}$	$5.2 \cdot 10^{-3}$	$4.2 \cdot 10^{-3}$	$6.3 \cdot 10^{-13}$
	1024	$3.5 \cdot 10^{-3}$	$1.4 \cdot 10^{-3}$	$1.1 \cdot 10^{-3}$	$6.0 \cdot 10^{-3}$
	2048	$8.6 \cdot 10^{-4}$	$3.7 \cdot 10^{-4}$	$3.2 \cdot 10^{-4}$	$1.3 \cdot 10^{-3}$
	4096	$2.2 \cdot 10^{-4}$	$8.1 \cdot 10^{-5}$	$7.2 \cdot 10^{-5}$	$3.7 \cdot 10^{-4}$
$\text{tol}_{\text{abs}} = 10^{-3}$					
	512	$1.4 \cdot 10^{-2}$	$7.2 \cdot 10^{-3}$	$5.5 \cdot 10^{-3}$	$6.5 \cdot 10^{-13}$
	1024	$3.5 \cdot 10^{-3}$	$1.3 \cdot 10^{-3}$	$1.9 \cdot 10^{-3}$	$4.4 \cdot 10^{-3}$
	2048	$9.2 \cdot 10^{-4}$	$5.7 \cdot 10^{-4}$	$5.2 \cdot 10^{-4}$	$8.2 \cdot 10^{-4}$
	4096	$2.3 \cdot 10^{-4}$	$9.8 \cdot 10^{-5}$	$1.2 \cdot 10^{-4}$	$3.6 \cdot 10^{-4}$

Table 3
Residuals on mesh $N = 8192$ for different values of tolerance

tol	$\|\text{res}_\varepsilon\|$	$\|\text{res}_\psi\|$	$\|\text{res}_p\|$
10^{-4}	$2.9 \cdot 10^{-3}$	$6.5 \cdot 10^{-13}$	$5.9 \cdot 10^{-15}$
10^{-6}	$1.5 \cdot 10^{-3}$	$6.3 \cdot 10^{-13}$	$4.2 \cdot 10^{-15}$
10^{-8}	$2.4 \cdot 10^{-5}$	$6.3 \cdot 10^{-13}$	$4.2 \cdot 10^{-15}$
10^{-10}	$5.2 \cdot 10^{-6}$	$6.2 \cdot 10^{-13}$	$4.1 \cdot 10^{-15}$
10^{-12}	$8.8 \cdot 10^{-7}$	$6.1 \cdot 10^{-13}$	$4.0 \cdot 10^{-15}$

Table 4
Computational efforts of the NCG scheme and the CNMS scheme for different choices of tolerance

tol_{abs}	$N = 2048$		$N = 4096$	
	CPU(NCG)	CPU(CNMS)	CPU(NCG)	CPU(CNMS)
10^{-4}	1.17	1.28	2.32	1.39
10^{-5}	4.32	12.63	9.26	15.92
10^{-6}	5.01	48.00	17.21	*no conv*

Table 5
Computational effort to solve for $\text{tol}_{\text{abs}} = 10^{-6}$; $\gamma_0 = 0.01$, $\alpha_3 = 0.05$; in C-NCG coarsest level $N = 1024$

N	$\gamma = 10^{-4}$		$\gamma = 10^{-6}$	
	CPU(NCG)	CPU(C-NCG)	CPU(NCG)	CPU(C-NCG)
4096	40.54	6.26	254.70	58.10
8192	112.57	12.71	319.46	134.00
16 384	312.17	27.42	626.84	279.46

While analysis of accuracy of numerical approximation of quantum optimal control problems deserves additional effort in a separate work, we provide results of experiments that demonstrate that using second-order marching schemes for the optimality system results in second-order accurate solutions. To show this fact, and since it is difficult to define an exact solution for this class of problems, we adopt the following strategy [8] that can be viewed as an additional criteria of convergence to the optimal solution.

