Quantum Optimal Control Theory of High Harmonic Generation

Thesis for the degree of
"Doctor of Philosophy"

By Ido Schaefer

Submitted to the Senate of the Hebrew University of Jerusalem
5/19

This work was carried out under the supervision of

Prof. Ronnie Kosloff

Abstract

High-harmonic-generation (HHG) is a highly nonlinear up-conversion process, in which an incident radiation field is converted into high multiples (harmonics) of the incoming frequency. Currently, HHG provides the only tabletop source of extreme UV (XUV) and soft X-ray coherent radiation. Important applications include attosecond pulse generation and high-energy atomic physics experiments.

The efficiency of the HHG is very low. This motivated the development of *quantum coherent control* schemes for the enhancement of the process. Learning algorithms were applied successfully for both theoretical and experimental optimization of HHG. However, this approach is limited by the low efficiency of the search process.

Quantum optimal-control-theory (OCT) is the most successful optimization method for quantum coherent control problems. It is based on the formulation of the control task as a maximization problem by the calculus of variation formalism. The search process of OCT utilizes, explicitly or implicitly, the gradient information of the optimization hypersurface.

In the present research, a theoretical method of optimization of HHG is developed in the framework of OCT. A numerical search is performed to locate an optimal driving pulse profile for a selective enhancement of harmonics. The optimal pulse is constructed from a predefined available band of the laser source.

The control requirements include the maximization of the emission in a selected frequency or spectral region, the minimization of the total energy of the driving pulse and

minimization of permanent ionization.

At the first stage, we address the more general problem of optimization of harmonic-generation (HG). After the establishment of the basic principles, we address issues which are unique to the HHG problem.

A major challenge in the optimization of HHG is a reliable numerical simulation of the HHG dynamics, which is known to be a difficult task. A reliable description requires highly accurate numerical tools.

The absorbing boundary conditions are realized by a complex absorbing potential (CAP). A new optimization method for the construction of the CAP is employed in order to minimize reflection and transmission from the boundaries.

The propagation is performed by a new highly accurate and efficient scheme, which is based on a semi-global propagation approach. The propagation method is adjusted to non-Hermitian dynamics, which characterize the dynamics under the influence of the CAP (Chapter 3). The application to the simulation of the dynamics of HHG is demonstrated.

A simple optimization scheme for the OCT formulation is employed in the earlier publication (Chapter 2). This scheme is later rejected for being too slow for the HHG problem, which is considerably more complex than the low-order harmonic problems. A more sophisticated quasi-Newton method (BFGS) is employed instead. It was required to adjust the optimization method to the specific control problem due to several unique issues.

The method is demonstrated in both low-order harmonic generation problems and HHG problems. The spectrum of the driving pulse is successfully restricted to the required spectral band. Selected frequencies are successfully enhanced, and the permanent ionization probability is successfully restricted. A comparison to reference pulses demonstrates that a significant enhancement is achieved by considerably lower total energy of the pulse and permanent ionization probability.

We demonstrate the ability of violation of the HHG selection rules, which confine the

emission spectrum to odd harmonics. This possibility is allowed due to the break of the characteristic symmetry of the typical pulses employed for HHG.

Contribution letter

Ido Schaefer is the main author of all papers in this collection.

The papers *Optimal control theory of harmonic generation* and *Optimization of high harmonic generation by optimal control theory—climbing a mountain in extreme conditions* represent the work of Ido Schaefer under the supervision of Prof. Ronnie Kosloff.

The paper *Semi-global approach for propagation of the time-dependent Schrödinger equation for time-dependent and nonlinear problems* gives a broad exposition of a propagation method developed by Prof. Hillel Tal-Ezer. The application to non-Hermitian dynamics, the application to the physical situation of HHG, the error analysis, and the mathematical and numerical insights are the work of Ido Schaefer under the supervision of Prof. Ronnie Kosloff.

Contents

1 Introduction 8

2 Optimal-control theory of harmonic generation 12

3 Semi-global approach for propagation of the time-dependent Schrödinger equation for time-dependent and nonlinear problems 26

4 Optimization of high harmonic generation by optimal control theory—climbing a mountain in extreme conditions 74

5 Discussion and conclusion 140

Chapter 1

Introduction

When high intensity light is irradiated on an atomic or molecular gas, higher frequencies are generated [11, 12, 15, 16]. This phenomenon is known as high harmonic generation (HHG) [7, 8, 13]. The current approach to attosecond pulse generation [3, 6, 14, 17] is based on this phenomenon.

It is of great interest to *control* the high harmonic generation phenomenon in order to increase the intensity of the emitted field at the frequencies of interest. Other desirable requirements that can be added to the control task are an economical use of the laser intensity, and limitation of the ionization rate.

The problem may be addressed by means of *quantum control*. Quantum control deals with the task of controlling quantum systems by time-dependent electric or magnetic fields. This can be achieved by an appropriate design of laser pulses, using pulse shaping techniques.

In order to control a quantum system it is desirable to find the *optimal field* for the task of interest. Two approaches have evolved for seeking the optimal field for quantum control problems:

1. Experimentally, by a sophisticated trial and error process using a genetic algorithm;

2. By theoretical calculation, using our knowledge about the quantum system.

The main importance of a theoretical method of calculation lies in the feasibility of investigation of the control mechanism. A theoretical calculation provides direct knowledge about the *evolution* of the quantum process. In contrary, experimental measurements are blind to the details of the evolution; any conclusion can only be deduced *indirectly* from the measured observables.

Another advantage of a theoretical method of calculation is related to the *optimization process*. The optimization method in a theoretical calculation relies on a knowledge about the *direction of search*, which is unavailable in the experimental approach. As a result, the theoretical optimization process is advantageous over the experimental one, which can rarely be exhausted with a practical number of experimental iterations.

Optimal control theory (OCT) is currently most successful theoretical method available for finding an optimal field in quantum control problems. In the framework of OCT, the problem is formulated as a maximization problem using the calculus of variation formalism (see [1, 10, 19, 24]).

Great progress in the task of controlling HHG has been achieved in the first decade of this century by using an experimental approach [18, 23, 25]. However, basic understanding of the mechanisms of the optimized fields is still lacking.

This motivates the development of a theoretical optimization method for HHG. However, the development of an OCT scheme poses several theoretical and numerical challenges, which delayed the development of a successful scheme until the current decade.

In principle, it is possible to perform theoretical optimization of HHG based on a genetic algorithm scheme, as in the experimental search. This approach has been applied in several theoretical studies [2, 5, 9, 21]. However, the computational simulation of the quantum dynamics is orders of magnitude slower compared to the experimental time-scales. This drawback combined with the low efficiency of the genetic algorithm optimization

results in an exceedingly slow optimization process. Consequently, the exhaustion of the search process becomes infeasible in reasonable time. Thus, in our view, this approach is inferior and should be avoided if possible. It is usually employed as a test of the control opportunities in experimental settings.

The current research aims at the development of an OCT scheme for the optimization of HHG. It represents the first successful effort of addressing this task. The task can be divided into two parts:

1. The development of an optimization scheme for the more general problem of optimization of harmonic generation; the HHG problem can be considered as a special case of the general problem.

2. The development of tools which are required specifically to the particular problem of HHG; several features of the HHG problem require a special treatment.

The original research goals were the following:

1. Development of theoretical tools for optimal control of high harmonic generation;

2. Exploration and investigation of new harmonic generation mechanisms;

3. Optional: Prediction of optimal fields for the generation of high harmonics experimentally.

The present thesis demonstrates the realization of the first research goal. The finding related to the second goal are not presented in the current framework. The realization of the third goal requires further work, and constitutes an important topic for a future research.

In Chapter 2, the general theory of OCT of harmonic generation is established. A simple optimization scheme is proposed. The method is applied to low-order harmonic

generation problems as well as to a below-threshold HHG problem. Chapter 3 describes the propagation method and its application to the current physical situation. In Chapter 4, the theory is further developed, with application to above-threshold HHG. An improved optimization scheme is also described. We conclude in Chapter 5 by a general discussion on the whole research, with an outlook for future directions of further research.

Chapter 2

Optimal-control theory of harmonic generation

Published; full citation:

Ido Schaefer and Ronnie Kosloff, *Optimal-control theory of harmonic generation*, Phys. Rev. A 86 (2012), 063417.

In the present chapter, the theory of OCT of harmonic generation (HG) is developed. A simple optimization scheme is applied, based on a relaxation process.

The paper focuses on the application to low-order HG problems. A single example of HHG control is also presented. However, the problem is an atypical HHG problem, where one of the Bohr-frequencies of the system is targeted (see Chapter 4). In the chosen HHG problem, the employment of absorbing boundaries can be avoided by the requirement of complete elimination of permanent ionization. This considerably simplifies the application. However, the requirement of complete elimination of permanent ionization was found to be unachievable in typical HHG problems, which necessitates the employment of absorbing boundaries. The application to typical HHG problems is thoroughly treated in Chapter 4.

The final application to HHG in Chapter 4 differs from that presented in the current

chapter in several aspects. In Chapter 4, the optimization method is replaced by a more sophisticated one. The method of restriction of permanent ionization is replaced by another method which allows a predefined permanent ionization probability. The required boundary conditions of the driving field are not imposed in the basic formulation presented in the present chapter; this is corrected in Chapter 4.

PHYSICAL REVIEW A **86**, 063417 (2012)

Optimal-control theory of harmonic generation

Ido Schaefer* and Ronnie Kosloff†

The Fritz Haber Research Center for Molecular Dynamics, The Institute of Chemistry, The Hebrew University of Jerusalem, Jerusalem 91904, Israel

(Received 22 October 2012; published 26 December 2012)

Coherent control of harmonic generation was studied theoretically. A specific harmonic order was targeted. An optimal control theory was employed to find the driving field where restrictions were imposed on the frequency band. Additional restrictions were added to suppress undesired outcomes, such as ionization and dissociation. The method was formulated in the frequency domain. An update procedure for the field based on relaxation was employed. The method was tested on several examples demonstrating the generation of high frequencies from a driving field with a restricted frequency band.

DOI: 10.1103/PhysRevA.86.063417 PACS number(s): 32.80.Qk, 37.10.Jk, 42.65.Ky

I. INTRODUCTION

When high intensity light is irradiated on an atomic or molecular gas, higher frequencies are generated [1,2]. This phenomenon is known as high harmonic generation [3–5]. The current approach to attosecond pulse generation [6] is based on this phenomenon. Significant effort has, therefore, been devoted to optimizing the process of harmonic generation [5,7,8]. Optimizing by control of the phase and amplitude of the incident light suggests coherent control [9].

The present paper addresses theoretically the issue of an optimal control strategy for harmonic generation. The target of optimization is the system's dipole operator. To be a source of radiation, the acceleration of the dipole operator should oscillate in the frequency which is much higher than the driving field frequency. The idea is to exploit the significant theoretical development in optimal control theory (OCT) [10–14]. The hope is that the optimized pulses will unravel specific mechanisms of harmonic generation.

OCT for harmonic generation poses two significant challenges:

(1) a constraint on the bandwidth of the incident control pulse has to be imposed;

(2) the target is designated in the frequency domain while OCT is typically formulated in the time domain.

