Monotonically convergent optimization in quantum control using Krotov's method

Daniel M. Reich, Mamadou Ndong, and Christiane P. Koch

Citation: The Journal of Chemical Physics **136**, 104103 (2012); doi: 10.1063/1.3691827

View online: http://dx.doi.org/10.1063/1.3691827

View Table of Contents: http://scitation.aip.org/content/aip/journal/jcp/136/10?ver=pdfcov

Published by the AIP Publishing

Monotonically convergent optimization in quantum control using Krotov's method

Daniel M. Reich,[1,2] Mamadou Ndong,[1,a)] and Christiane P. Koch[1,2,b)]
[1]*Institut für Theoretische Physik, Freie Universität Berlin, Arnimallee 14, 14195 Berlin, Germany*
[2]*Theoretische Physik, Universität Kassel, Heinrich-Plett-Str. 40, 34132 Kassel, Germany*

(Received 19 January 2012; accepted 14 February 2012; published online 9 March 2012)

The non-linear optimization method developed by A. Konnov and V. Krotov [Autom. Remote Cont. (Engl. Transl.) **60**, 1427 (1999)] has been used previously to extend the capabilities of optimal control theory from the linear to the non-linear Schrödinger equation [S. E. Sklarz and D. J. Tannor, Phys. Rev. A **66**, 053619 (2002)]. Here we show that based on the Konnov-Krotov method, monotonically convergent algorithms are obtained for a large class of quantum control problems. It includes, in addition to nonlinear equations of motion, control problems that are characterized by non-unitary time evolution, nonlinear dependencies of the Hamiltonian on the control, time-dependent targets, and optimization functionals that depend to higher than second order on the time-evolving states. We furthermore show that the nonlinear (second order) contribution can be estimated either analytically or numerically, yielding readily applicable optimization algorithms. We demonstrate monotonic convergence for an optimization functional that is an eighth-degree polynomial in the states. For the "standard" quantum control problem of a convex final-time functional, linear equations of motion and linear dependency of the Hamiltonian on the field, the second-order contribution is not required for monotonic convergence but can be used to speed up convergence. We demonstrate this by comparing the performance of first- and second-order algorithms for two examples. © *2012 American Institute of Physics*. [http://dx.doi.org/10.1063/1.3691827]

I. INTRODUCTION

Quantum control of light and matter uses external fields to engineer constructive and destructive interferences to steer a physical process into a desired direction.[1,2] The idea was pioneered in the 1980s (Refs. 3–6) and gained widespread attention with the advent of femtosecond lasers and pulse shaping techniques.[7] It was realized at about the same time that constructive and destructive interferences need not be devised by hand but can rather be obtained employing concepts from engineering such as feedback and optimization.[7–11] This has significantly broadened the range of quantum control problems that can be tackled. In particular, optimal control theory (OCT) has become a popular tool employed in areas as different as vibrational dynamics of complex molecules,[12,13] quantum dots and rings,[14,15] ultracold gases,[16–19] multi-photon excitations,[20,21] nuclear magnetic resonance,[22,23] and quantum information.[24–26]

All these applications are based on quantum dynamics, but differ in their (i) equation of motion, (ii) dependence of the Hamiltonian on the control, and (iii) target functional. (i) In standard applications of OCT to quantum control, the equation of motion is the linear Schrödinger equation, e.g., Refs. 8, 17, 24, and 25. OCT can also be applied to examples with a nonlinear equation of motion[13,16,19] which is obtained as an effective description in the framework of the mean-field approximation. (ii) The dependence of the Hamiltonian on the control is linear in most applications of OCT from atomic, molecular, and chemical physics, reflecting a laser field driving optically allowed transitions. This is changed to a nonlinear dependence if multi-photon excitations[20,21] or a parametrization of the control field[19] is considered. (iii) Linear and bilinear target functionals have commonly been used in application of OCT to quantum control to date.[27] However, target functionals that are higher order polynomials in the states of the system may be encountered in quantum information applications.[28] The type of dynamics and functionals translates into different requirements that must be met by the optimization algorithm. In particular for nonlinear equations of motion, nonlinear dependencies of the Hamiltonian on the control, time-dependent targets where the target operator is non-semi-definite and target functionals that are higher order polynomials in the states, it is not straightforward to construct *monotonically* convergent algorithms. This is the problem that we address here.

We utilize the nonlinear optimization algorithm by Konnov and Krotov[29] which had been translated to the language of quantum mechanics and first employed for solving a quantum control problem by Sklarz and Tannor.[16] Specifically, Sklarz and Tannor realized that a generalized form of the optimization functional yielding modified adjoint states is required in order to apply OCT to a nonlinear Schrödinger equation.[16] We show that the work by Sklarz and Tannor applies also to quantum control problems that are characterized by non-unitary time evolution, time-dependent targets, nonlinear dependencies of the Hamiltonian on the control and target functionals that depend to higher than second order

a)Present address: Laboratoire de Chimie Physique d'Orsay, Université de Paris Sud, 91405 Orsay, France.
b)Electronic mail: christiane.koch@uni-kassel.de.

on the time-evolving states. Translating the original proof by Konnov and Krotov[29] to quantum control, we furthermore show that the parameters of the optimization algorithm can be estimated analytically or numerically. We can thus give a ready-to-use prescription of the algorithm where the parameters are determined by the physics of the problem. For the case of a quadratic functional and linear equations of motion, a first-order construction of Krotov's method is sufficient to guarantee monotonic convergence.[16,27] We show that in this case, a second-order contribution may still be used to speed up convergence.

The paper is organized as follows. The optimization algorithm and the estimate of the algorithm's parameters are presented in Sec. II with all mathematical details of the derivation found in the appendices. Sections III and IV are devoted to applications of the optimization algorithm. We demonstrate monotonic convergence for one example that requires a second-order construction in Sec. III, and we present two examples in Sec. IV for which the second-order construction is not required but may be used to speed up convergence. Section V concludes.

II. OPTIMIZATION ALGORITHM

A. Control problem

The control problem is characterized by stating the control target and possible additional costs in functional form,

$$J = J_T\left[\{\varphi_k(T)\}\right] + \int_0^T J_t\left[\{\varphi_k(t)\}, \epsilon(t)\right]\, dt\,, \tag{1}$$

where we have separated final-time T and intermediate-time t "costs." $\{\varphi_k(t)\}$ denotes a set of complex state vectors[46] and $\epsilon(t)$ is the external field or control that will be optimized. The final-time cost is typically specified in terms of some desired unitary operator $\hat{\mathbf{O}}$, for example,[27]

$$\begin{aligned} J_T[\{\varphi_k(T)\}] &= -\frac{\lambda_0}{N^2}\left|\,\mathrm{Tr}\{\hat{\mathbf{O}}^\dagger\hat{\mathbf{P}}_\mathrm{N}\hat{\mathbf{U}}(\mathrm{T},0;\epsilon)\hat{\mathbf{P}}_\mathrm{N}\}\right|^2 \\ &= -\frac{\lambda_0}{N^2}\sum_{k,k'=1}^{N}\langle\varphi_k(t=0)|\hat{\mathbf{O}}^\dagger\hat{\mathbf{U}}(T,0;\epsilon)|\varphi_k(t=0)\rangle \\ &\quad\times\langle\varphi_{k'}(t=0)|\hat{\mathbf{U}}(T,0;\epsilon)^\dagger\hat{\mathbf{O}}|\varphi_{k'}(t=0)\rangle. \end{aligned} \tag{2}$$

Here, $\hat{\mathbf{U}}(T,0;\epsilon)$ denotes the time evolution operator from the initial time $t=0$ to the final time T under the action of the field ϵ, and λ_0 is a weight. The dimension of the subspace of the total Hilbert space, $\mathcal{H}$, on which the target operator $\hat{\mathbf{O}}$ acts, $\mathcal{H}_O$, is denoted by N, and $\hat{\mathbf{P}}_N$ is the projector onto $\mathcal{H}_O$. $\{|\varphi_k(t=0)\rangle\} \equiv \{|k\rangle\}$ is a suitable orthonormal basis spanning this subspace. A single state-to-state transition is obtained by taking $\hat{\mathbf{O}}$ to be the projector onto the target state, i.e., $N=1$. In order to optimize one-qubit or two-qubit quantum gates, $N=2$ and $N=4$, respectively. The functional is normalized by $1/N^2$ such that the optimum corresponds to the weight $J_T = -\lambda_0$.

The functional of Eq. (2) is quadratic in the states at final time $\{\varphi_k(T)\}$. A linear dependence is obtained by taking the real part instead of the square modulus of the trace and yields a phase-sensitive functional.[27] Generally, expressing the functionals in terms of expectation values yields at most a quadratic dependence on the states. A functional that is a higher order polynomial in the states is obtained in the context of quantum information when optimizing for a certain degree of entanglement rather than a specific unitary transformation,[28] see Sec. III below.

The intermediate-time cost J_t,

$$\begin{aligned} J_t\left[\{\varphi_k(t)\},\epsilon\right] &= g\left[\{|\varphi_k(t)\rangle\},\epsilon(t),t\right] \\ &= g_a\left[\epsilon(t),t\right] + g_b\left[\{\varphi_k(t)\},t\right]\,, \end{aligned} \tag{3}$$

is typically used to minimize the field intensity and to switch the field smoothly on and off,

$$g_a\left[\epsilon(t)\right] = \frac{\lambda_a}{S(t)}\left[\epsilon(t)-\epsilon_\mathrm{ref}(t)\right]^2\,, \tag{4}$$

where $\epsilon_\mathrm{ref}(t)$ denotes some reference field, $S(t)$ is a positive (shape) function, and λ_a a weight. J_t can also be used to formulate time-dependent targets[30–32] or constraints that depend on the state of the system at intermediate times[33] such as

$$g_b\left[\{\varphi_k(t)\}\right] = \frac{\lambda_b}{TN}\sum_{k=1}^{N}\langle\varphi_k(t)|\hat{\mathbf{D}}(t)|\varphi_k(t)\rangle\,, \tag{5}$$

where the dependence on the states again is quadratic. While complicated dependencies of $g_a[\epsilon]$ and $g_b[\{\varphi_k\}]$ are conceivable, we require $g_a[\epsilon]$ and $g_b[\{\varphi_k\}]$ to be additive, cf. Eq. (3). This assumption is typically justified since costs or penalties involving the field are usually not related to costs concerning the dynamics of the system.

The time evolution operator required to evaluate the functional, Eq. (2), can be obtained by solving the equation of motion for each of the basis states,

$$\begin{aligned} \frac{d}{dt}|\varphi_k(t)\rangle &= -\frac{i}{\hbar}\hat{\mathbf{H}}[\varphi_k,\epsilon]|\varphi_k(t)\rangle\,, \\ &= |f_k(\varphi_k,\epsilon)\rangle, \quad k=1,\ldots,N. \end{aligned} \tag{6}$$

An explicit dependency of the Hamiltonian on the state, $\hat{\mathbf{H}}[\varphi_k]$, will occur for nonlinear Schrödinger equations such as the Gross-Pitaevski equation or Hartree-Fock-like equations where the Hamiltonian is second order in the states.[13,16,19,34] The dependency of the Hamiltonian on the field can be linear, corresponding to one-photon dipole coupling, or higher order for non-resonant multi-photon transitions. Equation (6) can be extended to account for non-unitary time evolution by considering the density operator to be a vector in Liouville space and replacing the Hamiltonian by the Liouvillian.[35]

B. Optimization equations

The essence of Krotov's method[16,29] consists in disentangling the interdependency of the states and the control by constructing an auxiliary functional $L[\{\varphi_k\},\epsilon,\Phi]$ that depends on the states, the control, and an arbitrary scalar potential Φ such that for any Φ, $L[\{\varphi_k\},\epsilon,\Phi]$, and $J[\{\varphi_k\},\epsilon]$ are identical and minimization of $L[\{\varphi_k\},\epsilon,\Phi]$ is completely equivalent to minimization of $J[\{\varphi_k\},\epsilon]$. This is achieved by adding a

vanishing quantity,

$$L[\{\varphi_k\},\epsilon,\Phi] = G(\{\varphi_k(T)\}) - \Phi(\{\varphi_k(0)\},0) - \int_0^T R(\{\varphi_k(t)\},\epsilon(t),t)dt \tag{7}$$

with

$$\begin{aligned} G[\{\varphi_k(T)\}] &= J_T[\{\varphi_k(T)\}] + \Phi(\{\varphi_k(T)\},T), \\ R[\{\varphi_k(t)\},\epsilon(t),t] &= -(g_a[\epsilon(t),t] + g_b[\{\varphi_k(t)\},t]) \\ &\quad + \frac{\partial\Phi}{\partial t} + \sum_{k=1}^{N} \nabla_{|\varphi_k\rangle}\Phi|f_k[\varphi_k,\epsilon,t]\rangle \\ &\quad + \sum_{k=1}^{N} \langle f_k[\varphi_k,\epsilon,t]|\nabla_{\langle\varphi_k|}\Phi. \end{aligned}$$

It is sufficient to expand the functional $\Phi(\{|\varphi_k\rangle\},t)$ up to second order in the states to ensure the necessary monotonicity conditions for arbitrary change of the state vectors,

$$\Phi(\{\varphi_k\},t) = \sum_k [\langle\chi_k(t)|\varphi_k(t)\rangle + \langle\varphi_k(t)|\chi_k(t)\rangle] + \frac{1}{2}\sum_{kl}\langle\Delta\varphi_k(t)|\hat{\boldsymbol{\sigma}}(t)|\Delta\varphi_l(t)\rangle, \tag{8}$$

with first-order expansion coefficients $\chi_k(t)$ and $\Delta\varphi_k(t)$ the change in the state $\varphi_k(t)$, given by $\Delta\varphi_k(t) = \varphi_k^{(i+1)}(t) - \varphi_k^{(i)}(t)$ where the superscripts (i) and $(i+1)$ denote propagation under the old and new fields, $\epsilon^{(i)}(t)$ and $\epsilon^{(i+1)}(t)$, respectively. Due to the freedom of choice in Φ, the operator $\hat{\boldsymbol{\sigma}}(t)$ in the second order term can be chosen to ensure the extremum condition with respect to the state changes. If J is to be minimized, $\hat{\boldsymbol{\sigma}}(t)$ will be chosen to *maximize* it with respect to the change in the states. Any change in the field leads then to monotonic decrease of J.[16,29]