Consider a hierarchy of nested meshes with $N = 2^{\text{lev}}$ where lev is the level index. We take the solution on a fine mesh lev as the reference solution for computing the error on the next coarser mesh lev $-$ 1 as follows $e_{\psi^k,\text{lev}-1} = \psi_{k,\text{lev}-1} -$

Table 6
Optimization results depending on optimization parameters, $\mathrm{tol}_{\mathrm{abs}} = 10^{-7}$

γ	μ	α	$\lvert\psi(T) - \psi_d\rvert_{\mathbb{C}^3}$	J	CPU
10^{-7}	10^{-7}	0.05	$8.6 \cdot 10^{-4}$	$2.37 \cdot 10^{-3}$	19.6
10^{-7}	10^{-9}	0.05	$3.7 \cdot 10^{-4}$	$5.46 \cdot 10^{-4}$	55.6
10^{-7}	0	0.05	$6.9 \cdot 10^{-5}$	$1.41 \cdot 10^{-4}$	424.8
10^{-7}	0	0	$1.2 \cdot 10^{-3}$	$2.33 \cdot 10^{-6}$	763.1
10^{-4}	10^{-4}	0.05	$3.3 \cdot 10^{-2}$	$6.52 \cdot 10^{-2}$	47.3
10^{-4}	10^{-6}	0.05	$4.4 \cdot 10^{-3}$	$9.03 \cdot 10^{-3}$	42.3
10^{-4}	0	0.05	$2.7 \cdot 10^{-3}$	$5.68 \cdot 10^{-3}$	17.2
10^{-4}	0	0	$8.3 \cdot 10^{-3}$	$3.34 \cdot 10^{-4}$	5.5