Several suggestions for dealing with the first issue appear in the literature in a more general context [13–20]. It has been attempted to deal with the second issue by the means of the general OCT formulation of time-dependent targets [13,21]. However, no results from this approach have been reported.

In the present study, the two challenges are overcome by the formulation of the control problem in the frequency domain, replacing the formulation in the time domain. In this approach, the frequency requirements of the problem are expressed in a natural and direct way. A simple and effective optimization procedure, suitable for our formulation, is suggested.

II. OCT OF TIME DEPENDENT PROBLEMS

We first review the formulation of optimal control theory for time dependent problems. Let us denote the time-dependent state of the system by $|\psi(t)\rangle$, the drift Hamiltonian by $\hat{\mathbf{H}}_0$, and the driving field by $\epsilon(t)$. The dynamics of the system is governed by the time-dependent Schrödinger equation, under a given initial condition:

$$\frac{\partial|\psi(t)\rangle}{\partial t} = -i\hat{\mathbf{H}}(t)|\psi(t)\rangle, \quad |\psi(0)\rangle = |\psi_0\rangle \tag{1}$$

where $\hat{\mathbf{H}}(t) = \hat{\mathbf{H}}_0 - \hat{\mu}\epsilon(t)$. (Atomic units are used throughout, so we set $\hbar = 1$.) The optimization functional for time-dependent targets becomes (see [13,21–23])

$$J \equiv J_{\max} + J_{\text{bound}} + J_{\text{penal}} + J_{\text{con}}, \tag{2}$$

$$J_{\max} \equiv \int_0^T w(t)\langle\psi(t)|\hat{\mathbf{O}}(t)|\psi(t)\rangle\, dt, \quad w(t) \geqslant 0, \quad \int_0^T w(t)\, dt = 1 \tag{3}$$

$$J_{\text{bound}} \equiv \kappa\langle\psi(T)|\hat{\mathbf{O}}(T)|\psi(T)\rangle, \quad \kappa \geqslant 0 \tag{4}$$

$$J_{\text{penal}} \equiv -\alpha\int_0^T \epsilon^2(t)\, dt, \quad \alpha > 0 \tag{5}$$

$$J_{\text{con}} \equiv -2\text{Re}\int_0^T \langle\chi(t)|\frac{\partial}{\partial t} + i\hat{\mathbf{H}}(t)|\psi(t)\rangle\, dt, \tag{6}$$

where $\hat{\mathbf{O}}(t)$ is the time-dependent target operator, $|\chi(t)\rangle$ is a Lagrange-multiplier function, and T is the final time. $J_{\max}$ represents the target to be maximized. J_{bound} is a boundary term. The inclusion of this term prevents boundary problems (see [23], Sec. 2.2). J_{penal} is a penalty term on the intensity of $\epsilon(t)$. J_{con} represents the constraint on the dynamics of the system—the time-dependent Schrödinger equation.

The resulting Euler-Lagrange equations become

$$\frac{\partial|\psi(t)\rangle}{\partial t} = -i\hat{\mathbf{H}}(t)|\psi(t)\rangle, \quad |\psi(0)\rangle = |\psi_0\rangle \tag{7}$$

$$\frac{\partial|\chi(t)\rangle}{\partial t} = -i\hat{\mathbf{H}}(t)|\chi(t)\rangle - w(t)\hat{\mathbf{O}}(t)|\psi(t)\rangle, \quad |\chi(T)\rangle = \kappa\hat{\mathbf{O}}(T)|\psi(T)\rangle \tag{8}$$

$$\hat{\mathbf{H}}(t) = \hat{\mathbf{H}}_0 - \hat{\mu}\epsilon(t), \quad \epsilon(t) = -\frac{\text{Im}\langle\chi(t)|\hat{\mu}|\psi(t)\rangle}{\alpha}. \tag{9}$$

*ido.schaefer@mail.huji.ac.il
†ronnie@fh.huji.ac.il

This formulation is suitable for targets that are well defined in the time domain. For problems with frequency requirements this approach has to be modified.

III. OCT OF HARMONIC GENERATION

The harmonic generation problem may be divided into two distinct parts:

(1) The driving field spectrum has to be restricted to the frequency range available from the source.

(2) The intensity of the emission of the system in the desired portion of the spectrum has to be maximized.

These two parts will be treated separately in the next two subsections.

A. Restriction of the driving field spectrum

The task of restricting the driving field spectrum is of considerable importance in OCT; the reason is that most of the computed fields turn out to be too oscillatory to be produced experimentally. This problem may be overcome by limiting the spectrum of the field to sufficiently low frequencies.

Several approaches for achieving this goal have been proposed. In Ref. [13], a spectral filtration of the driving field is performed in each iteration of the Krotov algorithm. This approach leads to a nonmonotonic convergence of the optimization procedure. In Ref. [14], a two-dimensional penalty term is introduced in order to control the spectral properties of the driving field. This approach might lead to numerical instabilities or to nonmonotonic convergence of the optimization procedure (see [23], Sec. 3.2.1). In Ref. [15], the problem of optimization of a general driving field function is replaced by the optimization of the coefficients of a list of frequency terms.

In the present approach the restriction on the field spectrum is achieved by placing a penalty function on the undesirable frequency components of the field. The regular J_{penal} from Eq. (5) is replaced by a penalty term formulated in the frequency domain. In Ref. ([23], Sec. 3.1.2), it is shown that there is a close relationship between this formulation and the methods presented in Refs. [13–15].

The cosine transform is employed as a spectral tool. Other spectral transforms (i. e., the Fourier transform or the sine transform) could be used as well. For a typical signal, a cosine series is known to converge faster than a Fourier series or a sine series (see [23], Sec. 3.1.1).

The operation of the cosine transform on an arbitrary function $g(t)$ is denoted by the symbol $\mathcal{C}$, and the transformed function by $\bar{g}(\omega)$:

$$\bar{g}(\omega) \equiv \mathcal{C}[g(t)] \equiv \sqrt{\frac{2}{\pi}} \int_0^\infty g(t)\cos(\omega t)\,dt. \tag{10}$$

The inverse cosine transform will be denoted by $\mathcal{C}^{-1}$:

$$\mathcal{C}^{-1}[\bar{g}(\omega)] \equiv \sqrt{\frac{2}{\pi}} \int_0^\infty \bar{g}(\omega)\cos(\omega t)\,d\omega = g(t). \tag{11}$$

The driving field in the frequency domain is defined by a finite time cosine transform:

$$\bar{\epsilon}(\omega) = \sqrt{\frac{2}{\pi}} \int_0^T \epsilon(t)\cos(\omega t)\,dt. \tag{12}$$

The maximal cutoff driving frequency is denoted by Ω. The penalty term from Eq. (5) is modified to

$$J_{\text{penal}} \equiv -\alpha \int_0^\Omega \frac{1}{f_\epsilon(\omega)} \bar{\epsilon}^2(\omega)\,d\omega, \quad \alpha > 0 \tag{13}$$

where $f_\epsilon(\omega)$ is an adjustable function, which satisfies the conditions:

$$\int_0^\Omega f_\epsilon(\omega)\,d\omega = 1, \quad f_\epsilon(\omega) > 0. \tag{14}$$

α determines the cost of large fields; cf. Eq. (5). $f_\epsilon(\omega)$ is chosen to have small values for undesirable frequencies and regular values for the allowed frequency region. It may be interpreted as a filter function, as can be seen in Sec. III C. An additional envelope shape can be forced on the profile of the driving field spectrum by choosing an appropriate filter function. A complete filtration of undesirable frequencies is achieved in the limit $f_\epsilon(\omega) \longrightarrow 0$ for undesirable ω values. For practical purposes, $f_\epsilon(\omega)$ may be set to 0 for these values.

B. Optimization functional for harmonic generation

The field emitted by the system consists of the frequencies contained in the spectrum of the dipole expectation. In order to maximize the emission in a desired frequency region, the amplitude of the dipole moment oscillations in this frequency region is maximized. Thus, the physical quantity of interest is the dipole moment expectation value:

$$\langle\hat{\mu}\rangle(t) = \langle\psi(t)\,|\hat{\mu}|\,\psi(t)\rangle\,. \tag{15}$$

The dipole spectrum becomes

$$\overline{\langle\hat{\mu}\rangle}(\omega) = \mathcal{C}[\langle\hat{\mu}\rangle(t)] = \sqrt{\frac{2}{\pi}} \int_0^T \langle\hat{\mu}\rangle(t)\cos(\omega t)\,dt. \tag{16}$$

To maximize the emission in the desired region of the spectrum, the following functional is chosen:

$$J_{\max} \equiv \frac{1}{2}\lambda \int_0^\Omega f_\mu(\omega)\overline{\langle\hat{\mu}\rangle}^2(\omega)\,d\omega, \quad \lambda > 0 \tag{17}$$

where $f_\mu(\omega)$ satisfies the conditions:

$$\int_0^\Omega f_\mu(\omega)\,d\omega = 1, \quad f_\mu(\omega) \geqslant 0. \tag{18}$$

λ is an adjustable coefficient, which determines the relative importance of $J_{\max}$. λ is redundant with α, but from numerical stability considerations it is useful to vary it independently. $f_\mu(\omega)$ is a filter function, which has pronounced values in the frequency region of interest.

Equation (17) can be generalized to an arbitrary Hermitian operator $\hat{\mathbf{O}}$:

$$J_{\max} \equiv \frac{1}{2}\lambda \int_0^\Omega f_O(\omega)\overline{\langle\hat{\mathbf{O}}\rangle}^2(\omega)\,d\omega, \quad \lambda > 0 \tag{19}$$

$$\overline{\langle\hat{\mathbf{O}}\rangle}(\omega) = \mathcal{C}[\langle\hat{\mathbf{O}}\rangle(t)], \tag{20}$$

$$\int_0^\Omega f_O(\omega)\,d\omega = 1, \quad f_O(\omega) \geqslant 0. \tag{21}$$

One possible application of this generalization is in the case when it is desired to maximize emission with polarization other than that of the control field. For instance, if the control field is x polarized and we require the emission of the y polarized field, $\hat{\mu}$ in $\hat{\mathbf{H}}(t)$ is set to be $\hat{\mu}_x$, and $\hat{\mathbf{O}} \equiv \hat{\mu}_y$.