Requiring the first-order derivatives of L with respect to $|\varphi_k(t)\rangle$, $\langle\varphi_k(t)|$, and $\epsilon(t)$ to vanish, yields a set of coupled control equations for the first-order expansion coefficients $\chi_k(t)$ and the field. Since all involved functionals are real, it is sufficient to state the equation for the kets, the bra equations being given by the adjoints,

$$\begin{aligned} \frac{d}{dt}|\chi_k^{(i)}(t)\rangle &= -\frac{i}{\hbar}\hat{\mathbf{H}}^\dagger[\varphi_k^{(i)},\epsilon^{(i)}]|\chi_k^{(i)}\rangle \\ &\quad -\frac{i}{\hbar}\Bigg[\sum_l \langle\varphi_l^{(i)}|\nabla_{|\varphi_k\rangle}\hat{\mathbf{H}}^\dagger[\varphi_k^{(i)},\epsilon^{(i)}]|\chi_l^{(i)}\rangle \\ &\quad -\sum_l \langle\chi_l^{(i)}|\nabla_{\langle\varphi_k|}\hat{\mathbf{H}}[\varphi_k^{(i)},\epsilon^{(i)}]|\varphi_l^{(i)}\rangle\Bigg] \\ &\quad +\nabla_{\langle\varphi_k|}g_b|_{|\varphi_k^{(i)}(t)\rangle}, \end{aligned} \tag{9a}$$

$$|\chi_k^{(i)}(T)\rangle = -\nabla_{\langle\varphi_k|}J_T|_{|\varphi_k^{(i)}(T)\rangle}, \quad k=1,\ldots,N. \tag{9b}$$

The "initial" condition at the final time T is given in terms of the gradient with respect to the states of the final-time cost, J_T. Note that all gradients in Eqs. (9) are evaluated with the states $\varphi_k^{(i)}$ that are forward propagated under the old field, $\epsilon^{(i)}$ as indicated by the superscript (i). The prescription for the new field is obtained by evaluating the derivative of the constraints with respect to the field,

$$\begin{aligned} \frac{\partial g}{\partial\epsilon}\Big|_{\epsilon^{(i+1)},|\varphi^{(i+1)}\rangle} &= 2\Im\mathfrak{m}\Bigg[\sum_{k=1}^{N}\langle\chi_k^{(i)}(t)|\frac{\partial\hat{\mathbf{H}}}{\partial\epsilon}\Big|_{\epsilon^{(i+1)},\varphi^{(i+1)}}|\varphi_k^{(i+1)}(t)\rangle \\ &\quad +\frac{1}{2}\sigma(t)\sum_{k=1}^{N}\langle\Delta\varphi_k(t)|\frac{\partial\hat{\mathbf{H}}}{\partial\epsilon}|_{\epsilon^{(i+1)},\varphi^{(i+1)}}|\varphi_k^{(i+1)}(t)\rangle\Bigg]. \end{aligned} \tag{10}$$

It involves backward propagation of the adjoint states under the old field, $\chi_k^{(i)}(t)$, and forward propagation of the states, $\varphi_k^{(i+1)}(t)$, under the new field,

$$\frac{d}{dt}|\varphi_k^{(i+1)}(t)\rangle = -\frac{i}{\hbar}\hat{\mathbf{H}}[\varphi_k^{(i+1)},\epsilon^{(i+1)}]|\varphi_k^{(i+1)}(t)\rangle \tag{11a}$$

$$|\varphi_k^{(i+1)}(0)\rangle = |k\rangle, \quad k=1,\ldots,N. \tag{11b}$$

Equations (9)–(11) need to be solved simultaneously. The iteration is started by propagating Eqs. (11) under some guess field, $\epsilon^{(0)}(t)$, to obtain $\varphi_k^{(0)}(T)$ and evaluate Eq. (9b). The equation for the backward propagation, Eq. (9a), becomes an inhomogeneous Schrödinger equation for nonlinear equations of motion, cf. the terms in square brackets, or if the intermediate-time cost, g_b, depends on the states ($\nabla_{\langle\varphi_k|}g_b \neq 0$). If the time-dependent cost over the field takes the form of Eq. (4), the equation for the new field reads as

$$\begin{aligned} &\epsilon^{(i+1)}(t) \\ &= \epsilon_{\mathrm{ref}}(t) + \frac{S(t)}{\lambda_a}\Im\mathfrak{m}\Bigg\{\sum_{k=1}^{N}\langle\chi_k^{(i)}(t)|\frac{\partial\hat{\mathbf{H}}}{\partial\epsilon}\Big|_{\epsilon^{(i+1)},\varphi^{(i+1)}}|\varphi_k^{(i+1)}(t)\rangle \\ &\quad +\frac{1}{2}\sigma(t)\sum_{k=1}^{N}\langle\Delta\varphi_k(t)|\frac{\partial\hat{\mathbf{H}}}{\partial\epsilon}\Big|_{\epsilon^{(i+1)},\varphi^{(i+1)}}|\varphi_k^{(i+1)}(t)\rangle\Bigg\}. \end{aligned} \tag{12}$$

Moreover, for dipole transitions the Hamiltonian is given by $\hat{\mathbf{H}} = \hat{\mathbf{H}}_0 + \hat{\mu}\epsilon(t)$; and hence $\partial\hat{\mathbf{H}}/\partial\epsilon = \hat{\boldsymbol{\mu}}$. We thus recover the familiar prescription for the change in field obtained for a first-order construction of the algorithm[27] plus an additional second-order contribution, given in terms of the change in the states, $\Delta\varphi_k^{(i+1)}(t)$, with "weight" $\sigma(t)$.

A choice of $\sigma(t)$ that guarantees a maximum of L with respect to the states (i.e., a positive second derivative of R and negative second derivative of G) is given by[29]

$$\sigma(t) = e^{\bar{B}(T-t)}\left(\frac{\bar{C}}{\bar{B}} - \bar{A}\right) - \frac{\bar{C}}{\bar{B}} \quad \text{for} \quad \bar{B}\neq 0, \tag{13a}$$

$$\sigma(t) = \bar{C}(T-t) - \bar{A} \quad \text{for} \quad \bar{B} = 0. \tag{13b}$$

The physics of the problem, i.e., the dependency of the functional on the states, the dependency of the Hamiltonian on the control, (non-)linearity of the equation of motion

governing, and unitary or non-unitary of the time evolution, determine the parameters $\bar{A}$, $\bar{B}$, and $\bar{C}$,

$$\bar{A} = \max\left(\varepsilon_A, 2A + \varepsilon_A\right), \tag{14a}$$

$$\bar{B} = 2B + \varepsilon_B, \tag{14b}$$

$$\bar{C} = \min\left(-\varepsilon_C, 2C - \varepsilon_C\right), \tag{14c}$$

where the ε_i are non-negative numbers that can be used to enforce strict inequality.[47] It should be emphasized, that only in cases where A, B, and C all turn out to be zero, the linear version of Krotov's method[27] is sufficient to guarantee monotonic convergence.

C. Parameters of the second-order contribution

For quantum control problems the parameters A, B, and C can either be estimated analytically or approximated numerically by considering $\Delta G = G(\{\varphi_k^{(i+1)}(T)\}) - G(\{\varphi_k^{(i)}(T)\})$ and $\Delta R = R(\{\varphi_k^{(i+1)}(t)\}, \epsilon^{(i)}(t), t) - R(\{\varphi_k^{(i)}(t)\}, \epsilon^{(i)}(t), t)$ and guaranteeing their negativity and positivity, respectively. The analytical estimate is based on a worst case scenario and strictly guarantees monotonic convergence. The worst case scenario may, however, not always be required and a more efficient while less fail-proof approach is given by numerical approximation of A, B, and C.

The analytical estimate of A is obtained by requiring $\Delta G \leq 0$. This is guaranteed if

$$A = \sup_{\{\Delta\varphi_k\}} \frac{\sum_{k=1}^N \left[\langle \chi_k(T) \,|\, \Delta\varphi_k(T)\rangle + \langle \Delta\varphi_k(T) \,|\, \chi_k(T)\rangle\right] + J_T\left(\left\{\varphi_k^{(i)}(T) + \Delta\varphi_k(T)\right\}\right) - J_T\left(\left\{\varphi_k^{(i)}(T)\right\}\right)}{\sum_{k=1}^N \left[\langle \Delta\varphi_k(T) \,|\, \Delta\varphi_k(T)\rangle\right]}, \tag{15}$$

where the supremum needs to be taken over all sets of possible state change vectors $\{\Delta\varphi_k(T)\}$ with norm $\sum_{k=1}^N \langle \Delta\varphi_k(T) \,|\, \Delta\varphi_k(T)\rangle$ between zero and $2N$. Konnov and Krotov[29] proved that, under certain conditions which are almost trivially fulfilled for quantum systems, quantities like the one on the right-hand side of Eq. (15) and similar ones obtained for B and C exist and are well-defined, see also Appendix A. We discuss in Appendix B how Konnov's and Krotov's proof simplifies for quantum systems such that the supremum in Eq. (15) can be estimated. Specifically, we show in Appendix C that the argument of the supremum in Eq. (15) can be rewritten in terms of a Taylor series of J_T that starts at the second order. The series can be estimated by its Lagrange remainder,

$$A \leq \frac{1}{2} \sup_{\{\Delta\varphi_k\};|\alpha|=2} \partial^\alpha J_T(\{\Delta\varphi_k(T)\}), \tag{16}$$

i.e., A is given by the supremum over the second derivatives of the final-time functional, J_T, with respect to the states, $\varphi_k(T)$. The multi-index α and the derivative ∂^α are defined in Eqs. (C2) and (C3), respectively. For functionals J_T that are linear or convex in $\varphi_k(T)$, i.e., those for which the second derivatives vanish or are always non-positive, $A \leq 0$. For simplicity one can then choose $A = 0$, $\varepsilon_A = 0$ such that $\bar{A} = 0$.

The analytical estimates of B and C are obtained by requiring $\Delta R \geq 0$, where

$$\begin{aligned}\Delta R(\{\Delta\varphi(t)\}, t) = \sum_{k=1}^N \left[\langle \Delta\varphi_k(t) \,|\, \Delta\varphi_k(t)\rangle\right] \Bigg[&\frac{1}{2}\dot{\sigma}(t) + \sigma(t) \frac{\sum_{k=1}^N \left[\langle \Delta\varphi_k(t) \,|\, \Delta f_k(t)\rangle + \langle \Delta f_k(t) \,|\, \Delta\varphi_k(t)\rangle\right]}{\sum_{k=1}^N \left[\langle \Delta\varphi_k(t) \,|\, \Delta\varphi_k(t)\rangle\right]} \\ &+ \frac{\sum_{k=1}^N \left[\langle \dot{\chi}_k(t) \,|\, \Delta\varphi_k(t)\rangle + \langle \Delta\varphi_k(t) \,|\, \dot{\chi}_k(t)\rangle + \langle \chi_k(t) \,|\, \Delta f_k(t)\rangle + \langle \Delta f_k(t) \,|\, \chi_k(t)\rangle\right] - \Delta g}{\sum_{k=1}^N \left[\langle \Delta\varphi_k(t) \,|\, \Delta\varphi_k(t)\rangle\right]} \Bigg].\end{aligned} \tag{17}$$

The second summand in the square brackets determines B and the third one C. Δf_k and Δg describe the change, due to changes in the states, in the equations of motion,

$$\begin{aligned}|\Delta f_k(\Delta\varphi_k, t)\rangle = &\left| f_k\left(\varphi_k^{(i)}(t) + \Delta\varphi_k(t), \epsilon^{(i)}(t), t\right)\right\rangle \\ &- \left| f_k\left(\varphi_k^{(i)}(t), \epsilon^{(i)}(t), t\right)\right\rangle,\end{aligned} \tag{18}$$

and in the constraint,

$$\begin{aligned}\Delta g(\{\Delta\varphi_k\}, t) = &\, g\left(\left\{\varphi_k^{(i)}(t) + \Delta\varphi_k(t)\right\}, \epsilon^{(i)}(t), t\right) \\ &- g\left(\left\{\varphi_k^{(i)}(t)\right\}, \epsilon^{(i)}(t), t\right).\end{aligned} \tag{19}$$

B is given by

$$B = \sup_{\{\Delta\varphi_k\};t\in[0,T]} \left| \frac{\sum_{k=1}^{N}[\langle \Delta\varphi_k(t)|\Delta f_k(t)\rangle + \langle \Delta f_k(t)|\Delta\varphi_k(t)\rangle]}{\sum_{k=1}^{N}[\langle \Delta\varphi_k(t)|\Delta\varphi_k(t)\rangle]} \right|, \tag{20}$$

and can be rewritten in terms of a supremum over the Taylor expansion of the Hamiltonian starting at first order and a supremum over the Hamiltonian, cf. Appendix C. Estimating the Taylor series by its Lagrange remainder, we obtain

$$B \le 2\sqrt{N} \sup_{\substack{\Delta\varphi_k;|\alpha|=1\\ t\in[0,T]}} |\partial^\alpha \hat{\mathbf{H}}(\Delta\varphi_k(t))| + 2 \sup_{\{\Delta\varphi_k\};t\in[0,T]} \left| \frac{\sum_{k=1}^{N} \Im\mathrm{m}\langle \Delta\varphi_k(t)|\hat{\mathbf{H}}(\Delta\varphi_k(t),\epsilon^{(i)},t)|\Delta\varphi_k(t)\rangle}{\sum_{k=1}^{N}[\langle \Delta\varphi_k(t)|\Delta\varphi_k(t)\rangle]} \right|, \tag{21}$$

i.e., B can be estimated by the supremum over the first derivative of the Hamiltonian with respect to the states $\varphi(t)$ and a term which can be interpreted as twice the maximum absolute value of the imaginary part of the Hamiltonian's eigenvalues. For unitary time evolution governed by the standard (linear) Schrödinger equation, $B = 0$ can easily be proven and in this case $\sigma(t)$ is a linear function of time. It is also possible to take $B = 0$ for non-unitary time evolution with linear equations of motion provided A and C can taken to be zero. For non-linear equations of motion, the supremum of the first-order derivatives of the Hamiltonian needs to be evaluated explicitly. C is found to be

$$C = \inf_{\{\Delta\varphi_k\};t\in[0,T]} \frac{\sum_{k=1}^{N}[\langle \dot{\chi}_k(t)\,|\,\Delta\varphi_k(t)\rangle + \langle \Delta\varphi_k(t)\,|\,\dot{\chi}_k(t)\rangle + \langle \chi_k(t)\,|\,\Delta f_k(t)\rangle + \langle \Delta f_k(t)\,|\,\chi_k(t)\rangle] - \Delta g}{\sum_{k=1}^{N}[\langle \Delta\varphi_k(t)\,|\,\Delta\varphi_k(t)\rangle]}. \tag{22}$$