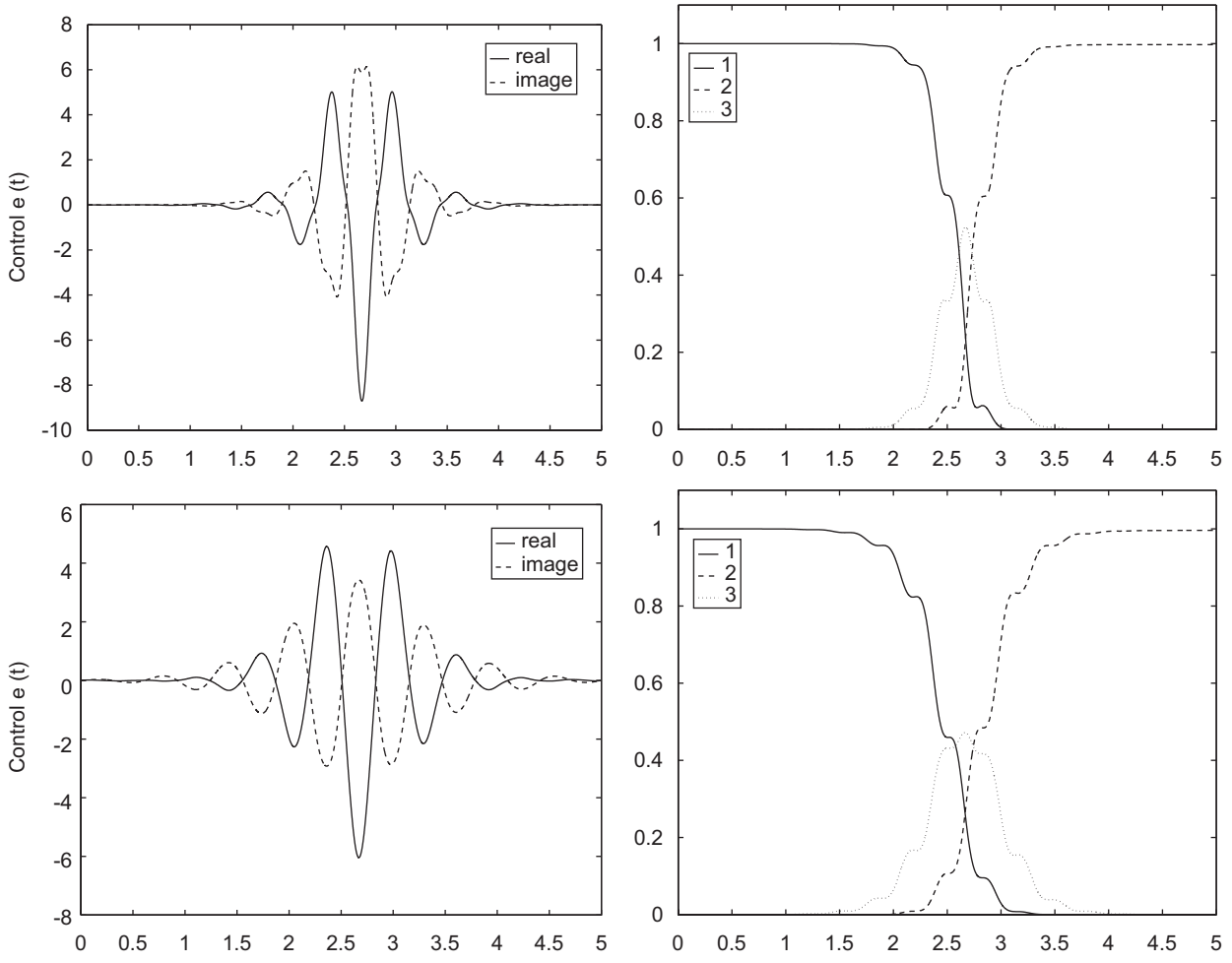

Fig. 3. Optimal solutions for $\gamma = 10^{-4}$ and $\mu = 0$ (top) and $\mu = 10^{-6}$ (bottom) $\alpha_3 = 0.01$. Left, the control field, right, the wave function.

$I_{\mathrm{lev}}^{\mathrm{lev}-1}\psi_{k,\mathrm{lev}}$, $k = 1, 2, 3$, and $e_{\varepsilon,\mathrm{lev}-1} = \varepsilon_{\mathrm{lev}-1} - I_{\mathrm{lev}}^{\mathrm{lev}-1}\varepsilon_{\mathrm{lev}}$ where $I_{\mathrm{lev}}^{\mathrm{lev}-1}$ is injection. In Table 2 we report results for different meshes and different values of tolerance obtained with the NCG scheme. We see that for sufficiently small $\mathrm{tol}_{\mathrm{abs}}$ second-order accuracy is attained. This result suggests a way to assess convergence of optimal solutions. In Table 3 we report values of norm of residuals. Because at each time step the exact solution of the implicit Crank–Nicholson scheme is performed, residuals of state and adjoint equations are zero. On the other hand, we see again that $\|\mathrm{res}_\varepsilon\|$ is representative of the attained accuracy of the solution.

In Table 4 results are reported to compare the computational performance of the NCG scheme (without cascadic acceleration) and the CNMS scheme for different choices of tolerance and mesh sizes. We see that the NCG scheme provides increasing better performance as $\mathrm{tol}_{\mathrm{abs}}$ is taken smaller and N is taken larger. Further experiments show lack of robustness of the CNMS scheme when γ is sufficiently small. That is, smaller time step-size δt is required for convergence with the consequence of slowing down of the iteration or else the algorithm may diverge.

The remaining part of this section is dedicated to the improvement of the NCG approach by cascadic acceleration and to the investigation of the tracking ability of our optimal control formulation.

In Table 5, the performance of the NCG scheme and its accelerated version, C-NCG, are compared. We see a dramatic improvement with the C-NCG version especially for moderate values of γ. Taking smaller γ the optimal control problem becomes more ill-conditioned (stiff) and more computational effort is required for convergence.

We now discuss the effect of different choices of values of the optimization parameters using the results reported in Table 6. As required we see that smaller values of $|\psi(T) - \psi_d|_{\mathbb{C}^3}$ are attained for smaller γ. We remark that $\gamma = 10^{-7}$ is quite small and that makes the problem quite stiff and ill-conditioned. The NCG algorithm appears to be robust with respect to changes of γ. We can also see the effect of the regularization parameter μ. As μ increases, $|\psi(T) - \psi_d|_{\mathbb{C}^3}$ increases, demonstrating that the additional smoothness of the control function (slightly) reduces the capability of tracking. Apparently, larger μ makes the problem behaving better, resulting in a smaller number of iterations. Concerning the parameter $\alpha_3 = \alpha$ we obtain better tracking for non-zero α. This is expected since we have $|\psi_d|_{\mathbb{C}^3} = 1$, whereas $|\psi(T)|_{\mathbb{C}^3} < 1$ whenever $\gamma_0 > 0$, because of dissipation. By taking $\alpha > 0$ dissipation is reduced and, therefore, better $|\psi(T) - \psi_d|_{\mathbb{C}^3}$ is possibly achieved (Fig. 3).