The full maximization functional for the harmonic generation problem becomes

$$J \equiv J_{\max} + J_{\text{penal}} + J_{\text{con}}, \tag{22}$$

$$J_{\max} \equiv \frac{1}{2}\lambda \int_0^\Omega f_O(\omega)\overline{\langle\hat{\mathbf{O}}\rangle}^2(\omega)\,d\omega, \tag{23}$$

$$J_{\text{penal}} \equiv -\alpha \int_0^\Omega \frac{1}{f_\epsilon(\omega)}\bar{\epsilon}^2(\omega)\,d\omega, \tag{24}$$

$$J_{\text{con}} \equiv -2\text{Re}\int_0^T \langle\chi(t)|\frac{\partial}{\partial t} + i\hat{\mathbf{H}}(t)|\psi(t)\rangle\,dt. \tag{25}$$

C. The Euler-Lagrange equations for harmonic generation

We choose the functional derivative of the objective in the frequency domain:

$$\frac{\delta J}{\delta\bar{\epsilon}(\omega)} = 0. \tag{26}$$

The resulting Euler-Lagrange equations become

$$\frac{\partial|\psi(t)\rangle}{\partial t} = -i\hat{\mathbf{H}}(t)|\psi(t)\rangle, \quad |\psi(0)\rangle = |\psi_0\rangle \tag{27}$$

$$\frac{\partial|\chi(t)\rangle}{\partial t} = -i\hat{\mathbf{H}}(t)|\chi(t)\rangle - \lambda\,\mathcal{C}^{-1}[f_O(\omega)\overline{\langle\hat{\mathbf{O}}\rangle}(\omega)]\hat{\mathbf{O}}|\psi(t)\rangle,$$
$$|\chi(T)\rangle = 0 \tag{28}$$

where $\hat{\mathbf{H}}(t) = \hat{\mathbf{H}}_0 - \hat{\mu}\epsilon(t)$ and

$$\bar{\epsilon}(\omega) = f_\epsilon(\omega)\mathcal{C}[\eta(t)], \quad \eta(t) \equiv -\frac{\text{Im}\langle\chi(t)\,|\hat{\mu}|\,\psi(t)\rangle}{\alpha} \tag{29}$$

$$\epsilon(t) = \mathcal{C}^{-1}[\bar{\epsilon}(\omega)] = \mathcal{C}^{-1}\left\{f_\epsilon(\omega)\mathcal{C}[\eta(t)]\right\}. \tag{30}$$

Note that the expression for $\eta(t)$ is the same as that for $\epsilon(t)$ in Eq. (9). $\epsilon(t)$ in Eq. (30) can be interpreted as the filtered field from the regular control problems, where $f_\epsilon(\omega)$ plays the role of a filter function. A comparison between Eqs. (8) and (28) leads to a similar interpretation of the inhomogeneous term in Eq. (28) (cf. [23], Sec. 3.3.1).

It is convenient to avoid normalizing $f_\epsilon(\omega)$ and $f_O(\omega)$, and substitute them with

$$\tilde{f}_\epsilon(\omega) \equiv \frac{f_\epsilon(\omega)}{\alpha}, \tag{31}$$

$$\tilde{f}_O(\omega) \equiv \lambda f_O(\omega). \tag{32}$$

D. Optional modifications for the optimization problem

1. *Prevention of dissociation*

Typically, when strong driving fields are employed the system dissociates or ionizes. This phenomenon can be avoided by restricting the system to localize in an "allowed" subspace of Hilbert space [22], for example, eliminating access to all eigenstates with energies above a threshold energy. Another option is to restrict the state vector to regions of space far from the threshold of the potential well. A similar idea is to restrict the dynamics to the allowed momentum values.

In order to restrict the system to the allowed subspace, two modifications in the maximization functional J are employed:

(1) $J_{\max}$ is modified to include the contribution only from the allowed states.

(2) A penalty term on the forbidden states is added to J.

The first modification is achieved by the replacement of the expectation $\langle\hat{\mathbf{O}}\rangle(t)$ in Eq. (20) by the expression:

$$\langle\hat{\mathbf{P}}_a\psi(t)|\hat{\mathbf{O}}|\hat{\mathbf{P}}_a\psi(t)\rangle,$$

where $\hat{\mathbf{P}}_a$ is the projection operator onto the allowed subspace. For instance, if all energies above the threshold E_L are restricted, then

$$\hat{\mathbf{P}}_a \equiv \sum_{n=0}^{L} |\varphi_n\rangle\langle\varphi_n|. \tag{33}$$

If the system is restricted in the x space to remain in the interval $[x_{\min}, x_{\max}]$, then:

$$\hat{\mathbf{P}}_a \equiv \int_{x_{\min}}^{x_{\max}} |x\rangle\langle x|\,dx. \tag{34}$$

If a smooth filtration of states is desired, $\hat{\mathbf{P}}_a$ may be generalized to a weighted projection operator, which will be denoted as $\hat{\mathbf{P}}^s_a$. For instance, $\hat{\mathbf{P}}_a$ from Eq. (33) is modified to

$$\hat{\mathbf{P}}^s_a \equiv \sum_{n=0}^{N-1} s_n|\varphi_n\rangle\langle\varphi_n|, \quad 0 \leqslant s_n \leqslant 1 \tag{35}$$

where s_n decreases gradually from 1 to 0 near the threshold. $\hat{\mathbf{P}}_a$ from Eq. (34) is modified to

$$\hat{\mathbf{P}}^s_a \equiv \int s(x)|x\rangle\langle x|\,dx = s(\hat{\mathbf{X}}), \quad 0 \leqslant s(x) \leqslant 1 \tag{36}$$

where $s(x)$ decays to 0 near the boundaries of the allowed x interval.

The resulting modified $J_{\max}$ becomes

$$J_{\max} \equiv \frac{1}{2}\int_0^\Omega \tilde{f}_O(\omega)\overline{\langle\hat{\mathbf{O}}_a\rangle}^2(\omega)\,d\omega, \quad \hat{\mathbf{O}}_a = \hat{\mathbf{P}}^s_a\hat{\mathbf{O}}\hat{\mathbf{P}}^s_a. \tag{37}$$

The second modification is the addition of the following penalty term to J (as suggested in Ref. [22]):

$$J_{\text{forb}} \equiv -\gamma\int_0^T \langle\psi(t)|\hat{\mathbf{P}}_f|\psi(t)\rangle\,dt, \quad \gamma > 0 \tag{38}$$

where $\hat{\mathbf{P}}_f$ is the projection onto the forbidden subspace, and γ is the penalty factor of the forbidden subspace.

It is possible to achieve a smooth filtration of states by using a state-dependent penalty factor. For instance, in the energy space, Eq. (38) is generalized to

$$J_{\text{forb}} \equiv -\int_0^T \langle\psi(t)|\hat{\mathbf{P}}^\gamma_f|\psi(t)\rangle\,dt, \tag{39}$$

$$\hat{\mathbf{P}}^\gamma_f \equiv \sum_{n=0}^{N-1}\gamma_n|\varphi_n\rangle\langle\varphi_n|, \quad \gamma_n \geqslant 0. \tag{40}$$

$\hat{\mathbf{P}}_f^{\gamma}$ is a skewed projection. γ_n is the penalty factor of the state $|\varphi_n\rangle$. It should be 0 for the allowed energy domain and increase gradually with n near the threshold energy. In the x space the skewed projection is

$$\hat{\mathbf{P}}_f^{\gamma} \equiv \int \gamma(x)|x\rangle\langle x|\,dx = \gamma(\hat{\mathbf{X}}), \quad \gamma(x) \geqslant 0 \tag{41}$$

where $\gamma(x)$ increases gradually in the forbidden x regions. A smooth penalization is recommended in order to decrease the difficulty in the optimization process.

When the Hilbert space is very large, it becomes impractical to compute all the eigenstates. Nevertheless, a restriction of the allowed subspace in the energy space is still possible. s_n and γ_n should be defined as functions of the energy, i.e.,

$$s_n = s(E_n), \quad \gamma_n = \gamma(E_n).$$

Then we have

$$\hat{\mathbf{P}}_a^{s} = s(\hat{\mathbf{H}}_0), \quad \hat{\mathbf{P}}_f^{\gamma} = \gamma(\hat{\mathbf{H}}_0). \tag{42}$$

$s(\hat{\mathbf{H}}_0)|\psi(t)\rangle$ and $\gamma(\hat{\mathbf{H}}_0)|\psi(t)\rangle$ can be approximated using standard methods.

When $\hat{\mathbf{P}}_a^{s} = \hat{\mathbf{I}}$ and $\hat{\mathbf{P}}_f^{\gamma} = \hat{\mathbf{0}}$, J reduces to Eq. (22).

After inserting these changes in J, the equation for $|\chi(t)\rangle$ [Eq. (28)] is modified in the following way:

$$\frac{\partial|\chi(t)\rangle}{\partial t} = -i\hat{\mathbf{H}}(t)|\chi(t)\rangle - \left\{\mathcal{C}^{-1}[\tilde{f}_O(\omega)\overline{\langle\hat{\mathbf{O}}_a\rangle}(\omega)]\hat{\mathbf{O}}_a - \hat{\mathbf{P}}_f^{\gamma}\right\}|\psi(t)\rangle. \tag{43}$$

2. *Prevention of boundary effects*

In practice, a finite time spectral transform is approximated by a discrete series. Care should be taken on possible boundary effects; otherwise noise ("ringing") throughout the spectral representation of the signal is generated. For a cosine representation this effect is relatively small. If necessary, it is possible to reduce this phenomenon by trying to enforce the boundary conditions. The appropriate boundary conditions for a cosine series representation of $\langle\hat{\mathbf{O}}\rangle(t)$ are

$$\frac{d\langle\hat{\mathbf{O}}\rangle(0)}{dt} = 0, \quad \frac{d\langle\hat{\mathbf{O}}\rangle(T)}{dt} = 0. \tag{44}$$

Usually, the condition at $t = 0$ is automatically satisfied because the initial state is typically chosen to be the ground state. The condition at $t = T$ may be enforced by an addition of the following penalty term to J:

$$J_{\text{bound}} \equiv -\frac{1}{2}\kappa\left[\frac{d\langle\hat{\mathbf{O}}\rangle(T)}{dt}\right]^2, \quad \kappa \geqslant 0. \tag{45}$$

κ is an adjustable parameter. When $\kappa = 0$, J reduces to Eq. (22).

In the special case that $[\hat{\mu},\hat{\mathbf{O}}] = \hat{\mathbf{0}}$, the insertion of J_{bound} results in a relatively simple modification of the Euler-Lagrange equations. The natural boundary condition for $|\chi(t)\rangle$ in Eq. (28) is replaced by

$$|\chi(T)\rangle = \kappa\langle[\hat{\mathbf{H}}_0,\hat{\mathbf{O}}]\rangle(T)[\hat{\mathbf{H}}_0,\hat{\mathbf{O}}]|\psi(T)\rangle. \tag{46}$$

The general case is more complex and will not be discussed here.

The derivation of the Euler-Lagrange equations is presented in Appendix A.

E. Optimization procedure

For current optimization posed both in time and frequency, the well established optimization procedures do not converge monotonically or converge very slowly. We therefore employed a more direct *relaxation method* to update the driving field from iteration to iteration. In the present context the method consists of the following update rule for the driving field:

$$\bar{\epsilon}^{\text{new}}(\omega) = K\bar{\epsilon}^{EL}(\omega) + (1-K)\bar{\epsilon}^{\text{old}}(\omega), \quad 0 < K \leqslant 1 \tag{47}$$

$$\bar{\epsilon}^{EL}(\omega) \equiv \tilde{f}_{\epsilon}(\omega)\mathcal{C}[-\text{Im}\langle\chi(t)|\hat{\mu}|\psi(t)\rangle|_{\bar{\epsilon}(\omega)=\bar{\epsilon}^{\text{old}}(\omega)}]. \tag{48}$$

The updated field is a mixture of the previous field and the field computed from the Euler-Lagrange equation [Eq. (29)], using the previous field for the computation of $|\chi(t)\rangle$ and $|\psi(t)\rangle$. K is the mixing parameter, which determines the weights of the two fields. The value of K is decreased when the optimization process progresses.