As shown in Appendix C, it can be rewritten in terms of suprema of the Taylor series of the Hamiltonian and the constraint, g, starting at first and second order, respectively. Estimating the Taylor series by their Lagrange remainder, we obtain

$$-C \ge 2\sum_{k=1}^{N}\left[\sqrt{\langle \chi_k^{(i)}(t)\,|\,\chi_k^{(i)}(t)\rangle} \sup_{\substack{\Delta\varphi_k;t\in[0,T]\\ |\alpha|=1}} [\partial^\alpha \hat{\mathbf{H}}(\Delta\varphi_k,t)]\right] + \sup_{\substack{\{\Delta\varphi_k\};t\in[0,T]\\ |\alpha|=2}} [\partial^\alpha g(\{\Delta\varphi_k\},t)], \tag{23}$$

i.e., C is given by the supremum of the first-order derivatives of the Hamiltonian multiplied by the norm of the costates $\chi_k(t)$ and the supremum of the first-order derivatives of the constraint, g. For linear equations of motion and g_b zero or linear in $\varphi_k(t)$, we find $C = 0$. The case of a quadratic dependency of g on the states φ_k, cf. Eq. (5), can also easily be handled. The second term in the right-hand side of Eq. (23) can then be estimated by the eigenvalue of the operator $\hat{\mathbf{D}}(t)$ with largest magnitude. For example, if $\hat{\mathbf{D}}(t)$ is the projector onto some subspace, $\hat{\mathbf{D}}(t) = \hat{\mathbf{P}}$, then

$$C \le -\frac{\lambda_b}{NT} \max_{\mathrm{EV}}[\hat{\mathbf{P}}] = -\frac{\lambda_b}{NT}. \tag{24}$$

In order to derive a numerical approximation for the parameters A, B, and C, we assume a finite time grid, $\{t_j\}$, $j = 1, \ldots, n$, and define, based on Eqs. (15), (20), and (22), parameters $A^{(i+1)}$, $B_j^{(i+1)}$, and $C_j^{(i+1)}$,

$$A^{(i+1)}(\{\Delta\varphi_k\}) = \frac{\sum_{k=1}^{N}[\langle \chi_k(T)\,|\,\Delta\varphi_k(T)\rangle + \langle \Delta\varphi_k(T)\,|\,\chi_k(T)\rangle] + J_T\left(\left\{\varphi_k^{(i)}(T) + \Delta\varphi_k(T)\right\}\right) - J_T\left(\left\{\varphi_k^{(i)}(T)\right\}\right)}{\sum_{k=1}^{N}[\langle \Delta\varphi_k(T)\,|\,\Delta\varphi_k(T)\rangle]}, \tag{25}$$

and

$$B_j^{(i+1)}(\{\Delta\varphi_k\}) = \frac{\sum_{k=1}^{N}[\langle \Delta\varphi_k(t_j)\,|\,\Delta f_k(\Delta\varphi_k,t_j)\rangle + \langle \Delta f_k(\Delta\varphi_k,t_j)\,|\,\Delta\varphi_k(t_j)\rangle]}{\sum_{k=1}^{N}[\langle \Delta\varphi_k(t_j)\,|\,\Delta\varphi_k(t_j)\rangle]},$$

$$C_j^{(i+1)}(\{\Delta\varphi_k\}) = \frac{1}{\sum_{k=1}^{N}[\langle \Delta\varphi_k(t_j)\,|\,\Delta\varphi_k(t_j)\rangle]}\left[\sum_{k=1}^{N}[\langle \dot{\chi}_k(t_j)\,|\,\Delta\varphi_k(t_j)\rangle + \langle \Delta\varphi_k(t_j)\,|\,\dot{\chi}_k(t_j)\rangle\right.$$

$$\left. + \langle \chi_k(t_j)\,|\,\Delta f_k(\Delta\varphi_k,t_j)\rangle + \langle \Delta f_k(\Delta\varphi_k,t_j)\,|\,\chi_k(t_j)\rangle] - \Delta g(\{\Delta\varphi_k\},t_j)\right].$$

At iteration step $(i+1)$, a numerical estimate for the second-order parameter, $\sigma^{(i+1)}(t)$, is obtained by replacing A, B, and C in Eqs. (14) by $A^{(i+1)}$, Eq. (25), together with

$$B^{(i+1)} = \sup_j B_j^{(i+1)}, \tag{26a}$$

$$C^{(i+1)} = \inf_j C_j^{(i+1)}, \tag{26b}$$

and inserting the resulting numerical estimates of $\bar{A}$, $\bar{B}$, and $\bar{C}$ into Eqs. (13). Since the new states $\varphi_k^{(i+1)}(t) = \varphi_k^{(i)}(t) + \Delta\varphi(t)$ are not known, $A^{(i+1)}$, $B_j^{(i+1)}$, and $C_j^{(i+1)}$ need to be approximated, for example, by the values of $A^{(i)}$, $B_j^{(i)}$, and $C_j^{(i)}$ calculated in the previous iteration. In rare cases, this approximation might lead to loss of monotonic convergence. The iteration then needs to be repeated with the values of $A^{(i+1)}$, $B_j^{(i+1)}$, and $C_j^{(i+1)}$ that were obtained during the failed step. Numerical estimation of the second-order parameters enforces monotonic convergence with respect to a change in the states as gently as possible, making use of the optimization history to find a compromise between monotonicity and speed of the convergence.

D. Monotonic convergence for arbitrary dependence of the Hamiltonian on the control

In any iterative optimization, convergence with respect to the field can only be expected towards a local extremum. Here, the local extremum condition on J translates into $\frac{\partial^2 R}{\partial \epsilon^2} < 0$, or,

$$\frac{\partial^2 g_a}{\partial \epsilon^2}\bigg|_{\varphi_k^{(i+1)},\epsilon^{(i+1)}} > \sum_{k=1}^{N} \left\langle \chi_k^{(i)}(t) \,\middle|\, \frac{\partial^2 \hat{\mathbf{H}}}{\partial \epsilon^2}\bigg|_{\epsilon^{(i+1)},\varphi^{(i+1)}} \,\middle|\, \varphi_k^{(i+1)}(t) \right\rangle + \sum_{k=1}^{N} \left\langle \varphi_k^{(i+1)}(t) \,\middle|\, \frac{\partial^2 \hat{\mathbf{H}}^\dagger}{\partial \epsilon^2}\bigg|_{\epsilon^{(i+1)},\varphi^{(i+1)}} \,\middle|\, \chi_k^{(i)}(t) \right\rangle$$
$$+ \frac{1}{2}\sigma(t) \sum_{k=1}^{N} \left[\langle \Delta\varphi_k(t) \,|\, \frac{\partial^2 \hat{\mathbf{H}}}{\partial \epsilon^2}\bigg|_{\epsilon^{(i+1)},\varphi^{(i+1)}} \,\middle|\, \varphi_k^{(i+1)}(t) \right\rangle + \left\langle \varphi_k^{(i+1)}(t) \,\middle|\, \frac{\partial^2 \hat{\mathbf{H}}^\dagger}{\partial \epsilon^2}\bigg|_{\epsilon^{(i+1)},\varphi^{(i+1)}} \,|\, \Delta\varphi_k(t) \right\rangle \Bigg]. \tag{27}$$

For a linear dependence of the Hamiltonian on the control, $\frac{\partial^2 \hat{\mathbf{H}}}{\partial \epsilon^2} = \frac{\partial^2 \hat{\mathbf{H}}^\dagger}{\partial \epsilon^2} = 0$, and a maximum in R requires simply $\frac{\partial^2 g_a}{\partial \epsilon^2}|_{\epsilon^{(i+1)}} > 0$. This translates into the sign of the weight λ_a for the typical quadratic dependence of g_a on ϵ, cf. Eq. (4). Inserting the corresponding derivative of g_a, we obtain

$$\epsilon^{(i+1)}(t) = \tilde{\epsilon}(t) + \frac{S(t)}{\lambda_a} \Im\mathrm{m} \left\{ \sum_{k=1}^{N} \left\langle \chi_k^{(i)}(t) \middle| \frac{\partial \hat{\mathbf{H}}}{\partial \epsilon}\bigg|_{\epsilon^{(i+1)},\varphi^{(i+1)}} \middle| \varphi_k^{(i+1)}(t) \right\rangle + \frac{1}{2}\sigma(t) \sum_{k=1}^{N} \langle \Delta\varphi_k(t) \,|\, \frac{\partial \hat{\mathbf{H}}}{\partial \epsilon}\bigg|_{\epsilon^{(i+1)},\varphi^{(i+1)}} \,\middle|\, \varphi_k^{(i+1)}(t) \right\rangle \right\}. \tag{28}$$

For a nonlinear dependency of the Hamiltonian on the control, we define the change in the intermediate-time functional due to changes in the control,

$$\Delta_\epsilon(t) = R\left(\left\{\varphi_k^{(i+1)}(t)\right\}, \epsilon^{(i+1)}(t), t\right) - R\left(\left\{\varphi_k^{(i+1)}(t)\right\}, \epsilon^{(i)}(t), t\right). \tag{29}$$

The strict maximum condition for R becomes $\Delta_\epsilon(t) > 0\ \forall t$. We assume J_t, cf. Eq. (3), to be additive. Rewriting $\hat{\mathbf{H}}[\varphi_k^{(i+1)}, \epsilon^{(i+1)}, t] - \hat{\mathbf{H}}[\varphi_k^{(i+1)}, \epsilon^{(i)}, t]$ in terms of the Taylor expansion of the Hamiltonian with respect to the control, we find the zeroth order term to vanish. The first-order derivative can be rewritten using Eq. (10). The remaining terms correspond to the Taylor series starting at second order which can be estimated by its Lagrange remainder,

$$\tilde{M}_2^\epsilon(t) = \sup_{\varphi_k;\epsilon} \left| \frac{\partial^2}{\partial \epsilon^2} \hat{\mathbf{H}}(\varphi_k, \epsilon, t) \right|. \tag{30}$$

Employing furthermore $\sum_{k=1}^{N} \sqrt{\langle \Delta\varphi_k(t) \,|\, \Delta\varphi_k(t) \rangle} \le 2\sqrt{N}$, we obtain

$$\Delta_\epsilon(t) > g_a(\epsilon^{(i)}, t) - g_a(\epsilon^{(i+1)}, t) + (\epsilon^{(i+1)}(t) - \epsilon^{(i)}(t)) \frac{\partial g_a}{\partial \epsilon}\bigg|_{\epsilon^{(i+1)}} + \left\{ \sqrt{N}\tilde{M}_2^\epsilon(t) \sum_{k=1}^{N} \sqrt{\left\langle \chi_k^{(i)}(t) \,\middle|\, \chi_k^{(i)}(t) \right\rangle} + |\sigma(t)| N \tilde{M}_2^\epsilon(t) \right\} \times (\epsilon^{(i+1)}(t) - \epsilon^{(i)}(t))^2,$$

The Lagrange remainder can be evaluated by taking the second derivative of the Hamiltonian with respect to the field and estimating the norm of the resulting operator by its spectral radius or its eigenvalue with largest square modulus. Since it is difficult to proceed without a specific dependence of g_a on ϵ, we assume a quadratic dependence, cf. Eq. (4), and find

$$\Delta_\epsilon(t) > \left[\frac{\lambda_a}{S(t)} - \left\{ \frac{1}{2}\sqrt{N} \sum_{k=1}^{N} \left[\sqrt{\left\langle \chi_k^{(i)}(t) \,\middle|\, \chi_k^{(i)}(t) \right\rangle} \right] \tilde{M}_2^\epsilon(t) + N|\sigma(t)|\tilde{M}_2^\epsilon(t) \right\} \right] (\epsilon^{(i+1)}(t) - \epsilon^{(i)}(t))^2. \tag{31}$$

Monotonic convergence is ensured by adjusting the shape function $S(t)$ and the parameter λ_a such that

$$\frac{\lambda_a}{S(t)} > \left\{ \frac{1}{2}\sqrt{N}\sum_{k=1}^{N}\left[\sqrt{\left|\chi_k^{(i)}(t)\,\middle|\,\chi_k^{(i)}(t)\right|}\right]\tilde{M}_2^{\epsilon}(t) + N|\sigma(t)|\tilde{M}_2^{\epsilon}(t)\right\}. \tag{32}$$

Unlike the algorithm of Ref. 21, the numerical effort in our approach to ensure monotonic convergence is independent of the order of the non-linearity of the Hamiltonian's dependence on the control.

To summarize Sec. II, a second-order construction of the optimization algorithm yields an additional contribution to the equation for the new field. Compared to the linear variant of Krotov's method,[27] it simply requires additional storage of the states from the previous iteration to determine $\Delta\varphi_k(t)$ and calculation of the function $\sigma(t)$. The parameters A, B, and C determining $\sigma(t)$ turn out to be zero, i.e., the second-order contribution vanishes, for the "standard" quantum control problem with bilinear final-time cost, intermediate-time costs that are independent of the states and linear equations of motion. A second-order contribution can nevertheless be invoked to study its influence on convergence. This is investigated below in Sec. IV B. A second-order contribution is required to guarantee monotonic convergence for final-time costs that are a polynomial of higher than second order in the states,[28] demonstrated below in Sec. III, and for general intermediate-time costs that depend on the states, studied below in Sec. IV C. A second-order construction is also required for non-unitary time evolution and for non-linear equations of motion.[16,34]

III. APPLICATION I: HIGHER ORDER POLYNOMIAL COSTS

A functional that is an eighth order polynomial in the states arises in quantum information when optimizing for a local equivalence class of two-qubit gates, $[\hat{\mathbf{O}}]$, rather than a specific two-qubit gate, $\hat{\mathbf{O}}$.[28] The functional is based on the local invariants of two-qubit gates[36] which uniquely characterize a local equivalence class. The explicit, somewhat lengthy expression of J_T^{LI} is given in Ref. 28.