6. Conclusions

Optimal control of finite-level quantum systems was investigated and iterative solution procedures were discussed.

It was shown that a cascadic non-linear conjugate gradient (C-NCG) approach provides a very efficient solution procedure that may outperform fast monotonic schemes. Moreover, it appeared that the C-NCG approach provides a robust iteration allowing to solve for very small values of the optimization parameters. This computational performance was exploited to investigate the tracking properties of our optimal control formulation. Results of numerical experiments were reported to demonstrate the efficiency and robustness of the proposed approach.

References

[1] R.A. Adams, Sobolev Spaces, Academic Press, New York, 1975.
[2] L. Baudouin, J. Salomon, Constructive solution of a bilinear quantum control problem, C.R. Math. Acad. Sci. Paris 342 (2) (2006) 119–124.
[3] K. Bergmann, H. Theuer, B.W. Shore, Coherent population transfer among quantum states of atoms and molecules, Rev. Modern Phys. 70 (1998) 1003.
[4] F.A. Bornemann, P. Deuflhard, The cascadic multigrid method for elliptic problems, Numer. Math. 75 (1996) 135–152.
[5] A. Borzì, G. Stadler, U. Hohenester, Optimal quantum control in nanostructures: theory and application to a generic three-level system, Phys. Rev. A 66 (2002) 053811.
[6] U. Boscain, G. Charlot, J.-P. Gauthier, S. Guerin, H.-R. Jauslin, Optimal control in laser-induced population transfer for two- and three-level quantum systems, J. Math. Phys. 43 (2002) 2107–2132.
[7] D. Bouwmeester, A. Ekert, A. Zeilinger (Eds)., The Physics of Quantum Information, Springer, Berlin, 2000.
[8] A. Brandt, Multigrid techniques: 1984 guide with applications to fluid dynamic, GMD-Studien no. 85, St. Augustin, Germany, 1984.
[9] P.W. Brumer, M. Shapiro, Principles of the Quantum Control of Molecular Processes, Wiley-VCH, Berlin, 2003.
[10] A.G. Butkovskiy, Yu.I. Samoilenko, Control of Quantum-Mechanical Processes and Systems, Kluwer Academic Publishers, Dordrecht, 1990.
[11] Y.H. Dai, Y. Yuan, A nonlinear conjugate gradient with a strong global convergence property, SIAM J. Optim. 10 (1999) 177–182.
[12] D.L. Elliott, Bilinear systems, in: J.G. Webster (Ed.), Wiley Encyclopedia of Electrical and Electronics Engineering, Wiley, New York, 1999.
[13] H.O. Fattorini, Infinite Dimensional Optimisation and Control Theory, Cambridge University Press, Cambridge, 1999.
[14] J.C. Gilbert, J. Nocedal, Global convergence properties of conjugate gradient methods for optimization, SIAM J. Optim. 2 (1992) 21–42.
[15] U. Hohenester, Optical properties of semiconductor nanostructures: decoherence versus quantum control, in: M. Rieth, W. Schommers (Eds.), Handbook of Theoretical and Computational Nanotechnology, American Scientific Publishers, 2005, to appear.
[16] U. Hohenester, F. Troiani, E. Molinari, Optically induced coherence effects in 'artificial atoms and molecules', in: L.C. Andreani, G. Benedek, E. Molinari (Eds.), Radiation-Matter Interaction in Confined Systems, Soc. Italiana di Fisica, Bologna, 2002.
[17] V.F. Krotov, Optimization methods of control with minimax criteria. I, Automat. Remote Control 34 (1973) 1863–1873.
[18] V.F. Krotov, Optimization methods in control processes with minimax criteria. II, Automat. Remote Control 35 (1974) 1–10.
[19] V.F. Krotov, Optimization methods of control with minimax criteria. III, Automat. Remote Control 35 (1974) 345–353.
[20] K. Kunisch, K. Ito, Optimal bilinear control of an abstract Schrödinger equation, SIAM J. Control Optim. 46 (2005) 274–287.