The scheme for the implementation of the relaxation method becomes

(1) Guess a driving field spectrum $\bar{\epsilon}^{(0)}(\omega)$.
(2) Set $\epsilon^{(0)}(t) = \mathcal{C}^{-1}[\bar{\epsilon}^{(0)}(\omega)]$.
(3) Guess an initial value for K.
(4) Propagate $|\psi^{(0)}(t)\rangle$ forward from $t = 0$ to $t = T$ according to Eq. (27), with $\epsilon^{(0)}(t)$.
(5) Calculate $J^{(0)}$ with $|\psi^{(0)}(t)\rangle$ and $\bar{\epsilon}^{(0)}(\omega)$.
(6) k = 0.
(7) Repeat the following steps until convergence:
  (a) Set $|\chi^{(k)}(T)\rangle$ according to Eq. (46), using $|\psi^{(k)}(T)\rangle$.
  (b) Propagate $|\chi^{(k)}(t)\rangle$ backward from $t = T$ to $t = 0$ according to Eq. (43), with $\epsilon^{(k)}(t)$.
  (c) Do the following steps, and repeat while $J^{\text{trial}} \leqslant J^{(k)}$:
    (i) Set a new field, using Eq. (47):

$$\bar{\epsilon}^{\text{trial}}(\omega) = K\tilde{f}_{\epsilon}(\omega)\mathcal{C}[-\text{Im}\langle\chi^{(k)}(t)|\hat{\mu}|\psi^{(k)}(t)\rangle] + (1-K)\bar{\epsilon}^{(k)}(\omega),$$
$$\epsilon^{\text{trial}}(t) = \mathcal{C}^{-1}[\bar{\epsilon}^{\text{trial}}(\omega)].$$

    (ii) Propagate $|\psi^{\text{trial}}(t)\rangle$ forward from $t = 0$ to $t = T$ according to Eq. (27), with $\epsilon^{\text{trial}}(t)$.
    (iii) Calculate J^{trial} with $|\psi^{\text{trial}}(t)\rangle$ and $\bar{\epsilon}^{\text{trial}}(\omega)$.
    (iv) If $J^{\text{trial}} \leqslant J^{(k)}$, then set $K = K/2$
  (d) Update all the variables:

$$\bar{\epsilon}^{(k+1)}(\omega) = \bar{\epsilon}^{\text{trial}}(\omega), \quad \epsilon^{(k+1)}(t) = \epsilon^{\text{trial}}(t),$$
$$|\psi^{(k+1)}(t)\rangle = |\psi^{\text{trial}}(t)\rangle, \quad J^{(k+1)} = J^{\text{trial}}.$$

  (e) $k = k + 1$.

It can be shown (see [23], Sec. 3.2.3) that the relaxation method, in the context of quantum-OCT problems, can be considered an approximated second order gradient method

TABLE I. Description of the notations in the tables.

Notation	Description
$\hat{\mathbf{H}}_0$	Unperturbed Hamiltonian
$\hat{\mu}$	Dipole moment operator
$\|\psi_0\rangle$	Initial state vector
T	Final time
$\tilde{f}_\epsilon(\omega)$	Scaled filter function of the driving field
$\tilde{f}_\mu(\omega)$	Scaled filter function of the dipole moment expectation value
$\bar{\epsilon}^0(\omega)$	Initial guess of the field
L	Index of the maximal allowed eigenstate
γ_n	Penalty factor of the state $\|\varphi_n\rangle$
$s(x)$	Projection function onto allowed x regions
$\gamma(x)$	Penalty function on forbidden x regions
K_i	Initial guess of K, for the relaxation method
x domain	Domain of the x grid
N_x	Number of equidistant points in the x grid
τ	Tolerance parameter of the optimization process

(quasi-Newton method), where the Hessian of J is approximated by the Hessian of J_{penal}.

IV. APPLICATION

Our method is demonstrated in four simple harmonic generation examples. The propagator for the Schrödinger equation is based on an efficient and highly accurate algorithm recently published [24].

The convergence condition of the optimization procedure is

$$\frac{\|\vec{\bar{\epsilon}}^{\,\text{new}} - \vec{\bar{\epsilon}}^{\,\text{old}}\|}{\|\vec{\bar{\epsilon}}^{\,\text{new}}\|} < \tau, \tag{49}$$

where $\vec{\bar{\epsilon}}$ is the discrete vector of frequency values on an equidistant ω grid and τ is a tolerance parameter. More numerical details may be found in Ref. ([23], Appendix B).

The important details of the problems and the computational process are presented in the tables. Atomic units are used throughout. The notations in the tables are described in Table I. $\Theta(x)$ denotes the Heaviside step function. The initial state in all problems is chosen to be the ground state, denoted as $|\varphi_0\rangle$.

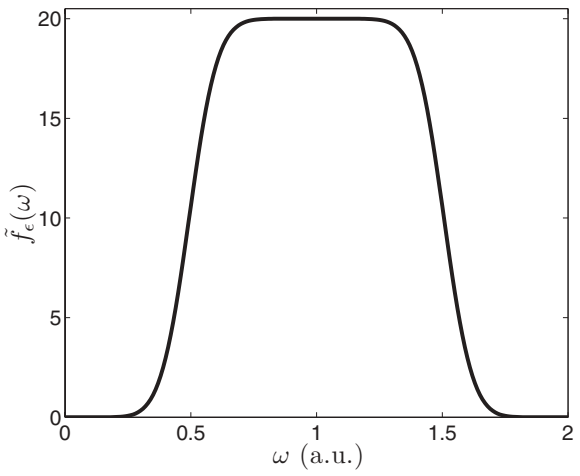

FIG. 1. $\tilde{f}_\epsilon(\omega)$ of the TLS problem.

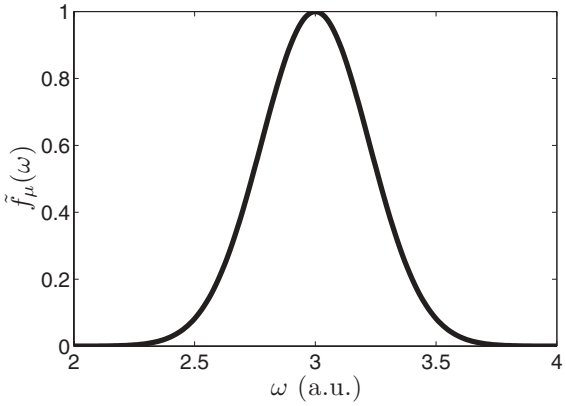

FIG. 2. $\tilde{f}_\mu(\omega)$ of the TLS problem.

A. Two level system

The first problem is tripling the driving frequency by a two level system (TLS). The unperturbed Hamiltonian is

$$\hat{\mathbf{H}}_0 = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}. \tag{50}$$

The dipole moment operator is chosen to be the x Pauli matrix:

$$\hat{\mu} = \sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \tag{51}$$

The driving field is restricted to be centered around $\omega = 1$ a.u., by a "hat" filter function (cf. Fig. 1). We require maximization of the emission in the region of the characteristic frequency of the system, $\omega_{1,0} = 3$ a.u. A Gaussian function is used for $\tilde{f}_\mu(\omega)$ (see Fig. 2).

The details of the problem are summarized in Table II.

The optimization process converges rapidly to a solution. The convergence curve is shown in Fig. 3.

The resulting spectra of the driving field and the dipole moment expectation value are shown in Fig. 4. $\bar{\epsilon}(\omega)$ is shown to be successfully restricted to the desired portion of the

TABLE II. The details of the TLS problem.

$\hat{\mathbf{H}}_0$	$\begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}$
$\hat{\mu}$	$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$
$\|\psi_0\rangle$	$\begin{bmatrix} 1 \\ 0 \end{bmatrix}$
T	100
$\tilde{f}_\epsilon(\omega)$	$20\,\text{sech}[20(\omega-1)^4]$
$\tilde{f}_\mu(\omega)$	$\exp[-10(\omega-3)^2]$
$\bar{\epsilon}^0(\omega)$	$\text{sech}[20(\omega-1)^4]$
K_i	0.5
τ	10^{-3}

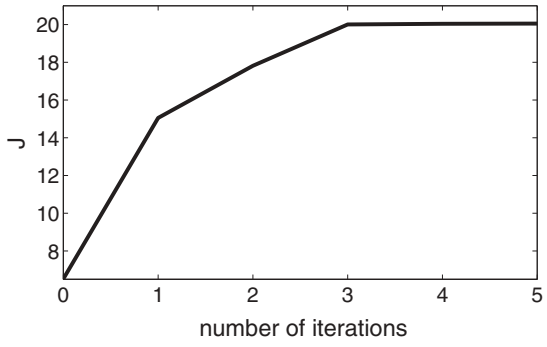

FIG. 3. The convergence curve of the TLS problem.

spectrum. The hat envelope shape is apparent. $\overline{\langle\hat{\mu}\rangle}(\omega)$ mainly consists of a large peak at $\omega_{1,0}$, as required.

B. Eleven level system

The second problem is of an eleven level system (11LS). The problem is designed for harmonic generation by a resonance mediated absorption mechanism (for example, see [25]).

The unperturbed Hamiltonian is

$$\hat{\mathbf{H}}_0 = \begin{bmatrix} 1 & & & & & & & & & & \\ & 2.1 & & & & & & & & & \\ & & 3 & & & & & & & & \\ & & & 3.9 & & & & & 0 & & \\ & & & & 5 & & & & & & \\ & & & & & 6.1 & & & & & \\ & & & & & & 7 & & & & \\ & & 0 & & & & & 8.1 & & & \\ & & & & & & & & 9 & & \\ & & & & & & & & & 9.9 & \\ & & & & & & & & & & 11 \end{bmatrix}. \tag{52}$$

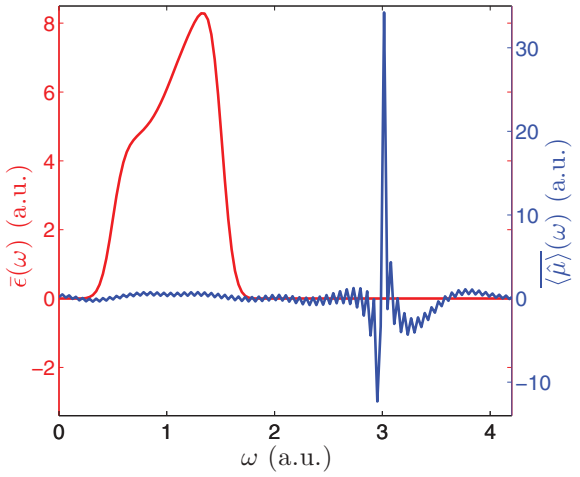

FIG. 4. (Color online) The spectra of the driving field $\bar{\epsilon}(\omega)$ (red, gray) and the dipole moment expectation spectra $\overline{\langle\hat{\mu}\rangle}(\omega)$ (blue, dark gray) for the TLS problem.

TABLE III. The details of the 11LS problem.