We employ J_T^{LI} to optimize for the local equivalence class of the B-gate,[37] given in the logical basis by

$$\hat{\mathbf{O}}_B = e^{\frac{i}{2}\left[\frac{\pi}{2}\hat{\boldsymbol{\sigma}}_x\otimes\hat{\boldsymbol{\sigma}}_x+\frac{\pi}{4}\hat{\boldsymbol{\sigma}}_y\otimes\hat{\boldsymbol{\sigma}}_y\right]} = \begin{pmatrix} \cos(\pi/8) & 0 & 0 & i\sin(\pi/8) \\ 0 & \sin(\pi/8) & i\cos(\pi/8) & 0 \\ 0 & i\cos(\pi/8) & \sin(\pi/8) & 0 \\ i\sin(\pi/8) & 0 & 0 & \cos(\pi/8) \end{pmatrix},$$

for an effective spin-spin system,

$$\hat{\mathbf{H}}_{eff} = \frac{\hbar\Omega(t)^2}{8}\sum_{i,j=0}^{3}\hat{\boldsymbol{\sigma}}_i a_{ij}(x_0)\hat{\boldsymbol{\sigma}}_j\,. \tag{33}$$

The effective Hamiltonian is derived, within second-order perturbation theory, for trapped polar molecules with $^2\Sigma_{1/2}$ electronic ground states, subject to near-resonant microwave driving that induces strong dipole-dipole coupling.[38] $\Omega(t)$ denotes the Rabi frequency that will be optimized, $\hat{\boldsymbol{\sigma}}_i$ are the 2×2 Pauli spin matrices, $i = 1, 2, 3 \equiv x, y, z$ with $\sigma_0 = \mathbb{1}_2$. The tensor a_{ij} depends on the distance x_0 between the molecules and on the polarization and detuning of the microwave field. We consider here CaCl molecules in an optical lattice with a lattice spacing of 300 nm, microwave radiation of about 9.13 GHz, polarizations $\alpha_\pm = 1/\sqrt{2}$, $\alpha_0 = 0$, and a detuning from the rotational transition of 1.2 kHz.

The higher than quadratic dependence of J_T^{LI} on the states leads to a non-zero analytical estimate of A, cf. Eq. (16). Specifically, one has to calculate all second derivatives of the eighth order polynomial and evaluate the supremum by considering an arbitrary change in the states up to $\|\Delta\varphi\| = \|\sum_k \Delta\varphi_k\| \le 2\sqrt{N}$, cf. Appendix B (Fig. 6). For our example, $N = 4$ and from the second derivatives, we obtain $\bar{A} = 45$ although in practice a value of $\bar{A} = 5$ was sufficient to preserve monotonic convergence. Figure 1 demonstrates that this choice of $\bar{A}$ (blue dashed line) indeed ensures monotonic convergence independent of other parameters of the algorithm such as the weight λ_a, cf. Eq. (4). Note that in this example, convergence of the final-time functional J_T and the complete functional J are equivalent due to our specific choice of g_a (Ref. 27) and $g_b = 0$. The latter yields $\bar{C} = 0$. Furthermore, $\bar{B} = 0$ since our equation of motion is linear in the states. The first-order algorithm (solid red line) fails completely for small λ_a and violates monotonic convergence for many iteration steps albeit finding an optimum eventually for intermediate λ_a. The weight λ_a determines the step size for changes in the field, cf. Eq. (12): small λ_a leads to large values of the field and thus a possibly more "erratic" optimization, while for large λ_a, optimization proceeds more cautiously, explaining that even the first-order construction is found to converge monotonically. This is, however, due to the simplicity of our control problem which is essentially one-dimensional since success is determined by the integrated field amplitude. For more complex Hamiltonians, optimization employing a first-order construction was found to always fail,[28] underlying the significance of the second-order construction. The analytical estimate of the second-order parameter takes all worst case scenarios into account. Therefore, values smaller than the analytical estimate might already be sufficient for monotonicity which is confirmed as shown in Fig. 1 by the blue dashed line representing $\bar{A} = 5$ and the black dotted line representing $\bar{A} = 1$ even though the analytical estimate is given by $\bar{A} = 45$. Such a less conservative estimate yields, moreover, significantly faster convergence for small λ_a, a fact that was also confirmed for more complex Hamiltonians.[28]

IV. APPLICATION II: BILINEAR COSTS

For bilinear costs, the supremum estimation of C yields zero. We consider the dynamics of our examples to be described by the standard Schrödinger equation such that $B = 0$ and $\bar{B}$ can also be chosen zero. A second-order construction of the optimization algorithm becomes necessary if bilinear intermediate-time costs are employed, $g_b \neq 0$, which

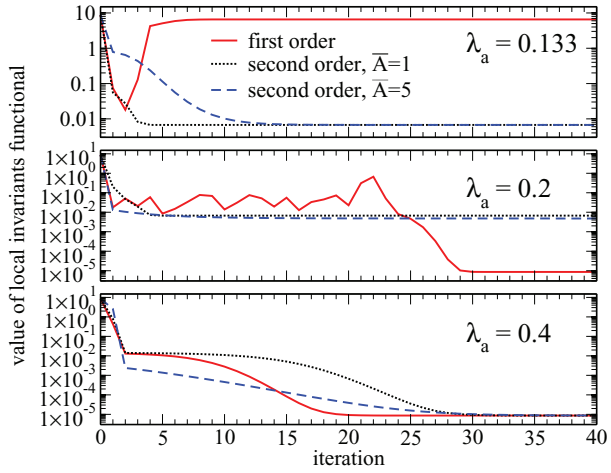

FIG. 1. Convergence of the local invariants functional that optimizes for a local equivalence class rather than a specific operator and is an eighth order polynomial in the states. The optimum corresponds to $J_T^{LI} = 0$.

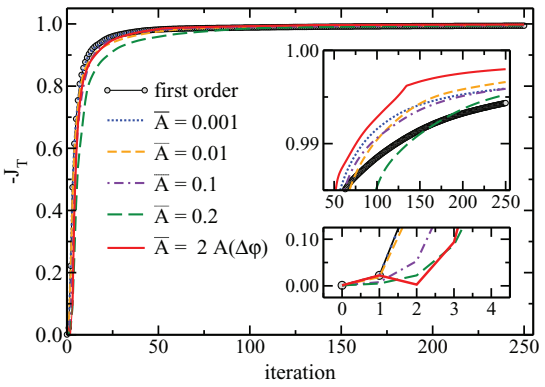

FIG. 2. Convergence of the first-order and second-order constructions of the optimization algorithm as measured by the final-time objective, J_T, for state-to-state transfer from vibrational level $v = 10$ to $v = 0$.

lead to non-zero C, cf. Eq. (23). If $g_b = 0$, a second-order construction is not required to ensure monotonicity but can be utilized to speed up convergence.

A. Model

A simplified model for the vibrations in a Rb_2 molecule that linearly interacts with a laser field takes three electronic states, $X^1\Sigma_g^+(5s+5s)$, $^1\Sigma_u^+(5s+5p)$, and $^1\Pi_g(5s+4d)$, into account,[33,39]

$$\hat{\mathbf{H}} = \sum_{i=1}^{3} \hat{\mathbf{H}}_i \otimes |e_i\rangle\langle e_i| + \hat{\mu}\,\epsilon(t)\cdot\left(|e_1\rangle\langle e_2| + |e_2\rangle\langle e_1| + |e_2\rangle\langle e_3| + |e_3\rangle\langle e_2|\right). \quad (34)$$

Here, $\hat{\mathbf{H}}_i$ denotes the vibrational Hamiltonians, $\hat{\mathbf{H}}_i = \hat{\mathbf{T}} + V_i(\hat{\mathbf{R}})$, with kinetic and potential operators $\hat{\mathbf{T}}$ and $V_i(\hat{\mathbf{R}})$, respectively, $\hat{\boldsymbol{\mu}}$ is the transition dipole operator, assumed to be independent of $\hat{\mathbf{R}}$, and $\epsilon(t)$ the electric field. The potentials are found in Ref. 40. The vibrational Hamiltonians are represented on a Fourier grid[41] with N_R grid points yielding a total Hilbert space dimension of $M = 3N_R$. The equations of motion are solved by the Chebychev propagator for homogeneous and inhomogeneous Schrödinger equations, respectively.[39,41] An initial field of the form

$$\epsilon^{(0)}(t) = \epsilon_0 s(t)\cos(\Omega t) \quad (35)$$

is employed with ϵ_0 the maximum amplitude and Ω the central frequency of the field. The shape function $s(t)$ is taken to be $s(t) = \sin^2(\pi t/T)$, where T corresponds to the optimization time. The weight of the final-time objective, λ_0, is taken to be one such that the optimum corresponds to $-J_T = 1$.

B. State-independent intermediate-time cost ($g_b = 0$)

We investigate optimization of a state-to-state transfer ($N = 1$), taking for simplicity only the electronic states $X^1\Sigma_g^+(5s+5s)$ and $^1\Sigma_u^+(5s+5p)$ into account. The initial and target states are taken to be the vibrational eigenstates $v = 10$ and $v = 0$ of the $X^1\Sigma_g^+(5s+5s)$ electronic ground state. With a vibrational period of the initial state of 614 fs, the optimization time is set to $T = 1$ ps. The central frequency and the maximum field amplitude are taken to be $\Omega = \omega_{v=10\to v'=0}$ and $\epsilon_0 = 1 \times 10^{-2}$ a.u.

Convergence of the final-time objective J_T is shown in Fig. 2, comparing first-order (black circles) and second-order constructions of the algorithm. The second-order construction is determined by the choice of $\bar{A}$ which can be taken to be equal to some non-negative number, ε_A, cf. Eq. (14) (dotted and dashed lines in Fig. 2) or $\bar{A} = 0$ (first-order optimization, black circles in Fig. 2). The numerical estimate of A, cf. Eq. (25), is represented by the solid red line in Fig. 2. The latter choice might speed up convergence, but is more risky: Since $\bar{A} = 2A^{(i+1)}(\Delta\varphi)$ can become negative, the condition for monotonic convergence may be violated. This is clearly seen in Fig. 2. In the lower inset, monotonic convergence is lost for one step after the first iteration step. We find in this case, that the state change is almost maximal, $\|\Delta\varphi\| = 1.95 \leq 2$, i.e., the worst possible case that the optimization algorithm must deal with is reached. While the first-order construction converges faster initially, the upper inset shows that all second-order constructions supersede the first order one as the optimum is approached. This is readily understood by inspection of Eq. (12): the first-order contribution to the change in the field is closely related to the gradient of the functional.[48] Since the gradient vanishes close to the optimum, convergence of the first-order construction slows down as the optimum is approached. Variation of the non-negative number, ε_A, shows that an optimal choice of ε_A exists. However, this optimal choice cannot be determined *a priori*. In terms of convergence speed close to the optimum, it is therefore recommendable to employ the numerical estimate of $\bar{A}$ (red solid line in Fig. 2). Very similar behavior is found for optimization of a unitary transformation, a Hadamard gate ($N = 2$) on the lowest two vibrational eigenstates of the electronic ground state (data not shown).

C. State-dependent intermediate-time cost ($g_b \neq 0$)

State-to-state transfer from $v = 0$ to $v = 1$ and the Hadamard gate on the lowest two vibrational eigenstates of

the electronic ground state,

$$\hat{\mathbf{O}} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},$$

are optimized, taking into account an additional state-dependent cost, $g_b(\varphi, t)$. If both a state-dependent cost and a final-time target are present, the algorithm seeks to optimize a compromise between the two goals. The parameters λ_0 and λ_b determine the relative weight of each target. Monotonic convergence always refers to the value that the total functional, J of Eq. (1), takes; and each separate contribution to J does not need to converge monotonically. Below we will discuss convergence of both J and J_T. In order to render the optimal value of J independent of the choice of the weights λ_0, λ_b, we define a normalized functional,

$$J_{\text{norm}} = \frac{J}{\lambda_b - \lambda_0}, \quad \lambda_b \leq 0\,, \tag{36a}$$

$$J_{\text{norm}} = 1 - \frac{J - \lambda_0}{\lambda_b - \lambda_0}, \quad \lambda_b \geq 0\,, \tag{36b}$$

that converges toward one.

The cost g_b is employed to avoid any population transfer to a forbidden subspace, taken to be the ${}^1\Pi_g(5s + 4d)$ state, at all times $t \in [0, T]$.[33] This can be expressed by taking the operator $\hat{\mathbf{D}}$ in Eq. (5) to be one of the two choices,

$$\hat{\mathbf{D}} = \hat{\mathbf{P}}_{\text{allow}} = |e_1\rangle\langle e_1| + |e_2\rangle\langle e_2|, \quad \lambda_b \leq 0 \tag{37a}$$

$$\hat{\mathbf{D}} = \hat{\mathbf{P}}_{\text{forbid}} = |e_3\rangle\langle e_3|, \quad \lambda_b \geq 0\,, \tag{37b}$$

where $\hat{\mathbf{P}}_{\text{allow}}$ and $\hat{\mathbf{P}}_{\text{forbid}}$ denote the projectors onto the allowed and forbidden subspaces, respectively. The allowed subspace is formed by the $X\,{}^1\Sigma_g^+(5s + 5s)$ and ${}^1\Sigma_u^+(5s + 5p)$ states. The different signs of the weight λ_b indicate maximization of $\hat{\mathbf{P}}_{\text{allow}}$ and minimization of $\hat{\mathbf{P}}_{\text{forbid}}$ which is physically equivalent. Mathematically, the equivalence does not hold since g_b is a constraint and, therefore, must have its sign opposite to that of J_T. This is possible only for the choice of Eq. (37a). Note that the "wrong" sign of Eq. (37b) exemplifies the more general class of indefinite operators $\hat{\mathbf{D}}(t)$ for which it is not possible to construct a monotonically convergent algorithm using the previously available tools by a simple change of sign.

With the choice of g_b according to Eqs. (37), the analytical estimate of the parameter C of the second-order contribution is given by Eq. (24). Writing explicitly the change in the intermediate-time contribution to the functional due to the change in the states for a first order algorithm,

$$\Delta_{\varphi(t)} = R\left(\{\varphi_k^{(i+1)}\}, \epsilon^{(i)}, t\right) - R\left(\{\varphi_k^{(i)}\}, \epsilon^{(i)}, t\right)$$

$$= -\lambda_b \frac{1}{NT}\sum_{k=1}^{N}\langle \Delta\varphi_k(t)|\hat{\mathbf{D}}|\Delta\varphi_k(t)\rangle\,.$$

We find the necessary condition for monotonic convergence, $\Delta_{\varphi(t)} \geq 0$, to be always fulfilled for $\hat{\mathbf{D}} = \hat{\mathbf{P}}_{\text{allow}}$. A second-order construction is therefore not required,[33] corresponding to $C = 0$ in accordance with Eq. (24). In this case, $\bar{C}$ can be set to $-\varepsilon_C$, cf. Eq. (14), where ϵ_C is a non-negative number to check for improved convergence with a second-order algorithm. However, if the projector onto the forbidden subspace is employed in g_b, $\Delta_{\varphi(t)} \geq 0$ is not necessarily fulfilled and a second-order construction is required to ensure monotonic convergence. Equation (24) now yields a large negative number for C since λ_b is negative, and $\bar{C}$ is determined by C. Our approach goes beyond the results of Ref. 33 where a convergent (first order) algorithm was obtained only for negative semi-definite operators $\lambda_b\hat{\mathbf{D}}(t)$ in the additional constraint g_b. The second-order construction of Eq. (28) allows for a larger class of operators $\hat{\mathbf{D}}(t)$ in the state-dependent constraint g_b.