[21] J.L. Lions, Optimal Control of Systems Governed by Partial Differential Equations, Springer, Berlin, 1971.
[22] Y. Maday, J. Salomon, G. Turinici, Monotonic time-discretized schemes in quantum control, Numer. Math. 103 (2006) 323–338.
[23] Y. Maday, G. Turinici, New formulations of monotonically convergent quantum control algorithms, J. Chem. Phys. 118 (2003) 8191–8196.
[24] H. Maurer, J. Zowe, First and second order necessary and sufficient optimality conditions for infinite-dimensional programming problems, Math. Programming 16 (1979) 98–110.
[25] R.M. Mohler, Bilinear Control Processes, Academic Press, New York, 1973.
[26] J. Nocedal, S.J. Wright, Numerical Optimization, Springer, New York, 1999.
[27] Y. Ohtsuki, G. Turinici, H. Rabitz, Generalized monotonically convergent algorithms for solving quantum optimal control problems, J. Chem. Phys. 120 (2004) 5509–5517.
[28] Y. Ohtsuki, W. Zhu, H. Rabitz, Monotonically convergent algorithm for quantum optimal control with dissipation, J. Chem. Phys. 110 (1999) 9825–9832.
[29] A.P. Peirce, M.A. Dahleh, H. Rabitz, Optimal control of quantum-mechanical systems: existence, numerical approximation, and applications, Phys. Rev. A 37 (1988) 4950–4964.
[30] W. Pötz, W.A. Schroeder (Eds.), Coherent Control in Atoms, Molecules, and Semiconductors, Kluwer Academic Publishers, Dordrecht, 1999.
[31] H. Rabitz, M. Hsieh, C. Rosenthal, Quantum optimally controlled transition landscapes, Science 303 (2004) 998.
[32] H. Rabitz, G. Turinici, E. Brown, Control of molecular motion: concepts, procedures, and future prospects, in: P. Ciarlet and J. Lions (Eds.), Handbook of Numerical Analysis, vol. X, Elsevier, Amsterdam, 2003.
[33] J. Salomon, Convergence of the time discretized monotonic schemes, Mathematical Modelling and Numerical Analysis (M2AN), to appear.
[34] J. Salomon, G. Turinici, On the relationship between the local tracking procedures and monotonic schemes in quantum optimal control, J. Chem. Phys. 124 (2006) 074102.
[35] M.O. Scully, M.S. Zubairy, Quantum Optics, Cambridge University Press, Cambridge, 1997.
[36] V.V. Shaidurov, Some estimates of the rate of convergence of the cascadic conjugate gradient method, Comput. Math. Appl. 31 (1996) 161–171.
[37] D.F. Shanno, Conjugate gradient methods with inexact searches, Math. Oper. Res. 3 (1978) 244–256.
[38] D. Tannor, V. Kazakov, V. Orlov, Control of photochemical branching: novel procedures for finding optimal pulses and global upper bounds, in: J. Broeckhove, L. Lathouwers (Eds.), Time Dependent Quantum Molecular Dynamics, Plenum, New York, 1992, pp. 347–360.
[39] S.H. Tersigni, P. Gaspard, 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) 1670–1680.
[40] G. Turinici, Monotonically convergent algorithms for bounded quantum controls, in: Proceedings of the LHMNLC03 IFAC conference, Sevilla, 2003, pp. 263–266.
[41] W. Zhu, H. Rabitz, A rapid monotonically convergent iteration algorithm for quantum optimal control over the expectation value of a positive definite operator, J. Chem. Phys. 109 (1998) 385–391.