$\hat{\mathbf{H}}_0$	Eq. (52)
$\hat{\mu}$	Eq. (53)
$\|\psi_0\rangle$	$\|\varphi_0\rangle$
T	100
$\tilde{f}_\epsilon(\omega)$	$50\,\Theta(1.3-\omega)$
$\tilde{f}_\mu(\omega)$	$\Theta(\omega-9.9)\,\Theta(10.1-\omega)$
$\bar{\epsilon}^0(\omega)$	$\Theta(1.3-\omega)$
K_i	1
τ	10^{-3}

The dipole moment operator is

$$\hat{\mu} = \begin{bmatrix} 0 & 1 & & & & & & & & & 1 \\ 1 & 0 & 1 & & & & & & & & \\ & 1 & 0 & 1 & & & & & & & \\ & & 1 & 0 & 1 & & & & & & \\ & & & 1 & 0 & 1 & & & 0 & & \\ & & & & 1 & 0 & 1 & & & & \\ & & 0 & & & 1 & 0 & 1 & & & \\ & & & & & & 1 & 0 & 1 & & \\ & & & & & & & 1 & 0 & 1 & \\ & & & & & & & & 1 & 0 & \\ 1 & & & & & & & & & & \end{bmatrix}. \tag{53}$$

This $\hat{\mu}$ couples between neighboring eigenstates and between the outer eigenstates, $|\varphi_0\rangle$ and $|\varphi_{10}\rangle$. The driving field is restricted so as not to exceed the region of the resonance frequencies of the neighboring levels, $\omega_{n+1,n} = 1$ a.u. We require an emission in the neighborhood of the Bohr frequency of the outer energy levels, $\omega_{10,0} = 10$ a.u. $\tilde{f}_\epsilon(\omega)$ and $\tilde{f}_\mu(\omega)$ are chosen to be rectangular functions.

The details of the problem are summarized in Table III.

The convergence curve is shown in Fig. 5. The resulting $\bar{\epsilon}(\omega)$ and $\overline{\langle\hat{\mu}\rangle}(\omega)$ are shown in Fig. 6. Our method is shown again to be quite efficient in maximizing the emission in the required region.

C. Anharmonic oscillator—the HCl molecule

In this example an anharmonic oscillator is used to double the incoming frequency. The oscillator chosen is an

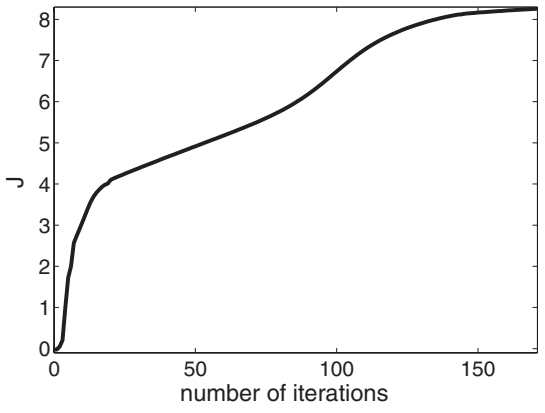

FIG. 5. The convergence curve of the 11LS problem.

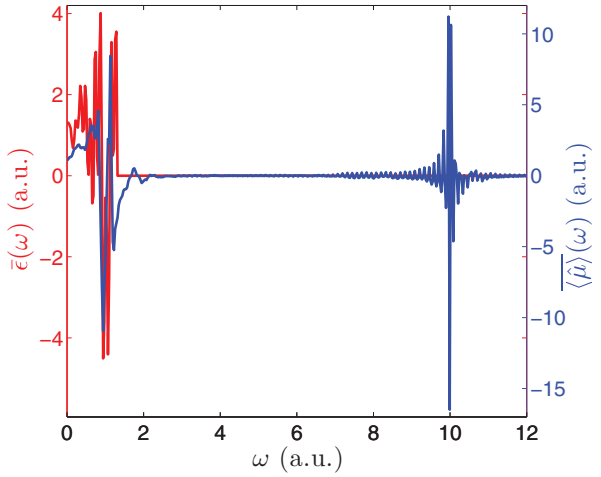

FIG. 6. (Color online) The spectra of the driving field $\bar{\epsilon}(\omega)$ (red, gray) and the dipole moment expectation value $\overline{\langle\hat{\mu}\rangle}(\omega)$ (blue, dark gray) for the 11LS problem.

approximation of the H ^{35}Cl molecule (see Appendix B). As in the previous problem, the intended mechanism is of harmonic generation by resonance mediated absorption.

The coordinate of the one-dimensional oscillator is the displacement of the internuclei distance, r_{H-Cl}, from the bottom of the well (r^*):

$$x = r_{H-Cl} - r^*. \tag{54}$$

The approximated potential function $V(x)$ and dipole moment function $\mu(x)$ are shown in Fig. 7.

$\bar{\epsilon}(\omega)$ is restricted not to exceed much the characteristic frequency of the bottom of the well:

$$\omega_0 = 1.35 \times 10^{-2} \text{ a.u.}$$

We require maximization of the emission in the neighborhood of the second harmonic:

$$\omega_{2,0} = E_2 - E_0 = 2.54 \times 10^{-2} \text{ a.u.}$$

$\tilde{f}_\epsilon(\omega)$ and $\tilde{f}_\mu(\omega)$ are chosen to be rectangular functions.

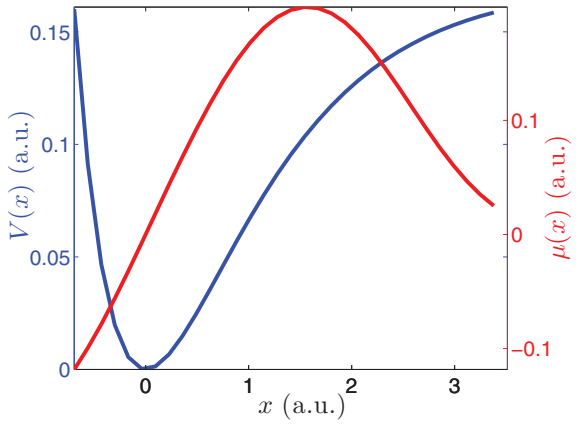

FIG. 7. (Color online) The approximated potential $V(x)$ (blue, dark gray) and dipole function $\mu(x)$ (red, gray) curves for the HCl molecule.

TABLE IV. The details of the HCl problem.

$\hat{\mathbf{H}}_0$	$\frac{\hat{\mathbf{P}}^2}{2\cdot 1785} + 0.171[\exp(-0.975\,\hat{\mathbf{X}}) - \hat{\mathbf{I}}]^2$
$\hat{\mu}$	$(0.19309\,\hat{\mathbf{X}})\{\hat{\mathbf{I}} - \mathrm{Re}[\tanh((0.17069 + 0.056854\,i)(\hat{\mathbf{X}} - 0.10630\,\hat{\mathbf{I}})^{1.8977})]\}$
$\lvert\psi_0\rangle$	$\lvert\varphi_0\rangle$
T	10^4
$\tilde{f}_\epsilon(\omega)$	$2500\,\Theta(0.015-\omega)$
$\tilde{f}_\mu(\omega)$	$100\,\Theta(\omega-0.025)\,\Theta(0.027-\omega)$
L	19
γ_n	$\begin{cases} 0 & n \leqslant 19 \\ (n-19)^2 & n > 19 \end{cases}$
$\bar{\epsilon}^0(\omega)$	$\Theta(0.015-\omega)$
K_i	1
x domain	$[-0.69407, 3.51178)$
N_x	32
τ	10^{-3}

In order to prevent dissociation of the molecule, the energies above the dissociation threshold were restricted (see Sec. III D1).

The details of the problem are summarized in Table IV.

The convergence curve is shown in Fig. 8. The resulting $\bar{\epsilon}(\omega)$ and $\overline{\langle\hat{\mu}\rangle}(\omega)$ are shown in Fig. 9. $\overline{\langle\hat{\mu}\rangle}(\omega)$ mainly consists of a large linear response to the driving field, as could be expected. However, there is also a significant nonlinear response in the region of the second harmonic frequency, as required.

A detailed analysis of the results of the first three examples may be found in Ref. ([23], Chap. 4).

D. One-dimensional particle in a truncated Coulomb potential

The last example demonstrates the application of our method in a system with stronger nonlinearity. A driven electron in a Coulomb potential is the system studied. The model has been extensively studied in the context of harmonic generation [4,5]. Typically, for strong driving fields a comb of odd frequencies is generated up to a cutoff. In the present example the target is the emission of a single high harmonic of the driving frequency. This target has similarities to the experiment in the Joint Institute for Laboratory Astrophysics

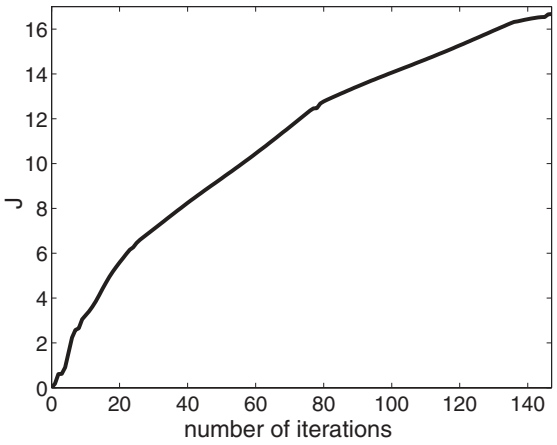

FIG. 8. The convergence curve of the HCl problem.

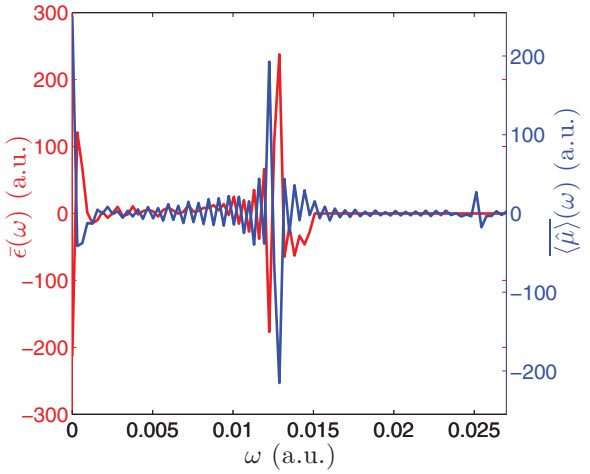

FIG. 9. (Color online) The spectra of the driving field $\bar{\epsilon}(\omega)$ (red, gray) and the dipole expectation spectrum $\overline{\langle\hat{\mu}\rangle}(\omega)$ (blue, dark gray) for the HCl problem. Two significant peaks appear: A large linear response of the dipole to the driving field, and a significant nonlinear response in the neighborhood of the second harmonic. Notice the low frequency component of the driving field which changes the static part of the Hamiltonian.

(JILA) [26] where the emission of the 27th harmonic was enhanced relative to its neighbors using a genetic algorithm optimization.

Our model consists of a particle of unit mass and charge placed in a truncated Coulomb potential constrained to one dimension (see Fig. 10):

$$V(x) = 1 - \frac{1}{\sqrt{x^2+1}}. \tag{55}$$

The dipole operator is $\hat{\mu} = \hat{\mathbf{X}}$.