The final time is set to $T = 2$ ps, the central frequency of the guess field is chosen to be $\Omega = \omega_{v=0\to v'=10}$ and $\epsilon_0 = 2 \times 10^{-4}$ a.u. $\bar{A}$ is taken to be zero since our emphasis is on the choice of the parameter $\bar{C}$.

We first study the case $\hat{\mathbf{D}} = \hat{\mathbf{P}}_{\text{allow}}$ where the second-order construction is not required but may improve convergence, in analogy to Sec. IV B. Figure 3 compares the convergence of first- and second-order constructions. Taking $\bar{C}$ to be equal to $-\varepsilon_C$, cf. Eq. (14) (dotted and dashed lines in Fig. 3), does not affect monotonicity. However, it also does not yield faster convergence than the first-order algorithm (black circles). The numerical estimate, $\bar{C} = 2C^{(i+1)}(\Delta\varphi)$ (solid red line in Fig. 3), cf. Eq. (26b), neglecting ε_C in Eq. (14) is somewhat risky since $C^{(i+1)}(\Delta\varphi)$ can become positive such that this choice of $\bar{C}$ does not guarantee monotonic convergence. Indeed, small violations of monotonicity, for example, between steps 1180 and 1300, are observed in Fig. 3. However, this is more than compensated for by the improved speed of convergence as compared to the first order and the conservative choices $\bar{C} = -\varepsilon_C$. Very similar behavior is found for optimization of a state-to-state transition (data not shown).

For $\hat{\mathbf{D}} = \hat{\mathbf{P}}_{\text{forbid}}$, monotonic convergence needs to be ensured by a second-order construction with C given by Eq. (24). This choice corresponds to the green dashed line in Fig. 4 which studies convergence of the final-time objective, J_T, and the complete functional, J_{norm}, for a state-to-state

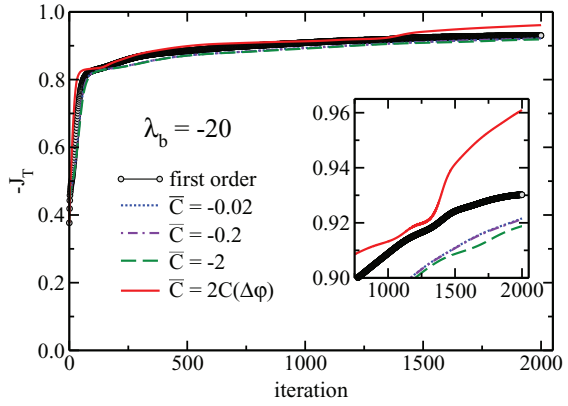

FIG. 3. Convergence of the first-order and second-order constructions for a Hadamard gate with a state-dependent cost ($\hat{\mathbf{D}} = \hat{\mathbf{P}}_{\text{allow}}$, i.e., the second-order construction is not required).

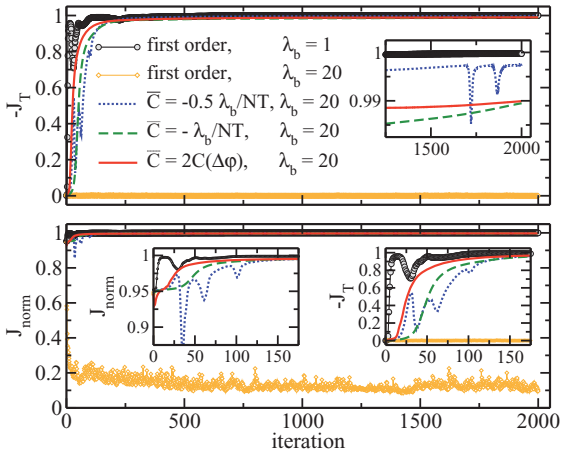

FIG. 4. Convergence of the first-order and second-order constructions for state-to-state transfer with state-dependent cost. The operator $\hat{\mathbf{D}}$ is taken to be the projector onto a forbidden subspace, i.e., the second-order construction *is* required.

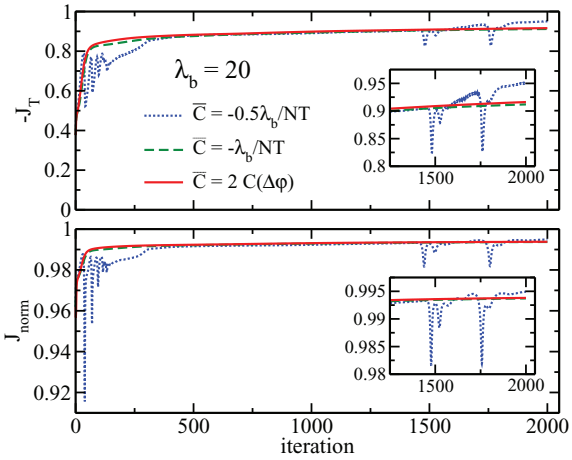

FIG. 5. Convergence of second-order constructions for a Hadamard gate with a state-dependent cost. The operator $\hat{\mathbf{D}}$ is taken to be the projector onto a forbidden subspace, i.e., the second-order construction *is* required.

optimization. Taking C somewhat smaller than the estimate of Eq. (24) may yield non-monotonic behavior, cf. the blue dotted line in Fig. 4. If one neglects the second-order contribution and the weight λ_b is large, the algorithm completely fails (orange diamonds in Fig. 4). For a small weight λ_b, the algorithm converges to an optimum but non-monotonic behavior is observed at intermediate iterations (black circles). Note that for small λ_b, the constraint g_b is almost not enforced due to insufficient weight. The best compromise between monotonic convergence and high fidelity is obtained for $\bar{C} = 2C^{(i+1)}(\Delta\varphi)$ (red solid line in Fig. 4). For the parameters for which the complete functional, J_{norm}, converges monotonically (green dashed and solid red lines in Fig. 4), monotonic behavior is observed also for the final-time objective, J_T. In general, this need not be the case. We attribute it in the current example to our choice of the guess field which is relatively weak such that the forbidden subspace is not strongly populated. The algorithm, therefore, starts out in the "right" direction for optimizing both targets, J_T and g_b, and it does not need to optimize one target at the expense of the other.

Figure 5 presents convergence of the final-time objective, J_T, and the complete functional, J_{norm}, for optimization of the Hadamard gate. Similarly to Fig. 4, convergence is almost identical for the analytical estimate of C based on Eq. (24) (green dashed line in Fig. 4) and the numerical estimate according to Eq. (26b) (red solid line). A small violation of the analytical estimate (blue dotted line in Fig. 5) leads to non-monotonic behavior but may yield larger fidelities after many iterations.

To summarize our numerical investigations, a second-order contribution can be employed to enforce monotonic convergence for functionals that are higher order polynomials in the states or correspond to expectation values of non-semidefinite operators. The numerical estimate of the second-order parameters might slightly violate monotonicity but yields the highest fidelities, especially as the optimum is approached. If a second-order contribution is not required by the functional, it may nevertheless be used to improve convergence. Also in this case the numerical estimate of the second-order parameters $\bar{A}$ and $\bar{C}$ turns out to be most efficient.

V. SUMMARY AND CONCLUSIONS

Applying Krotov's method[29] to quantum control, we have shown that monotonically convergent optimization algorithms are obtained for *any* quantum control problem provided that a second-order construction is employed. The equation for the optimized field then contains an additional term. Compared to a first-order algorithm,[27] only storage of the quantum states of the previous iteration and calculation of the second-order weight are additionally required. We have shown that the parameters for the second-order contribution can be estimated analytically based on the final-time target and intermediate-time "costs," the equations of motion and the dependence of the Hamiltonian on the control field or calculated numerically from the optimization history. This is due to the normalization of quantum state vectors and finiteness of physical control fields, implying that optimization is performed over compact sets of candidate states and controls, which has allowed us to significantly relax the conditions for Krotov's constructive proof.[29]

We have illustrated the power of our approach by applying it to two control problems for which no monotonically convergent algorithm existed – to target functionals that are higher order polynomials in the states and to state-dependent constraints expressed as expectation value of a non-semidefinite operator. Target functionals that are higher order polynomials in the states arise for optimization towards an equivalence class of operators rather than a specific operator. This is particularly relevant in quantum information where one is interested in the optimal evolution of a primary system alone, irrespective of its environment,[42] or in the entangling content of two-qubit gates.[28]

Our numerical examples illustrate that an analytical estimate of the algorithm parameters ensures monotonic convergence by taking all worst case scenarios for optimization into account. However, if the worst case scenarios are not

encountered, the analytical estimate imposes limits which are too severe, slowing down convergence. Estimating the algorithm parameters numerically based on the optimization history turns out to be a more efficient choice. The numerical estimate of the second-order parameters can also be employed to speed up convergence, in particular close to the optimum where the first-order contribution vanishes, for optimization problems where a second-order construction is not strictly required.

Note that the overall performance of the algorithm still depends on the the weights of the each term in the optimization functional and the optimization time T. The latter has little influence on the convergence once it is larger than the quantum speed limit.[43] The role of the weights is more subtle, in particular for multi-optimization problems.[44] A special role is taken by the weight, λ_a, of the term minimizing the field intensity: It determines the magnitude of the first-order contribution to the new field analogously to the step size in gradient-type algorithms.[23] Since its modulus, $|\lambda_a|$, is a free parameter of the algorithm, its choice may be used to further improve the convergence speed. An efficient optimization method is obtained by choosing $|\lambda_a|$ based on information from the second-order derivative of the functional with respect to the field, estimated with the Broyden-Fletcher-Goldfarb-Shanno algorithm.[45] Note that in terms of convergence speed this goes beyond our approach which only makes use of information from the second-order derivatives with respect to the states.

The work presented here opens up a whole range of new applications for quantum optimal control. It provides a general set of tools to study optimization of final-time functionals that are higher order polynomials in the states,[28] or optimization of time-dependent expectation values that were suggested for the control of high-harmonic emission,[32] or optimization for nonlinear equations of motion such as the Gross-Pitaevski equation, time-dependent Hartree-Fock equations or time-dependent density functional theory.[15,19,34] This set of tools allows for designing novel optimization functionals that capture the relevant physics without restriction to bilinear functionals.

ACKNOWLEDGMENTS

We would like to thank Reuven Eitan and David Tannor for their helpful comments on the manuscript. Financial support from the Deutsche Forschungsgemeinschaft (DFG) is gratefully acknowledged.

APPENDIX A: PROOF OF OPTIMALITY OF THE SECOND ORDER ANSATZ

Konnov and Krotov base their proof upon the following condition set $\mathfrak{A}$:[29]

1. The right-hand side of the equation of motion, $f(\vec{\varphi}, \epsilon, t)$, is bounded, i.e., $\exists\, K, L < \infty : \forall (\vec{\varphi}, \epsilon, t) \in \mathbb{R}^{2NM} \times \mathbb{E} \times [0, T]$, and $\|x\| \geq M \Rightarrow f(\vec{\varphi}, \epsilon, t) \leq L\|x\|$.
2. The Jacobian of the equation of motion is bounded, i.e., $\exists\, A < \infty : \forall (\vec{\varphi}, \epsilon, t)\, \mathbb{R}^{2NM} \times \mathbb{E} \times \in [0, T] : \|J\| \leq A$.
3. The functionals $J_T(\vec{\varphi})$ and $g(\vec{\varphi}, \epsilon, t)$ are twice differentiable and bounded, i.e., $\exists\, K, L < \infty : \forall \vec{\varphi} \in \mathbb{R}^{2NM}$, $\|\vec{\varphi}\| \geq L \Rightarrow J_T(\vec{\varphi}) \leq K\|\vec{\varphi}\|^2$, and $|g(\vec{\varphi}, \epsilon, t)| \leq K\|\vec{\varphi}\|^2\, \forall (t, \epsilon) \in [0, T] \times \mathbb{E}$.

Here, the real state vector $\vec{\varphi}(t)$ is a piecewise differentiable function for all $t \in [0, T]$ and the control ϵ is an element of the Banach space of continuous real valued functions on the interval $[0, T]$ with supremum norm $\|f(x)\|_\infty = \sup_{x\in[0,T]} |f(x)|$, i.e., $\epsilon \in (C[0, T], \|\cdot\|_\infty)$ with $\|\epsilon\|_\infty \leq E < \infty$. In quantum control, the components of $\vec{\varphi}(t)$ are the real and imaginary parts of the projections of N Hilbert space vectors on a suitable orthonormal basis of the M-dimensional Hilbert space. The norm of the Jacobian is then the matrix norm of the $2NM \times 2NM$ matrix, either column or row norm.

Based on $\mathfrak{A}$, Konnov and Krotov proved the following theorem:[29]

Theorem 1: *If the conditions $\mathfrak{A}$ hold, then for each process $w^{(i)}(\vec{\varphi}^{(i)}(t), \epsilon^{(i)}(t))$, there exists a solution for Φ to the extremization problem,*

$$R(\vec{\varphi}^{(i)}(t), \epsilon^{(i)}(t), t; \Phi) = \min_{\vec{\varphi}} R(\vec{\varphi}(t), \epsilon^{(i)}(t), t) \quad \forall\, t \in [0, T], \tag{A1}$$

$$G(\vec{\varphi}^{(i)}(T); \Phi) = \max_{\vec{\varphi}} G(\vec{\varphi}(T)), \tag{A2}$$

of the form

$$\Phi(\vec{\varphi}, \vec{\chi}, t) = \vec{\chi}(t) \cdot \vec{\varphi}(t) + \frac{1}{2} \Delta\vec{\varphi}(t) \cdot \hat{\boldsymbol{\sigma}}(t) \cdot \Delta\vec{\varphi}(t),$$

and the matrix function $\hat{\boldsymbol{\sigma}}(t)$ can be represented as

$$\hat{\boldsymbol{\sigma}}(t) = (\alpha(e^{\gamma(T-t)} - 1) + \beta) \cdot \mathbb{1} \equiv \sigma(t) \cdot \mathbb{1},$$

where α, $\beta < 0$, and $\gamma > 0$ are constants.

With this theorem, the crucial part of constructing the algorithm, i.e., determination of $\sigma(t)$ and thus Φ, is reduced to the determination of the constants α, β, and γ. Since the proof of Theorem 1 shows how to estimate the values of α, β, and γ, we will sketch it here. It is based on the following lemma which indicates why the conditions $\mathfrak{A}$ need to be imposed.

Lemma 1: Let the function $h(\vec{\varphi}) : \mathbb{R}^n \to \mathbb{R}$ satisfy the following conditions: $h \in C(\mathbb{R}^n)$ with $C(\mathbb{R}^n)$ denoting the space of continuous functions over $\mathbb{R}^n$, h is twice differentiable at $\vec{0}$ with $h(\vec{0}) = 0$ and $\nabla_{\vec{\varphi}} h(\vec{\varphi})|_{\vec{\varphi}=\vec{0}} = \vec{0}$ and $\exists\, K, L < \infty : \|\vec{\varphi}\| \geq L \Rightarrow h(\vec{\varphi}) \leq K\|\vec{\varphi}\|^2$. Then,

$$\sup_{\vec{\varphi}\in\mathbb{R}^n} \frac{h(\vec{\varphi})}{\|\vec{\varphi}\|^2} < \infty.$$

When using Lemma 1, two problems may arise in ensuring that the supremum is finite. (i) Small values of $\|\vec{\varphi}\|$ create a small denominator which may lead to large values of $\frac{h(\vec{\varphi})}{\|\vec{\varphi}\|^2}$ near $\vec{\varphi} = \vec{0}$. This is eliminated by the conditions of Lemma 1, $h(\vec{0}) = 0$ and $\nabla_{\vec{\varphi}} h(\vec{\varphi})|_{\vec{\varphi}=\vec{0}} = \vec{0}$. We show in

Appendices A 1 and A 2 that all quantities taking the role of $h(\vec{\varphi})$ satisfy these conditions so that we can employ Lemma 1. (ii) Large values of $\|\vec{\varphi}\|$ may cause large values of $h(\vec{\varphi})$ which in turn can lead to large values of $\frac{h(\vec{\varphi})}{\|\vec{\varphi}\|^2}$ as $\|\vec{\varphi}\| \to \infty$. The second problem is handled by imposing that $h(\vec{\varphi})$ grows at most quadratically with $\|\vec{\varphi}\|$ which is guaranteed by the conditions $\mathfrak{A}$. Continuity of $h(\vec{\varphi})$ then ensures finiteness of the supremum's argument for all intermediate values of $\|\vec{\varphi}\|$, $0 < \|\vec{\varphi}\| < \infty$.

1. Final-time contribution to the second-order ansatz: Proof of existence of finite *A*

Without any assumption on the normalization, i.e., without any reference to quantum control, we verify that A of Eq. (15) is well-defined with $A < \infty$ based on the conditions $\mathfrak{A}$ and using Lemma 1. Let $\vec{\psi}$ denote the change in the state vector. Specifically, we have to check that, for

$$h(\vec{\psi}) = \vec{\psi} \cdot \Delta \vec{f}$$

with Δf defined in Eq. (18), (i) h is twice differentiable at $\vec{0}$ with $h(\vec{0}) = 0$, (ii) $\nabla_{\vec{\psi}} h(\vec{\psi})|_{\vec{\psi}=\vec{0}} = \vec{0}$, and (iii) $h(\vec{\psi})$ is bounded, i.e., finite constants K and L exist such that $\exists\, K, L < \infty : \|\vec{\psi}\| \geq L \Rightarrow h(\vec{\psi}) \leq K\|\vec{\psi}\|^2$. Obviously h is twice differentiable since $\vec{f}$ is twice differentiable and $\vec{\chi}(T)$ is a constant with respect to $\vec{\psi}$. Since $J_T(\vec{\varphi}^{(i)} + \vec{\psi})|_{\vec{\psi}=0} - J_T(\vec{\varphi}^{(i)}) = J_T(\vec{\varphi}^{(i)}) - J_T(\vec{\varphi}^{(i)}) = 0$, we find $h(\vec{0}) = 0$. We check whether $\nabla_{\vec{\psi}} h(\vec{\psi})|_{\vec{\psi}=\vec{0}} = \vec{0}$,

$$\nabla_{\vec{\psi}} h(\vec{\psi})|_{\vec{\psi}=\vec{0}} = \left(\frac{\vec{\psi}}{\|\vec{\psi}\|}\right) \cdot \vec{\chi} + \left.\frac{\partial J_T(\vec{\varphi}^{(i)} + \vec{\psi})}{\partial \vec{\psi}}\right|_{\vec{\psi}=\vec{0}} \cdot \left(\frac{\vec{\psi}}{\|\vec{\psi}\|}\right)$$

$$= \vec{\chi} \cdot \hat{e}_{\vec{\psi}} + \frac{\partial J_T}{\partial \vec{\varphi}}(\vec{\varphi}^{(i)}) \cdot \hat{e}_{\vec{\psi}}.$$

Since we have to check these conditions at $t = T$ and remembering that $\vec{\chi}(T) = -\frac{\partial J_T}{\partial \vec{\varphi}}(\vec{\varphi}^{(i)}(T))$, we obtain the desired result,

$$\nabla_{\vec{\psi}} h(\vec{\psi})|_{\vec{\psi}=\vec{0},t=T} = \vec{\chi} \cdot \hat{e}_{\vec{\psi}(T)} + \frac{\partial J_T}{\partial \vec{\varphi}}(\vec{\varphi}^{(i)}(T)) \cdot \hat{e}_{\vec{\psi}(T)}$$

$$= \vec{\chi} \cdot \hat{e}_{\vec{\psi}(T)} - \vec{\chi} \cdot \hat{e}_{\vec{\psi}(T)} = \vec{0}.$$

Finally, the condition set $\mathfrak{A}$ tells us that $J_T(\vec{\varphi}) \overset{\|\vec{\psi}\|\to\infty}{=} O(\|\vec{\psi}\|^2)$, hence

$$h(\vec{\psi}) = (\vec{\chi} \cdot \vec{\psi}) + J_T(\vec{\varphi}^{(i)} + \vec{\psi}) - J_T(\vec{\varphi}^{(i)}) \overset{\|\vec{\psi}\|\to\infty}{=} O(\|\vec{\psi}\|) + O(\|\vec{\psi}\|^2) + O(1) = O(\|\vec{\psi}\|^2)$$

and condition (iii) is fulfilled. We may thus use Lemma 1 which guarantees the existence of A.

2. Intermediate-time contribution to the second-order ansatz: Proof of existence of finite *B*, *C*

Analogously to Appendix A 1, we now verify that B and C are well-defined with $0 < B < \infty$ and $C > -\infty$ without any assumption on the normalization of the state vector. The constant B defined in Eq. (20) is easily checked using Lemma 1. Similarly to proving the existence of finite A, we have to check whether the conditions for Lemma 1 are fulfilled for

$$h(\vec{\psi}(t)) = \vec{\psi}(t) \cdot \Delta \vec{f}.$$

In analogy to Appendix A 1, it is obvious that h is twice differentiable at $\vec{0}$ with $h(\vec{0}) = 0$. In order to check that $\nabla_{\vec{\psi}(t)} h(\vec{\psi}(t))|_{\vec{\psi}(t)=\vec{0}} = 0$, we use the product rule,

$$\nabla_{\vec{\psi}(t)} h(\vec{\psi}(t))|_{\vec{\psi}(t)=\vec{0}} = \left(\frac{\vec{\psi}(t)}{\|\vec{\psi}(t)\|}\right) \cdot \underbrace{\Delta \vec{f}(\vec{\psi}(t))|_{\vec{\psi}(t)=0}}_{=\vec{0}} + \underbrace{\vec{\psi}(t) \cdot \nabla_{\vec{\psi}(t)} (\Delta \vec{f})|_{\vec{\psi}(t)=0}}_{=\vec{0}} = \vec{0}.$$

Finally, the condition set $\mathfrak{A}$ tells us that $\|\vec{f}(\vec{\varphi})\| \overset{\|\vec{\psi}\|\to\infty}{=} O(\|\vec{\psi}\|)$, hence

$$h(\vec{\psi}) = \vec{\psi} \cdot \Delta \vec{f} \overset{\|\vec{\psi}\|\to\infty}{=} O(\|\vec{\psi}\|^2),$$

i.e., h is indeed quadratically bounded. We may thus use Lemma 1 which guarantees the existence of $B < \infty$. Moreover, $B > 0$ due to taking the square modulus inside the supremum.

To prove the existence of finite C defined by Eq. (22), we distinguish the two cases for large and small $\|\vec{\psi}\|$. For large $\|\vec{\psi}(t)\|$, the argument of the infimum is finite. This follows from $\dot{\vec{\chi}}(t)$ and $\vec{\chi}(t)$ being constant with respect to $\vec{\psi}(t)$ and the fact that for large $\|\vec{\varphi}\|$, we have $\|\vec{f}\| \leq K_1\|\vec{\varphi}\|$ and $|g| \leq K_2\|\vec{\varphi}\|^2$. Hence

$$\dot{\vec{\chi}}(t) \cdot \vec{\psi}(t) + \vec{\chi}(t) \cdot \Delta \vec{f} - \Delta g \overset{\|\vec{\psi}(t)\|\to\infty}{\leq} O(\|\vec{\psi}(t)\|^2).$$

For small $\|\vec{\psi}(t)\|$, we have to check that the zeroth and first-order terms in the denominator disappear. Inserting the expression for $\dot{\vec{\chi}}$ into the definition of C yields

$$C = \inf_{\vec{\psi}(t)\in\mathbb{R}^{2NM};t\in[0,T]} \frac{(-\nabla_{\vec{\varphi}(t)} \vec{f}^{\,T} \cdot \vec{\chi}(t)) \cdot \vec{\psi}(t) + \vec{\chi}(t) \cdot \Delta \vec{f} + \nabla_{\vec{\varphi}} g \cdot \vec{\psi}(t) - \Delta g}{(\vec{\psi}(t) \cdot \vec{\psi}(t))}, \tag{A3}$$

For sufficiently small ε and $\|\vec{\psi}(t)\| < \varepsilon$, we may approximate Δf and Δg to the first order,

$$\Delta \vec{f} \simeq \nabla_{\vec{\varphi}(t)} \vec{f} \cdot \vec{\psi}(t) + O(\|\vec{\psi}\|^2),$$
$$\Delta g \simeq \nabla_{\vec{\varphi}(t)} g \cdot \vec{\psi}(t) + O(\|\vec{\psi}\|^2).$$

This yields

$$\frac{(-\nabla_{\vec{\varphi}(t)} \vec{f}^T \cdot \vec{\chi}(t)) \cdot \vec{\psi}(t) + \nabla_{\vec{\varphi}(t)} g \cdot \vec{\psi}(t) + \vec{\chi}(t) \cdot \Delta \vec{f} - \Delta g}{(\vec{\psi}(t) \cdot \vec{\psi}(t))}$$
$$\xrightarrow{\|\vec{\psi}(t)\| \to 0} \frac{(-\nabla_{\vec{\varphi}(t)} \vec{f}^T \cdot \vec{\chi}(t)) \cdot \vec{\psi}(t) + \nabla_{\vec{\varphi}(t)} g \cdot \vec{\psi}(t) + \vec{\chi}(t) \cdot (\nabla_{\vec{\varphi}(t)} \vec{f} \cdot \vec{\psi}(t)) - \nabla_{\vec{\varphi}(t)} g \cdot \vec{\psi}(t) + O(\|\vec{\psi}(t)\|^2)}{(\vec{\psi}(t) \cdot \vec{\psi}(t))}$$
$$\xrightarrow{\|\vec{\psi}(t)\| \to 0} \frac{O(\|\vec{\psi}(t)\|^2)}{(\vec{\psi}(t) \cdot \vec{\psi}(t))} = O(1),$$

i.e., C remains finite also for small $\|\vec{\psi}(t)\|$.

APPENDIX B: APPLYING KROTOV's PROOF TO QUANTUM CONTROL PROBLEMS

We adapt Konnov and Krotov's results[29] to the case that $\vec{\varphi}$ is composed of the M real and imaginary expansion coefficients of N normalized quantum state vectors. In quantum control applications, some of the conditions $\mathfrak{A}$, cf. Appendix A, are trivially fulfilled. Specifically, the state vector $\vec{\varphi}$ is inherently bounded for any field ϵ if $\vec{\varphi}$ is obtained from the equation of motion for a given ϵ due to normalization. In particular, $\|\vec{\varphi}\|$ cannot become larger than $\sqrt{N}$ and we may reduce the candidate space $\mathbb{R}^{2NM}$ for $\vec{\varphi}$ to a *compact* subset of $\mathbb{R}^{2NM}$, namely, $\mathbb{X} = \{\vec{\varphi} \in \mathbb{R}^{2NM} | \|\vec{\varphi}\| \leq \sqrt{N}\} \subset \mathbb{R}^{2NM}$. The conditions concerning the behavior for large $\|\vec{\varphi}\|$ are then trivially fulfilled because the complete space of interest $[0, T] \times \mathbb{X} \times \mathbb{E}$, that is the set of all $\{(t, \vec{\varphi}, \epsilon)\}$, is compact. The boundedness of the right-hand side of the equations of motion, the Jacobian, and the functionals is then guaranteed by simply asking $f(t, \vec{\varphi}, \epsilon)$, $\mathbf{J}$, $J_T(\vec{\varphi})$, and $g(\epsilon, \vec{\varphi}, t)$ to be *regular*.