The driving field is restricted to frequencies which are much lower than the resonance frequencies of the system. $\bar{\epsilon}(\omega)$ is restricted so as not to exceed the region of $\omega = 0.07$ a.u. The shape of $\tilde{f}_\epsilon(\omega)$ (see Fig. 11) induces a smooth filtration of higher frequencies. We require maximization of the emission

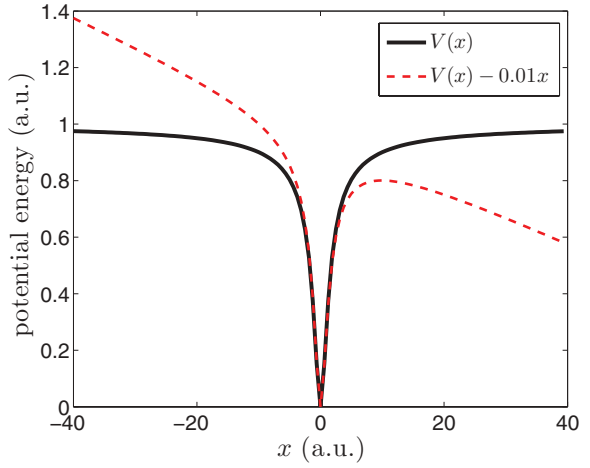

FIG. 10. (Color online) The truncated Coulomb potential [Eq. (55), solid black] and the potential energy of the system under the influence of a strong field (dashed red).

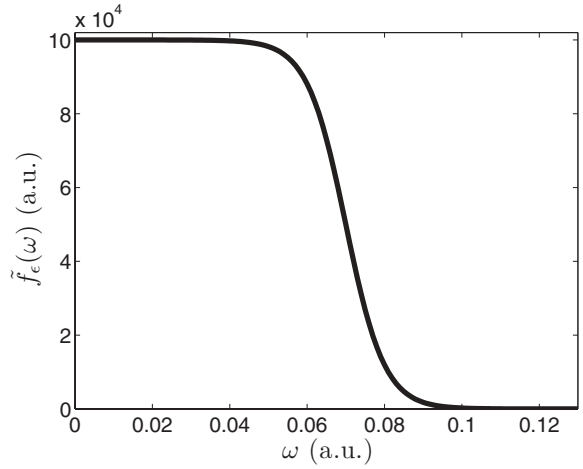

FIG. 11. $\tilde{f}_\epsilon(\omega)$ of the truncated Coulomb potential problem.

in the region of one of the Bohr frequencies of the system, $\omega_{5,0} = 0.624$ a.u. For the filter $\tilde{f}_\mu(\omega)$ a rectangular function is employed.

The edges of the x grid are restricted using the method presented in Sec. III D1. $s(x)$ and $\gamma(x)$ are shown in Fig. 12.

The details of the problem are summarized in Table V.

The convergence curve is shown in Fig. 13.

The resulting field $\bar{\epsilon}(\omega)$ and dipole $\overline{\langle\hat{\mu}\rangle}(\omega)$ are shown in Fig. 14. The largest peak of $\overline{\langle\hat{\mu}\rangle}(\omega)$ is located near the fundamental frequency of the system, $\omega_{1,0} = 0.395$ a.u. The response in the desired frequency is marked on the figure. The method is shown to be effective also in the production of frequencies considerably higher than those of the driving field.

The final solution obtained up to the chosen τ is not a converged solution, as can be deduced from the convergence curve shape at the end of the optimization process (see Fig. 13). If the optimization process is carried on to smaller τ, contributions to $\overline{\langle\hat{\mu}\rangle}(\omega)$ from interactions with the boundaries of the x grid begin to be significant. Nevertheless, a converged solution which does not involve such undesirable effects can be obtained. This is achieved by the increment of the $\gamma(x)$ values during the optimization process before the spurious effects appear.

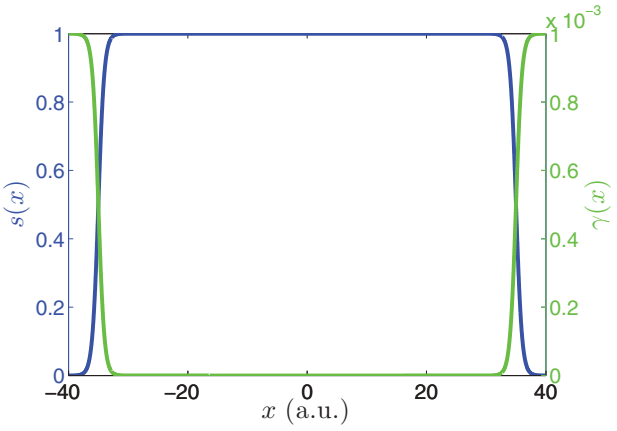

FIG. 12. (Color online) $s(x)$ (blue, dark gray) and $\gamma(x)$ (green, light gray) of the truncated Coulomb potential problem.

TABLE V. The details of the truncated Coulomb potential problem.

$\hat{\mathbf{H}}_0$	$\frac{\hat{\mathbf{P}}^2}{2} + \hat{\mathbf{I}} - \frac{1}{\sqrt{\hat{\mathbf{X}}^2+\hat{\mathbf{I}}}}$
$\hat{\mu}$	$\hat{\mathbf{X}}$
$\|\psi_0\rangle$	$\|\varphi_0\rangle$
T	2000
$\tilde{f}_\epsilon(\omega)$	$5 \times 10^4 \{1 - \tanh[100(\omega - 0.07)]\}$
$\tilde{f}_\mu(\omega)$	$\Theta(\omega - 0.61)\,\Theta(0.63 - \omega)$
$s(x)$	$0.5\,[\tanh(x + 35) - \tanh(x - 35)]$
$\gamma(x)$	$10^{-3}\,[1 - s(x)]$
$\bar{\epsilon}^0(\omega)$	$0.3\,\{1 - \tanh[100(\omega - 0.07)]\}$
K_i	10^{-6}
x domain	$[-40, 40)$
N_x	128
τ	5×10^{-4}

Maximizing the response of an arbitrary frequency, which is not one of the Bohr frequencies, was also achieved. However, the resulting amplitude of the response was found to be considerably smaller than the response for a Bohr frequency. Larger response was obtained when partial ionization was allowed. Technically, this requires absorbing boundary conditions. This topic is still under investigation, and is therefore not presented.

V. CONCLUSION

Optimizing harmonic generation is one of the most difficult tasks in the context of quantum control. A major obstacle is that the target objective cannot be formulated in the time domain. Additional restrictions have to be added to suppress ionization or dissociation. A theoretical method of calculation for optimal control of harmonic generation was studied. The task was addressed by the means of OCT using a frequency domain formulation. The relaxation method was used as the iterative optimization procedure.

For low and intermediate control fields, fast convergence was obtained when the emitted high harmonic fit a fundamental Bohr frequency of the system. Stronger fields can modify the system thus allowing harmonic emission in other frequencies.

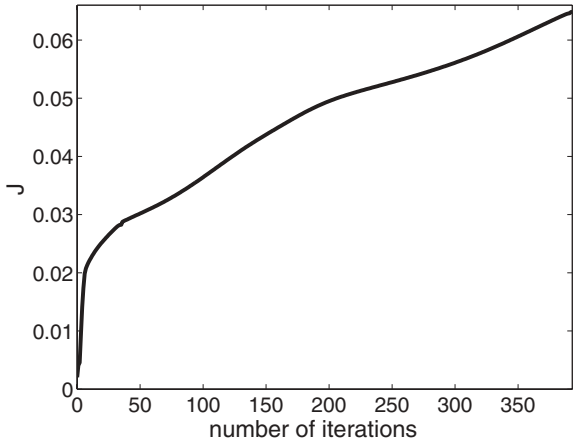

FIG. 13. The convergence curve of the truncated Coulomb potential problem.

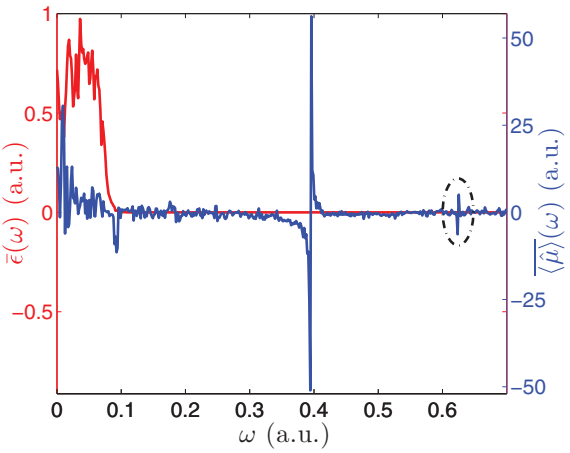

FIG. 14. (Color online) The spectra of the driving field $\bar{\epsilon}(\omega)$ (red, gray) and the dipole expectation spectrum $\overline{\langle\hat{\mu}\rangle}(\omega)$ (blue, dark gray) for the truncated Coulomb potential problem; the response in the desired frequency, $\omega_{5,0} = 0.624$ a.u., is marked by a black ellipse. The largest peak of $\overline{\langle\hat{\mu}\rangle}(\omega)$ is in the fundamental frequency, $\omega_{1,0} = 0.395$ a.u.

The difficulty we found in locating such solutions was that they competed with dissociation or ionization. Thus, the method should be modified to allow partial dissociation or ionization.

This paper focuses on the control aspect of our method. However, the physical interpretation of the results is of interest. Significant physical insight can be deduced from the optimized fields, unraveling new harmonic generation mechanisms, as was demonstrated in Ref. ([23], Chap. 4). Further studies employing the current approach will contribute to the understanding of harmonic generation processes, in particular nonadiabatic mechanisms which go beyond the three-step model [3].

ACKNOWLEDGMENTS

We thank Christiane Koch, Hardy Gross, and Nimrod Moiseyev for helpful discussions and criticism. We gratefully acknowledge financial support from the Israel Science Foundation. The Fritz Haber Center is supported by the Minerva Gesellschaft für die Forschung GmbH, München, Germany.

APPENDIX A: THE DERIVATION OF THE EULER-LAGRANGE EQUATIONS

The general maximization functional is rewritten in its full form, for convenience:

$$J \equiv J_{\max} + J_{\text{bound}} + J_{\text{forb}} + J_{\text{penal}} + J_{\text{con}}, \tag{A1}$$

$$J_{\max} \equiv \frac{1}{2}\int_0^\Omega \tilde{f}_O(\omega)\overline{\langle\hat{\mathbf{O}}_a\rangle}^2(\omega)\,d\omega, \quad \tilde{f}_O(\omega) \geqslant 0 \tag{A2}$$

$$\overline{\langle\hat{\mathbf{O}}_a\rangle}(\omega) \equiv \sqrt{\frac{2}{\pi}}\int_0^T \langle\hat{\mathbf{O}}_a\rangle(t)\cos(\omega t)\,dt, \tag{A3}$$

$$J_{\text{bound}} \equiv -\frac{1}{2}\kappa\left[\frac{d\langle\hat{\mathbf{O}}\rangle(T)}{dt}\right]^2, \quad \kappa \geqslant 0 \tag{A4}$$

$$J_{\text{forb}} \equiv -\int_0^T \langle\psi(t)|\hat{\mathbf{P}}_f^\gamma|\psi(t)\rangle\,dt, \tag{A5}$$

$$J_{\text{penal}} \equiv -\int_0^{\Omega} \frac{1}{\tilde{f}_{\epsilon}(\omega)} \bar{\epsilon}^2(\omega)\, d\omega, \quad \tilde{f}_{\epsilon}(\omega) > 0 \tag{A6}$$

$$\bar{\epsilon}(\omega) \equiv \sqrt{\frac{2}{\pi}} \int_0^{T} \epsilon(t) \cos(\omega t)\, dt, \tag{A7}$$

$$J_{\text{con}} \equiv -2\text{Re} \int_0^{T} \langle \chi(t)| \frac{\partial}{\partial t} + i\hat{\mathbf{H}}(t)|\psi(t)\rangle\, dt, \tag{A8}$$

$$\hat{\mathbf{H}}(t) = \hat{\mathbf{H}}_0 - \hat{\mu}\epsilon(t) = \hat{\mathbf{H}}_0 - \hat{\mu}\left(\sqrt{\frac{2}{\pi}} \int_0^{\Omega} \bar{\epsilon}(\omega) \cos(\omega t)\, d\omega\right). \tag{A9}$$

The constraint equations are

$$\frac{\partial |\psi(t)\rangle}{\partial t} = -i\hat{\mathbf{H}}(t)|\psi(t)\rangle, \quad |\psi(0)\rangle = |\psi_0\rangle \tag{A10}$$

$$\frac{\partial \langle\psi(t)|}{\partial t} = i \langle\psi(t)| \hat{\mathbf{H}}(t), \quad \langle\psi(0)| = \langle\psi_0|. \tag{A11}$$

Equations (A10) and (A11) ensure that

$$\langle\psi(t)| = |\psi(t)\rangle^{\dagger}.$$

Assuming this, all the computations can be performed using (A10) only.