Due to the restriction of the states to the compact subset $\mathbb{X} \subset \mathbb{R}^{2NM}$, the proof simplifies for quantum control applications which we use to obtain straightforward estimates of the constants α, β, and γ. The changes in R and G due to variation of the state from the extremal point, $\vec{\varphi}^{(i)}$, to all possible states, $\vec{\varphi}^{(i)} + \vec{\psi}$, is measured by

$$\Delta G(\vec{\psi}) = G(\vec{\varphi}^{(i)}(T) + \vec{\psi}) - G(\vec{\varphi}^{(i)}(T)), \tag{B1}$$

$$\Delta R(\vec{\psi}(t), t) = R(\vec{\varphi}^{(i)}(t) + \vec{\psi}(t), \epsilon^{(i)}(t), t) - R(\vec{\varphi}^{(i)}(t), \epsilon^{(i)}(t), t). \tag{B2}$$

Then the global extremum conditions, Eqs. (A1) and (A2), correspond to $\Delta G(\vec{\psi}) \leq 0$ and $\Delta R(\vec{\psi}(t), t) \geq 0$. In quantum control applications, any state $\vec{\varphi}^{(i)} + \vec{\psi}$, that is, candidate for $\vec{\varphi}^{(i+1)}$ must also be normalized. Geometrically, all states $\vec{\varphi}^{(i)}$ and $\vec{\varphi}^{(i)} + \vec{\psi}$ lie therefore on a sphere of radius X. The norm of the vectors $\vec{\psi}$ varies between zero and $2X$ since for any two vectors $\vec{\varphi}^{(i)}$, $\vec{\varphi}^{(i+1)}$ in $\mathbb{X}$ the minimum distance is zero while the maximum distance is $2X$. This is illustrated in Fig. 6. The space for the state change vectors $\vec{\psi}$ is then given by $\mathbb{Y} = \{\vec{\psi} \in \mathbb{R}^{2NM} | \exists \vec{\varphi}^{(i)} \in \mathbb{X} : \vec{\varphi}^{(i)} + \vec{\psi} \in \mathbb{X}\} \subset \mathbb{R}^{2NM}$. In quantum control applications, it is thus sufficient to vary the vector $\vec{\varphi}^{(i+1)}$ over the sphere of radius X around the origin instead of the full $\mathbb{R}^{2NM}$.

With the definitions of A, B, and C, cf. Eqs. (15), (20) and (22), a strict maximum condition for G and minimum condition for R is transformed into

$$\sigma(T) < -2A, \tag{B3}$$

$$\frac{1}{2}\dot{\sigma}(t) - |\sigma(t)| B + C > 0. \tag{B4}$$

A solution $\sigma(t)$ fulfilling these inequalities exists and it is straightforward to check that the Konnov and Krotov's ansatz equation (13) satisfies (B3) and (B4). More generally, the ansatz[29]

$$\sigma(t) = \alpha(e^{\gamma(T-t)} - 1) + \beta$$

with $\alpha, \beta < 0$, and $\gamma > 0$ lead to the inequalities

$$\beta + 2A < 0, \tag{B5a}$$

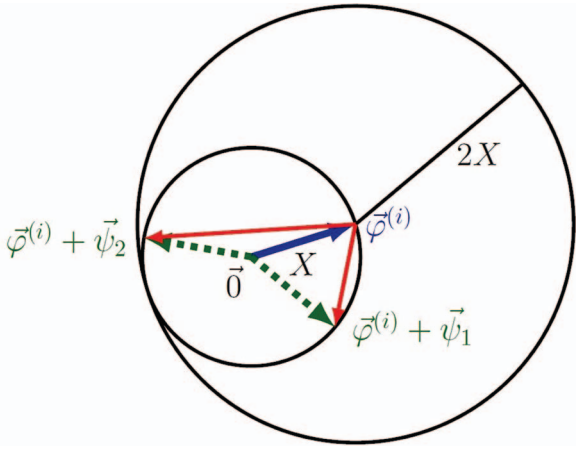

FIG. 6. All admissible states, $\vec{\varphi}^{(i)}$ and $\vec{\varphi}^{(i+1)} = \vec{\varphi}^{(i)} + \vec{\psi}$ lie on a sphere of radius $X = \sqrt{N}$ around the origin. Therefore, the norm of the change in the states, $\|\vec{\psi}\|$, varies between 0 and $2X$.

$$-\frac{\alpha\gamma}{2} - B\beta + C > 0\,, \tag{B5b}$$

which have at least one specific solution, namely, Eq. (13), with

$$\alpha = \frac{\bar{C}}{\bar{B}} - \bar{A} \quad \beta = -\bar{A} \quad \gamma = \bar{B}\,.$$

Note that if a set $\{\alpha_0, \beta_0, \gamma_0\}$ fulfills the inequalities (B5), then any other set $\{\alpha, \beta, \gamma\}$ with $\alpha \le \alpha_0 < 0$, $\beta \le \beta_0 < 0$, and $\gamma \ge \gamma_0 > 0$ does so, too. This flexibility can be utilized to estimate the constants A, B, and C. Therefore, the suprema in Eqs. (15) and (20) and the infimum in Eq. (22) can be estimated analytically.

APPENDIX C: ANALYTICAL ESTIMATE OF THE PARAMETERS OF THE SECOND-ORDER CONTRIBUTION

The arguments of the suprema, respectively of the infimum, in A, B, and C can be expressed in terms of the Taylor series starting at the first, respectively second, order. Evaluating the remainder term of these Taylor series, we obtain estimates for A, B, and C. Let $W(\vec{\varphi})$ be a scalar, vector, or matrix depending on $\vec{\varphi}$. $W_n(\vec{\varphi})$ denotes the Taylor expansion of W around $\vec{\varphi}^{(i)}$ starting at the n-th order or, in other words, $W_n(\vec{\varphi})$ equals $W(\vec{\varphi})$ minus the first $(n-1)$ terms of its Taylor expansion around $\vec{\varphi}^{(i)}$. For example, we obtain for a scalar field,

$$W_0(\vec{\varphi}) = W(\vec{\varphi}^{(i)})\,,$$
$$W_1(\vec{\varphi}) = W(\vec{\varphi}) - W(\vec{\varphi}^{(i)})\,,$$
$$W_2(\vec{\varphi}) = W(\vec{\varphi}) - W(\vec{\varphi}^{(i)}) - \nabla_\varphi W(\vec{\varphi}^{(i)})\cdot(\vec{\varphi} - \vec{\varphi}^{(i)})\,,$$

and so forth. The Taylor series starting at the n-th order can be approximated by evaluating the remainder term. For a scalar field $W(\vec{\varphi})$, this is given by

$$W_n(\vec{\varphi}) = \mathcal{R}^W_{\vec{\varphi}^{(i)},n}(\vec{\psi}) = \sum_{|\alpha|=n} \frac{1}{\alpha!}\partial^\alpha W(\vec{\varphi}^{(i)} + c\vec{\psi})\vec{\psi}^\alpha \tag{C1}$$

for a $c \in (0, 1)$ and with $\vec{\psi} = \vec{\varphi} - \vec{\varphi}^{(i)}$. Here α is a multi-index representing the $2NM$-tuple of natural numbers including zero,

$$|\alpha| = \sum_{i=1}^{2NM}\alpha_i, \quad \frac{1}{\alpha!} = \frac{1}{\prod_{i=1}^{2NM}\alpha_i!}, \quad \vec{\psi}^\alpha = \prod_{i=1}^{2NM}\psi_i^{\alpha_i}\,, \tag{C2}$$

and

$$(\partial^\alpha W)(\vec{\varphi}^{(i)}) = \prod_{i=1}^{2NM}\left.\frac{\partial^{\alpha_i} W(\vec{\varphi})}{\partial\varphi_i^{\alpha_i}}\right|_{\vec{\varphi}=\vec{\varphi}^{(i)}}. \tag{C3}$$

The remainder can be estimated by

$$\mathcal{R}^W_{\vec{\varphi}^{(i)},n}(\vec{\psi}) \le \frac{1}{\alpha!}M_n^W(\vec{\varphi}^{(i)})\vec{\psi}^\alpha, \quad |\alpha| = n\,, \tag{C4}$$

with

$$M_n^W(\vec{\varphi}^{(i)}) = \sup_{\vec{\psi}\in\mathbb{Y};|\alpha|=n}\partial^\alpha W(\vec{\varphi}^{(i)} + \vec{\psi})\,.$$

An estimate, that is, independent of the state $\vec{\varphi}^{(i)}$ is obtained by taking the supremum over all possible $\vec{\varphi}^{(i)}$, i.e., we define

$$\begin{aligned} M_n^W &\equiv \sup_{\vec{\varphi}^{(i)}\in\mathbb{X}} M_n(\vec{\varphi}^{(i)}) \\ &= \sup_{\vec{\varphi}^{(i)}\in\mathbb{X};\vec{\psi}\in\mathbb{Y};|\alpha|=n}\partial^\alpha W(\vec{\varphi}^{(i)} + \vec{\psi}) \\ &= \sup_{\vec{\Psi}\in\mathbb{X};|\alpha|=n}\partial^\alpha W(\vec{\Psi})\,. \end{aligned} \tag{C5}$$

This method to estimate the Lagrange remainder term lends itself to an intuitive geometrical interpretation, cf. Fig. 6. For given $\vec{\varphi}^{(i)}$, the Taylor expansion of W around $\vec{\varphi}^{(i)}$ starting at the n-th order, W_n, can be estimated by the supremum of the nth derivatives taken over the sphere around this state with radius $\|\vec{\psi}\| = y_0$. To give an estimate of W_n that holds for any $\vec{\varphi}^{(i)}$, we have to calculate the supremum around all possible state vectors. Since all $\vec{\varphi}^{(i)}$ are located on a sphere around the origin with radius $\sqrt{N}$, we simply need to take the supremum of the nth derivatives over all vectors within a ball around the origin with radius $2\sqrt{N}$. Since the Lagrange form of the remainder is based on the mean value theorem, the difference between two values of a function $W(\vec{\varphi}^{(i)} + \vec{\psi}) - W(\vec{\varphi}^{(i)})$ is estimated by the first derivative of the function at some point between $(\vec{\varphi}^{(i)} + \vec{\psi})$ and $\vec{\varphi}^{(i)}$. Since both $\vec{\varphi}^{(i)}$ and $(\vec{\varphi}^{(i)} + \vec{\psi})$ are located on the sphere of radius $\sqrt{N}$ around the origin, any difference of a function between these two points can only concern values of derivatives *inside* this sphere. Therefore, we can restrict the supremum to be taken over all states in $\mathbb{X}$ in the last line of Eq. (C5).

We now apply the estimate of the Lagrange remainder to derive the constants A, B, and C and use, instead of the vector notation above, the bracket notation in the following. Considering in Eq. (15), the Taylor expansion of $J_T(\{\varphi_k^{(i)}(T) + \Delta\varphi_k\})$ in $\Delta\varphi_k$ around $\varphi_k^{(i)}(T)$, the first order term cancels with $\langle\chi_k(T)|\Delta\varphi_k(T)\rangle + c.c.$ since $|\chi_k(T)\rangle$ is given in terms of the gradient of J_T, cf. Eq. (9b). The zeroth order term is nullified by $-J_T(\{\varphi_k^{(i)}(T)\})$,

$$\begin{aligned} A &= \sup_{\{\Delta\varphi_k\}}\frac{J_T\left(\left\{\varphi_k^{(i)}(T) + \Delta\varphi_k(T)\right\}\right) - J_T\left(\left\{\varphi_k^{(i)}(T)\right\}\right) + \sum_{k=1}^N\left[\langle\chi_k(T)|\Delta\varphi_k(T)\rangle + \langle\Delta\varphi_k(T)|\chi_k(T)\rangle\right]}{\sum_{k=1}^N\langle\Delta\varphi_k(T)\,|\,\Delta\varphi_k(T)\rangle} \\ &= \sup_{\{\Delta\varphi_k\}}\frac{J_{T,2}(\{\Delta\varphi_k(T)\})}{\sum_{k=1}^N\langle\Delta\varphi_k(T)\,|\,\Delta\varphi_k(T)\rangle}\,. \end{aligned}$$

The argument of the supremum in the definition of A, Eq. (15), can therefore be viewed as the Taylor expansion of $J_T(\{\varphi_k^{(i)}(T) + \Delta\varphi_k\})$ around the $\varphi_k^{(i)}(T)$ starting at the second order, divided by $\sum_{k=1}^N \langle \Delta\varphi_k(T) \,|\, \Delta\varphi_k(T)\rangle$. It is now possible to estimate $J_{T,2}$ by its Lagrange remainder according to Eq. (C4),

$$\begin{aligned} A &= \sup_{\{\Delta\varphi_k\}} \frac{\mathcal{R}^{J_T}_{\{\varphi_k^{(i)}(T)\},2}(\{\Delta\varphi_k(T)\})}{\sum_{k=1}^N \langle \Delta\varphi_k(T) \,|\, \Delta\varphi_k(T)\rangle} \\ &\le \sup_{\{\Delta\varphi_k\}} \frac{\frac{1}{2} M_2^{J_T}\left(\{\varphi_k^{(i)}(T)\}\right) \sum_{k=1}^N \langle \Delta\varphi_k(T) \,|\, \Delta\varphi_k(T)\rangle}{\sum_{k=1}^N \langle \Delta\varphi_k(T) \,|\, \Delta\varphi_k(T)\rangle} \\ &= \frac{1}{2} M_2^{J_T}\left(\{\varphi_k^{(i)}(T)\}\right) = \frac{1}{2} \sup_{\{\Delta\varphi_k\};|\alpha|=2} \partial^\alpha J_T(\{\Delta\varphi_k(T)\}), \end{aligned}$$

yielding Eq. (16). Note that for functionals J_T that are quadratic in the states, the *global* convergence condition (B3) coincides with the local one, $\nabla_\varphi^2 G < 0$, that was used in Ref. 16. This can be seen by inserting the ansatz for Φ, Eq. (8), into $\nabla_\varphi^2 G < 0$,

$$\left\| \nabla_\varphi^2 J_T\left(\varphi^{(i)}(T)\right) \right\| + \sigma(T) < 0\,.$$

To find an expression for the constant B defined by Eq. (20), we rewrite the change Δf_k in the equations of motion due to changes in the states, cf. Eq. (18), in terms of the Taylor expansion of the Hamiltonian in $\Delta\varphi_k(t)$ around the $\varphi_k^{(i)}(t)$ starting at the first order, $\hat{\mathbf{H}}_1$, and obtain

$$\begin{aligned} B \le & \sup_{\{\Delta\varphi_k\};t\in[0,T]} \left| \frac{\sum_{k=1}^N \left[\langle \Delta\varphi_k(t) \,|\, \hat{\mathbf{H}}_1(\varphi_k + \Delta\varphi_k, \epsilon^{(i)}) \;\big|\; \varphi_k^{(i)}(t)\rangle + \langle \varphi_k(t) \,|\, \hat{\mathbf{H}}_1^\dagger(\varphi_k + \Delta\varphi_k, \epsilon^{(i)}) \;\big|\; \Delta\varphi_k^{(i)}(t)\rangle \right]}{\sum_{k=1}^N \langle \Delta\varphi_k(t) \,|\, \Delta\varphi_k(t)\rangle} \right| \\ & + 2 \sup_{\{\Delta\varphi_k\};t\in[0,T]} \left| \frac{\sum_{k=1}^N \mathfrak{Im}\left\langle \Delta\varphi_k(t) \,\middle|\, \hat{\mathbf{H}}\left(\varphi_k^{(i)}(t) + \Delta\varphi_k(t), \epsilon^{(i)}, t\right) \,\middle|\, \Delta\varphi_k(t)\right\rangle}{\sum_{k=1}^N \left[\langle \Delta\varphi_k(t) \,|\, \Delta\varphi_k(t)\rangle\right]} \right|. \end{aligned}$$