The extremum conditions are

$$\frac{\delta J}{\delta \bar{\epsilon}(\omega)} = 0, \tag{A12}$$

$$\frac{\delta J}{\delta |\psi(t)\rangle} = 0, \tag{A13}$$

$$\frac{\delta J}{\delta \langle\psi(t)|} = 0, \tag{A14}$$

$$\frac{\delta J}{\delta |\psi(T)\rangle} = 0, \tag{A15}$$

$$\frac{\delta J}{\delta \langle\psi(T)|} = 0. \tag{A16}$$

After integrating by parts the following expression in J_{con}:

$$\int_0^{T} \left\langle \chi(t) \middle| \frac{\partial \psi(t)}{\partial t} \right\rangle dt,$$

we obtain

$$J_{\text{con}} = -2\text{Re}\Bigg[\langle\chi(T)|\psi(T)\rangle - \langle\chi(0)|\psi(0)\rangle - \int_0^{T} \left\langle \left(\frac{\partial}{\partial t} + i\hat{\mathbf{H}}(t)\right)\chi(t) \middle| \psi(t) \right\rangle dt\Bigg]. \tag{A17}$$

For simplicity, we will assume that J_{bound} has no explicit dependence on $\bar{\epsilon}(\omega)$. This requires that $[\hat{\mu}, \hat{\mathbf{O}}] = \hat{\mathbf{0}}$ [see Eq. (A31)], or that $\kappa = 0$. The expression for the left-hand side (LHS) of (A12) is obtained using (A6), (A17), and (A9):

$$\frac{\delta J}{\delta \bar{\epsilon}(\omega)} = \frac{\delta J_{\text{penal}}}{\delta \bar{\epsilon}(\omega)} + \frac{\delta J_{\text{con}}}{\delta \bar{\epsilon}(\omega)}, \tag{A18}$$

$$\frac{\delta J_{\text{penal}}}{\delta \bar{\epsilon}(\omega)} = -\frac{2}{\tilde{f}_{\epsilon}(\omega)} \bar{\epsilon}(\omega), \tag{A19}$$

$$\begin{aligned}\frac{\delta J_{\text{con}}}{\delta \bar{\epsilon}(\omega)} &= 2\,\text{Re}\left[-i \int_0^{T} \langle\chi(t)| \frac{\delta \hat{\mathbf{H}}(t)}{\delta \bar{\epsilon}(\omega)} |\psi(t)\rangle\, dt\right] \\ &= -2\,\text{Im}\left[\sqrt{\frac{2}{\pi}} \int_0^{T} \langle\chi(t)|\hat{\mu}|\psi(t)\rangle \cos(\omega t)\, dt\right] \\ &= -2\,\text{Im}\{\mathcal{C}[\langle\chi(t)|\hat{\mu}|\psi(t)\rangle]\}.\end{aligned} \tag{A20}$$

From (A12) and (A18)–(A20), we obtain the following expression for $\bar{\epsilon}(\omega)$:

$$\bar{\epsilon}(\omega) = \tilde{f}_{\epsilon}(\omega)\mathcal{C}[-\text{Im}\langle\chi(t)|\hat{\mu}|\psi(t)\rangle]. \tag{A21}$$

In order to derive the LHS of (A13), we first write the explicit expression of $J_{\max}$ as a functional of $|\psi(t)\rangle$:

$$J_{\max} = \frac{1}{\pi} \int_0^{\Omega} \int_0^{T} \int_0^{T} \tilde{f}_O(\omega) \langle\psi(t)|\hat{\mathbf{O}}_a|\psi(t)\rangle \langle\psi(t')|\hat{\mathbf{O}}_a|\psi(t')\rangle \times \cos(\omega t) \cos(\omega t')\, dt\, dt'\, d\omega. \tag{A22}$$

The expression for the LHS of (A13) is obtained using (A22), (A5), and (A17):

$$\frac{\delta J}{\delta |\psi(t)\rangle} = \frac{\delta J_{\max}}{\delta |\psi(t)\rangle} + \frac{\delta J_{\text{forb}}}{\delta |\psi(t)\rangle} + \frac{\delta J_{\text{con}}}{\delta |\psi(t)\rangle}, \tag{A23}$$

$$\begin{aligned}\frac{\delta J_{\max}}{\delta |\psi(t)\rangle} &= \frac{2}{\pi} \int_0^{\Omega} \int_0^{T} \tilde{f}_O(\omega) \langle\psi(t)|\hat{\mathbf{O}}_a \langle\psi(t')|\hat{\mathbf{O}}_a|\psi(t')\rangle \times \cos(\omega t) \cos(\omega t')\, dt'\, d\omega \\ &= \sqrt{\frac{2}{\pi}} \int_0^{\Omega} \tilde{f}_O(\omega) \overline{\langle\hat{\mathbf{O}}_a\rangle}(\omega) \cos(\omega t)\, d\omega \langle\psi(t)|\hat{\mathbf{O}}_a \\ &= \mathcal{C}^{-1}[\tilde{f}_O(\omega)\overline{\langle\hat{\mathbf{O}}_a\rangle}(\omega)]\langle\psi(t)|\hat{\mathbf{O}}_a,\end{aligned} \tag{A24}$$

$$\frac{\delta J_{\text{forb}}}{\delta |\psi(t)\rangle} = -\langle\psi(t)| \hat{\mathbf{P}}_f^{\gamma}, \tag{A25}$$

$$\frac{\delta J_{\text{con}}}{\delta |\psi(t)\rangle} = \frac{\partial \langle\chi(t)|}{\partial t} + \langle i\hat{\mathbf{H}}(t)\chi(t)|. \tag{A26}$$

Using (A13) and (A23)–(A26), we obtain

$$\frac{\partial \langle\chi(t)|}{\partial t} = -\langle i\hat{\mathbf{H}}(t)\chi(t)| - \langle\psi(t)| \times \left\{\mathcal{C}^{-1}[\tilde{f}_O(\omega)\overline{\langle\hat{\mathbf{O}}_a\rangle}(\omega)]\hat{\mathbf{O}}_a - \hat{\mathbf{P}}_f^{\gamma}\right\}. \tag{A27}$$

Equation (A14) gives the adjoint of (A27):

$$\frac{\partial |\chi(t)\rangle}{\partial t} = -i\hat{\mathbf{H}}(t)|\chi(t)\rangle - \left\{\mathcal{C}^{-1}[\tilde{f}_O(\omega)\overline{\langle\hat{\mathbf{O}}_a\rangle}(\omega)]\hat{\mathbf{O}}_a - \hat{\mathbf{P}}_f^{\gamma}\right\}|\psi(t)\rangle. \tag{A28}$$

In order to derive the expression of the LHS of (A15), we write (A4) in a more useful form. Taking the expectation value of both sides of the Heisenberg equation, we have

$$\frac{d\langle\hat{\mathbf{O}}\rangle(T)}{dt} = i\langle[\hat{\mathbf{H}}(T), \hat{\mathbf{O}}]\rangle(T). \tag{A29}$$

In the special case that $[\hat{\mu}, \hat{\mathbf{O}}] = \hat{\mathbf{0}}$, we have

$$\frac{d\langle\hat{\mathbf{O}}\rangle(T)}{dt} = i\langle[\hat{\mathbf{H}}_0, \hat{\mathbf{O}}]\rangle(T). \tag{A30}$$

In this case, J_{bound} becomes

$$J_{\text{bound}} = \frac{\kappa}{2}\langle\psi(T)|[\hat{\mathbf{H}}_0,\hat{\mathbf{O}}]|\psi(T)\rangle^2. \tag{A31}$$

The LHS of (A15) is

$$\frac{\delta J}{\delta|\psi(T)\rangle} = \frac{\delta J_{\text{bound}}}{\delta|\psi(T)\rangle} + \frac{\delta J_{\text{con}}}{\delta|\psi(T)\rangle}, \tag{A32}$$

$$\frac{\delta J_{\text{bound}}}{\delta|\psi(T)\rangle} = \kappa\langle\psi(T)|[\hat{\mathbf{H}}_0,\hat{\mathbf{O}}]|\psi(T)\rangle\langle\psi(T)|[\hat{\mathbf{H}}_0,\hat{\mathbf{O}}], \tag{A33}$$

$$\frac{\delta J_{\text{con}}}{\delta|\psi(T)\rangle} = -\langle\chi(T)|. \tag{A34}$$

Using (A15) and (A32)–(A34), we obtain

$$\begin{aligned}\langle\chi(T)| &= \kappa\langle\psi(T)|[\hat{\mathbf{H}}_0,\hat{\mathbf{O}}]|\psi(T)\rangle\langle\psi(T)|[\hat{\mathbf{H}}_0,\hat{\mathbf{O}}]\\ &= \kappa\langle[\hat{\mathbf{H}}_0,\hat{\mathbf{O}}]\rangle(T)\langle\psi(T)|[\hat{\mathbf{H}}_0,\hat{\mathbf{O}}].\end{aligned} \tag{A35}$$

Equation (A16) gives the adjoint of (A35):

$$|\chi(T)\rangle = \kappa\langle[\hat{\mathbf{H}}_0,\hat{\mathbf{O}}]\rangle(T)[\hat{\mathbf{H}}_0,\hat{\mathbf{O}}]|\psi(T)\rangle. \tag{A36}$$

Equations (A27), (A28), (A35), and (A36) ensure that

$$\langle\chi(t)| = |\chi(t)\rangle^\dagger.$$

Assuming this, all the computations can be performed using (A28) and (A36) only.