Using the Cauchy Schwarz inequality for the scalar products in the argument of the first supremum and Eq. (C5) for the second supremum yields

$$\begin{aligned} B \le & \sup_{\{\Delta\varphi_k\};t\in[0,T]} \left(\frac{2}{\sum_{k=1}^N \langle \Delta\varphi_k(t) \,|\, \Delta\varphi_k(t)\rangle} \sum_{k=1}^N \left[\sqrt{\langle \Delta\varphi_k(t) \,|\, \Delta\varphi_k(t)\rangle} \cdot \|\hat{\mathbf{H}}_1(\varphi_k + \Delta\varphi_k, \epsilon^{(i)})\| \cdot \sqrt{\langle \varphi_k^{(i)}(t) | \varphi_k^{(i)}(t)\rangle} \right] \right) \\ & + 2 \sup_{\{\Delta\varphi_k\};t\in[0,T]} \left| \frac{\sum_{k=1}^N \mathfrak{Im} \langle \Delta\varphi_k(t) \,|\, \hat{\mathbf{H}}(\Delta\varphi_k(t), \epsilon^{(i)}, t) \,|\, \Delta\varphi_k(t)\rangle}{\sum_{k=1}^N \langle \Delta\varphi_k(t) \,|\, \Delta\varphi_k(t)\rangle} \right|. \end{aligned}$$

To evaluate the first supremum, we estimate the Taylor expansion of the Hamiltonian starting at the first order, $\hat{\mathbf{H}}_1$, by its Lagrange remainder,

$$\|\hat{\mathbf{H}}_1(\Delta\varphi_k)\| \le \sum_{k=1}^N M_1^{\hat{\mathbf{H}}|,k} \cdot \sqrt{\langle \Delta\varphi_k \,|\, \Delta\varphi_k\rangle} = \sum_{k=1}^N \sup_{\Delta\varphi_k;|\alpha|=1} |\partial^\alpha \hat{\mathbf{H}}(\Delta\varphi_k)| \cdot \sqrt{\langle \Delta\varphi_k \,|\, \Delta\varphi_k\rangle}. \tag{C6}$$

Note that the absolute value of the derivatives is required in the estimation of the Lagrange remainder since we need to estimate the norm of $\hat{\mathbf{H}}_1$. With $\sum_{k=1}^N \sqrt{\langle \varphi_k^{(i)}(t) | \varphi_k^{(i)}(t)\rangle} = \sqrt{N}$, we obtain Eq. (21) for B. In particular, for Hamiltonians that do not depend on the state, $\hat{\mathbf{H}}(\varphi_k^{(i)}(t) + \Delta\varphi_k(t), \epsilon^{(i)}) - \hat{\mathbf{H}}(\varphi_k^{(i)}(t), \epsilon^{(i)}) = 0$, and the first supremum can taken to be zero. Then B is given by the second supremum alone which is twice the maximum absolute value of the imaginary part the Hamiltonian's eigenvalues. That is, the second supremum is non-zero only for non-unitary time evolution. In summary, for linear equations of motion and unitary time evolution, $B = 0$. The differential inequality system for $\sigma(t)$ then reduces to

$$\begin{aligned} \frac{1}{2}\sigma(T) + A &< 0\,, \\ \frac{1}{2}\dot{\sigma}(t) + C &> 0\,, \end{aligned}$$

such that we obtain a linear solution for $\sigma(t)$,

$$\sigma(t) = \bar{C}(T - t) - \bar{A}\,. \tag{C7}$$

To estimate the constant C defined by Eq. (22), we rewrite it, using Eq. (A3),

$$C = \inf_{\{\Delta\varphi_k\};t\in[0,T]} \frac{\sum_{k=1}^N \left[\left\langle \chi_k^{(i)}(t) \right| f_{k,2}\left(\varphi_k^{(i)} + \Delta\varphi_k, \epsilon^{(i)}\right)\rangle + \left\langle f_{k,2}\left(\varphi_k^{(i)} + \Delta\varphi_k, \epsilon^{(i)}\right) \,\middle|\, \chi_k^{(i)}(t)\right\rangle \right] - g_2\left(\left\{\varphi_k^{(i)} + \Delta\varphi_k\right\}, t\right)}{\sum_{k=1}^N \langle \Delta\varphi_k(t) \,|\, \Delta\varphi_k(t)\rangle}, \tag{C8}$$

where $|f_{k,2}\rangle$ and g_2 are the Taylor expansions of $|f_k\rangle$ and g in $\Delta\varphi_k(t)$ around $\varphi_k^{(i)}(t)$ starting at the second order. Expressing $|f_{k,2}\rangle$

in terms of the Taylor expansion of the Hamiltonian starting at first order, $|f_{k,2}\rangle = \hat{\mathbf{H}}_1(\varphi_k^{(i)} + \Delta\varphi_k, \epsilon^{(i)})|\Delta\varphi_k(t)\rangle$, and introducing the approximation,

$$-C \geq \sup_{\{\Delta\varphi_k\};t\in[0,T]} -\frac{\sum_{k=1}^{N}\left[\langle\chi_k^{(i)}(t)|\,\hat{\mathbf{H}}_1\left(\varphi_k^{(i)}+\Delta\varphi_k,\epsilon^{(i)}\right)|\,\Delta\varphi_k(t)\rangle + \langle\Delta\varphi_k(t)|\,\hat{\mathbf{H}}_1^{+}\left(\varphi_k^{(i)}+\Delta\varphi_k,\epsilon^{(i)}\right)|\,\chi_k^{(i)}(t)\rangle\right]}{\sum_{k=1}^{N}\langle\Delta\varphi_k(t)|\,\Delta\varphi_k(t)\rangle}$$

$$+ \sup_{\{\Delta\varphi_k\};t\in[0,T]} \frac{g_2\left(\left\{\varphi_k^{(i)}+\Delta\varphi_k\right\},t\right)}{\sum_{k=1}^{N}\langle\Delta\varphi_k(t)|\,\Delta\varphi_k(t)\rangle},$$

we may reuse our results for $\hat{\mathbf{H}}_1$, cf. Eq. (C6), together with $M_1^{-}\hat{\mathbf{H}}^{,k} = -M_1^{\hat{\mathbf{H}}^{,k}}$ to estimate the first term. The estimation of the second term involving g_2 proceeds analogously to that of $J_{T,2}$, with M_2^g given by

$$M_2^g = \sup_{\substack{\{\Delta\varphi_k\};\,t\,\in\,[0,\,T]\\ |\alpha|=2}} \partial^\alpha g\left(\{\Delta\varphi_k\},t\right).$$

We thus obtain Eq. (23) for C. For Hamiltonians that depend on the state, the first term in the right-hand side of Eq. (23) is non-zero. Note that the norm of all adjoint vectors $\chi_k^{(i)}$ is equal to $\sqrt{N}$ only for state-independent constraint g. For state-dependent constraint g, $\chi_k^{(i)}$ is the solution of an inhomogeneous Schrödinger equation and its norm may be smaller or larger than $\sqrt{N}$. However, the norm of all adjoint vectors $\chi_k^{(i)}$ is always known since the $\chi_k^{(i)}$ are calculated in the previous iteration step, i. So in order to estimate the first term of the right-hand side of Eq. (23), we only need to determine $\|\partial\hat{\mathbf{H}}\|$ which is also needed for estimating B. In this case, $-C$ is then given by the sum of the spectral radius of the operator $\|\partial\hat{\mathbf{H}}\|$ and the eigenvalue of $\hat{\mathbf{D}}(t)$ with largest magnitude.

[1] S. A. Rice and M. Zhao, *Optical Control of Molecular Dynamics* (Wiley, New York, 2000).
[2] P. Brumer and M. Shapiro, *Principles and Applications of the Quantum Control of Molecular Processes* (Wiley Interscience, New York, 2003).
[3] D. Tannor and S. Rice, J. Chem. Phys. **83**, 5013 (1985).
[4] D. Tannor, R. Kosloff, and S. Rice, J. Chem. Phys. **85**, 5805 (1986).
[5] P. Brumer and M. Shapiro, J. Chem. Phys. **84**, 4103 (1986).
[6] P. Brumer and M. Shapiro, Chem. Phys. Lett. **126**, 541 (1986).
[7] W. Warren, H. Rabitz, and M. Dahleh, Science **259**, 1581 (1993).
[8] D. Tannor, V. Kazakov, and V. Orlov, in *Time-dependent Quantum Molecular Dynamics*, edited by J. Broeckhove and L. Lathouwers (Plenum, New York, 1992), pp. 347–360.
[9] R. S. Judson and H. Rabitz, Phys. Rev. Lett. **68**, 1500 (1992).
[10] J. Somlói, V. A. Kazakov, and D. J. Tannor, Chem. Phys. **172**, 85 (1993).
[11] W. Zhu, J. Botina, and H. Rabitz, J. Chem. Phys. **108**, 1953 (1998).
[12] N. Doslić, K. Sundermann, L. González, O. Mó, J. Giraud-Girard, and O. Kühn, Phys. Chem. Chem. Phys. **1**, 1249 (1999).
[13] L. Wang, H.-D. Meyer, and V. May, J. Chem. Phys. **125**, 014102 (2006).
[14] U. Hohenester and G. Stadler, Phys. Rev. Lett. **92**, 196801 (2004).
[15] E. Räsänen, A. Castro, J. Werschnik, A. Rubio, and E. K. U. Gross, Phys. Rev. Lett. **98**, 157404 (2007).
[16] S. E. Sklarz and D. J. Tannor, Phys. Rev. A **66**, 053619 (2002).
[17] C. P. Koch, J. P. Palao, R. Kosloff, and F. Masnou-Seeuws, Phys. Rev. A **70**, 013402 (2004).
[18] U. Hohenester, P. K. Rekdal, A. Borzì, and J. Schmiedmayer, Phys. Rev. A **75**, 023602 (2007).
[19] J. Grond, G. von Winckel, J. Schmiedmayer, and U. Hohenester, Phys. Rev. A **80**, 053625 (2009).
[20] V. May, D. Ambrosek, M. Oppel, and L. González, J. Chem. Phys. **127**, 144102 (2007).
[21] Y. Ohtsuki and K. Nakagami, Phys. Rev. A **77**, 033414 (2008).
[22] T. E. Skinner, T. O. Reiss, B. Luy, N. Khaneja, and S. J. Glaser, J. Magn. Reson. **163**, 8 (2003).
[23] N. Khaneja, T. Reiss, C. Kehlet, T. Schulte-Herbrüggen, and S. J. Glaser, J. Magn. Reson. **172**, 296 (2005).
[24] J. P. Palao and R. Kosloff, Phys. Rev. Lett. **89**, 188301 (2002).
[25] C. Tesch and R. de Vivie-Riedle, Phys. Rev. Lett. **89**, 157901 (2002).
[26] P. Treutlein, T. W. Hänsch, J. Reichel, A. Negretti, M. A. Cirone, and T. Calarco, Phys. Rev. A **74**, 022312 (2006).
[27] J. P. Palao and R. Kosloff, Phys. Rev. A **68**, 062308 (2003).
[28] M. M. Müller, D. Reich, H. Yuan, K. B. Whaley, J. Vala, T. Calarco, and C. P. Koch, Phys. Rev. A **84**, 042315 (2011).
[29] A. Konnov and V. Krotov, Autom. Remote Cont. (Engl. Transl.) **60**, 1427 (1999).
[30] Y. Ohtsuki, G. Turinici, and H. Rabitz, J. Chem. Phys. **120**, 5509 (2004).
[31] A. Kaiser and V. May, J. Chem. Phys. **121**, 2528 (2004).
[32] I. Şerban, J. Werschnik, and E. K. U. Gross, Phys. Rev. A **71**, 053810 (2005).
[33] J. P. Palao, R. Kosloff, and C. P. Koch, Phys. Rev. A **77**, 063412 (2008).
[34] M. Mundt and D. J. Tannor, New J. Phys. **11**, 105038 (2009).
[35] S. Mukamel, *Principles of Nonlinear Optics and Spectroscopy* (Oxford University Press, Oxford, 1995).
[36] J. Zhang, J. Vala, S. Sastry, and K. B. Whaley, Phys. Rev. A **67**, 042313 (2003).
[37] J. Zhang, J. Vala, S. Sastry, and B. Whaley, Phys. Rev. Lett. **93**, 020502 (2004).
[38] A. Micheli, G. K. Brennen, and P. Zoller, Nat. Phys. **2**, 341 (2006).
[39] M. Ndong, H. Tal-Ezer, R. Kosloff, and C. P. Koch, J. Chem. Phys. **130**, 124108 (2009).
[40] S. J. Park, S. W. Suh, Y. S. Lee, and G. H. Jeung, J. Mol. Spectrosc. **207**, 129 (2001).
[41] R. Kosloff, J. Phys. Chem. **92**, 2087 (1988).
[42] M. D. Grace, J. Dominy, R. L. Kosut, C. Brif, and H. Rabitz, New J. Phys. **12**, 015001 (2010).
[43] T. Caneva, M. Murphy, T. Calarco, R. Fazio, S. Montangero, V. Giovannetti, and G. E. Santoro, Phys. Rev. Lett. **103**, 240501 (2009).
[44] R. Chakrabarti, R. Wu, and H. Rabitz, Phys. Rev. A **78**, 033414 (2008).
[45] R. Eitan, M. Mundt, and D. J. Tannor, Phys. Rev. A **83**, 053426 (2011).

[46] In principle, all functionals depend on the state vectors and their complex conjugates, $|\varphi_k(t)\rangle$ and $\langle\varphi_k(t)|$, and differentiation needs to be carried out with respect to the state vectors and their complex conjugates. To simplify notation, we indicate by $\varphi_k(t)$ the dependence on both the bra and the ket vectors.
[47] Note that the choice $\varepsilon_i = 0$ ($i = A, B, C$), in the worst case, allows for iterations in which no change towards the objective due to the change in the state variables is achieved. However, we will ensure strict monotonic convergence with respect to the change in the field such that in practice even the choice $\varepsilon_i = 0$ will almost always lead to strict monotonic convergence.
[48] Specifically, the gradient in an implementation of the GRAPE algorithm employing a sequential update of the field coincides with the first order term of Eq. 12 in Ref. 23.