We collect the resulting equations, (A21), (A28), and (A36), together with the constraint (A10):

$$\frac{\partial|\psi(t)\rangle}{\partial t} = -i\hat{\mathbf{H}}(t)|\psi(t)\rangle, \quad |\psi(0)\rangle = |\psi_0\rangle \tag{A37}$$

$$\begin{aligned}\frac{\partial|\chi(t)\rangle}{\partial t} &= -i\hat{\mathbf{H}}(t)|\chi(t)\rangle - \left\{\mathcal{C}^{-1}[\tilde{f}_O(\omega)\overline{\langle\hat{\mathbf{O}}_a\rangle}(\omega)]\hat{\mathbf{O}}_a - \hat{\mathbf{P}}_f^\gamma\right\}\\ &\quad\times|\psi(t)\rangle,\\ |\chi(T)\rangle &= \kappa\langle[\hat{\mathbf{H}}_0,\hat{\mathbf{O}}]\rangle(T)[\hat{\mathbf{H}}_0,\hat{\mathbf{O}}]|\psi(T)\rangle\end{aligned} \tag{A38}$$

$$\begin{aligned}\hat{\mathbf{H}}(t) &= \hat{\mathbf{H}}_0 - \hat{\mu}\epsilon(t)\\ \bar{\epsilon}(\omega) &= \tilde{f}_\epsilon(\omega)\mathcal{C}\left[-\text{Im}\langle\chi(t)\,|\hat{\mu}|\,\psi(t)\rangle\right],\end{aligned} \tag{A39}$$

$$\epsilon(t) = \mathcal{C}^{-1}[\bar{\epsilon}(\omega)]. \tag{A40}$$

These are the Euler-Lagrange equations of the problem.

APPENDIX B: THE APPROXIMATION OF THE HCl MOLECULE

The potential of the H-Cl bond was obtained by adjusting the parameters of the Morse potential

$$V(x) = D_0[\exp(-ax) - 1]^2 \tag{B1}$$

to experimental data on HCl—the atomization energy of HCl and the frequency of vibration, using the infrared (IR) absorption frequency for the transition to the fundamental state. We made a few reasonable approximations. The resulting potential is presented in Table IV (the second term of $\hat{\mathbf{H}}_0$).

The dipole function was obtained by adjusting experimental data to a reasonable functional form. The experimental data is the first four derivatives of $\mu(x)$ at equilibrium [27]:

$$\left(\frac{d^n\mu}{dx^n}\right)_{\text{eq}}, \quad n = 1,2,3,4.$$

The functional form is

$$\mu(x) = a_1 x\{1 - \tanh[a_2(x - a_3)^{a_4}]\}. \tag{B2}$$

We made the approximation:

$$\left(\frac{d^n\mu}{dx^n}\right)_{x=0} \approx \left(\frac{d^n\mu}{dx^n}\right)_{\text{eq}}.$$

The resulting system of equations was solved using the Symbolic Math Toolbox of MATLAB. The resulting function is complex. We take its real part (see Table IV).

[1] N. H. Burnett, H. A. Baldis, M. C. Richardson, and G. D. Enright, Appl. Phys. Lett. **31**, 172 (1977).

[2] M. Ferray, A. Lhuillier, X. F. Li, L. A. Lompre, G. Mainfray, and C. Manus, J. Phys. B **21**, L31 (1988).

[3] P. B. Corkum, Phys. Rev. Lett. **71**, 1994 (1993).

[4] M. Lewenstein, P. Balcou, M. Y. Ivanov, A. L'Huillier, and P. B. Corkum, Phys. Rev. A **49**, 2117 (1994).

[5] E. Constant, D. Garzella, P. Breger, E. Mével, C. Dorrer, C. Le Blanc, F. Salin, and P. Agostini, Phys. Rev. Lett. **82**, 1668 (1999).

[6] J. Mauritsson, P. Johnsson, E. Gustafsson, A. L'Huillier, K. J. Schafer, and M. B. Gaarde, Phys. Rev. Lett. **97**, 013001 (2006).

[7] J. Itatani, D. Zeidler, J. Levesque, M. Spanner, D. M. Villeneuve, and P. B. Corkum, Phys. Rev. Lett. **94**, 123902 (2005).

[8] P. Salieres, P. Antoine, A. de Bohan, and M. Lewenstein, Phys. Rev. Lett. **81**, 5544 (1998).

[9] C. Winterfeldt, C. Spielmann, and G. Gerber, Rev. Mod. Phys. **80**, 117 (2008).

[10] A. P. Peirce, M. A. Dahleh, and H. Rabitz, Phys. Rev. A **37**, 4950 (1988).

[11] Ronnie Kosloff, Stuart A. Rice, Pier Gaspard, Sam Tersigni, and David Tannor, Chem. Phys. **139**, 201 (1989).

[12] José P. Palao and Ronnie Kosloff, Phys. Rev. A **68**, 062308 (2003).

[13] J. Werschnik and E. K. U. Gross, J. Phys. B **40**, R175 (2007).

[14] I. Degani, A. Zanna, L. Sælen, and R. Nepstad, SIAM J. Sci. Comput. **31**, 3566 (2009).

[15] T. E. Skinner, and N. I. Gershenzon, J. Mag. Res. **204**, 248 (2010).

[16] C. Gollub, M. Kowalewski, and R. de Vivie-Riedle, Phys. Rev. Lett. **101**, 073002 (2008).

[17] C. Gollub, M. Kowalewski, Markus S. Thallmair, and R. de Vivie-Riedle, Phys. Chem. Chem. Phys. **12**, 15780 (2010).

[18] M. Lapert, R. Tehini, G. Turinici, and D. Sugny, Phys. Rev. A **79**, 063411 (2009).

[19] M. Schroeder and A. Brown, New J. Phys. **11**, 105031 (2009).

[20] F. Motzoi, J. M. Gambetta, S. T. Merkel, and F. K. Wilhelm, Phys. Rev. A **84**, 022307 (2011).

[21] I. Serban, J. Werschnik, and E. K. U. Gross, Phys. Rev. A **71**, 053810 (2005).
[22] José P. Palao, Ronnie Kosloff, and Christiane P. Koch, Phys. Rev. A **77**, 063412 (2008).
[23] I. Schaefer, arXiv:1202.6520v1.
[24] Hillel Tal-Ezer, Ronnie Kosloff, and Ido Schaefer, J. Sci. Comput. **53**, 211 (2012).
[25] L. Rybak, L. Chuntonov, A. Gandman, N. Shakour, and Z. Amitay, Opt. Express **16**, 21738 (2008).
[26] R. A. Bartels, M. M. Murnane, H. C. Kapteyn, I. Christov, and H. Rabitz, Phys. Rev. A **70**, 043404 (2004).
[27] E. W. Kaiser, J. Chem. Phys. **53**, 1686 (1970).

Chapter 3

Semi-global approach for propagation of the time-dependent Schrödinger equation for time-dependent and nonlinear problems

Published; full citation:

Ido Schaefer, Hillel Tal-Ezer, and Ronnie Kosloff, *Semi-global approach for propagation of the time-dependent Schrödinger equation for time-dependent and nonlinear problems*, Journal of Computational Physics 343 (2017), 368 – 413.

The present chapter discusses the development of a highly accurate propagation method, and its application to non-Hermitian dynamics in general, and the physical situation of HHG in particular. The dynamics of the HHG problem becomes non-Hermitian due to the employment of absorbing boundary conditions.

The HHG simulation requires highly accurate numerical tools, due to the sensitivity of the high-harmonic spectrum to tiny numerical artefacts (see Chapter 4). The propagation

method described in the present chapter enables minimization of propagation artefacts. Another source of inaccuracy is imperfection in the absorption capabilities of the absorbing boundaries. This topic is also addressed in the present chapter.

Journal of Computational Physics 343 (2017) 368–413

Contents lists available at ScienceDirect

Journal of Computational Physics

www.elsevier.com/locate/jcp

Semi-global approach for propagation of the time-dependent Schrödinger equation for time-dependent and nonlinear problems

Ido Schaefer, Hillel Tal-Ezer, Ronnie Kosloff

ARTICLE INFO

Article history:
Received 15 December 2016
Received in revised form 1 March 2017
Accepted 4 April 2017
Available online xxxx

Keywords:
Time-dependent Schrödinger equation
Propagation
System of ODE's
Quantum dynamics

ABSTRACT

A detailed exposition of highly efficient and accurate method for the propagation of the time-dependent Schrödinger equation [50] is presented. The method is readily generalized to solve an arbitrary set of ODE's. The propagation is based on a global approach, in which large time-intervals are treated as a whole, replacing the local considerations of the common propagators. The new method is suitable for various classes of problems, including problems with a time-dependent Hamiltonian, nonlinear problems, non-Hermitian problems and problems with an inhomogeneous source term. In this paper, a thorough presentation of the basic principles of the propagator is given. We give also a special emphasis on the details of the numerical implementation of the method. For the first time, we present the application for a non-Hermitian problem by a numerical example of a one-dimensional atom under the influence of an intense laser field. The efficiency of the method is demonstrated by a comparison with the common Runge–Kutta approach.

1. Introduction

The time-dependent Schrödinger equation (TDSE),

$$\frac{d\psi(t)}{dt} = -\frac{i}{\hbar}\hat{\mathbf{H}}(t)\psi(t) \tag{1}$$

is a central pillar in quantum dynamics. Solution of the equation supplies insight on fundamental quantum processes. For the majority of problems closed form solutions do not exist. An alternative is to develop numerical schemes able to simulate from first principle quantum processes. In the present paper we concentrate on the issue of time-propagation.

We assume that the formal operation $\phi = \hat{\mathbf{H}}\psi$ can be carried out in a matrix vector representation, $\vec{\mathbf{v}} = H\vec{\mathbf{u}}$. Hence, the problem of solving Eq. (1) becomes a special case of the problem of solving a general set of ordinary differential equations (ODE's).

Historically, the common practice in early studies was either to solve Eq. (1) by general methods for solving a set of ODE's, or by methods which were developed particularly for Eq. (1). General solvers for a set of ODE's rely on approximations to a low order Taylor expansion of $\psi(t)$, derived from Eq. (1). This leads to the necessity of a time-step propagation scheme (see Sec. 2.2). The most popular methods for general applications are the Runge–Kutta methods [7]. Another method, which became very popular in early quantum studies, is second order differencing [23].

E-mail addresses: ido.schaefer@mail.huji.ac.il (I. Schaefer), hillel@mta.ac.il (H. Tal-Ezer), ronnie@fh.huji.ac.il (R. Kosloff).

http://dx.doi.org/10.1016/j.jcp.2017.04.017
0021-9991/

Commonly, early researchers preferred other methods, which were intended specifically for quantum applications. These methods have conservation properties of certain physical quantities, such as norm or energy. The popular methods are Crank–Nicolson implicit scheme [40] and split operator exponentiation [12] (it is noteworthy that the second order differencing method has also special conservation properties). These methods are also equivalent to an approximation of a low order Taylor expansion of $\psi(t)$, and lead to a time-step scheme. The advantage of these methods over the general methods is questionable, since the overall quality of the obtained solution is not improved over the general methods. The convergence becomes non-uniform—the error is accumulated in the physical quantities which are not conserved by the propagation scheme. In particular, the norm conservation leads to larger accumulation of errors in phase.

All the methods that were mentioned can be classified as *local methods*. They all share the common property of being equivalent to a low order Taylor expansion in time. The Taylor expansion has slow convergence properties, which limit its application for approximation purposes to low orders. This leads to the locality of the solution, and consequently, to the time-step integration scheme. The drawback of a time-step scheme is the accumulation of the errors in each time step, which limits the accuracy for realistic times. The time step Δt is limited by the spectral range of $\hat{\mathbf{H}}$, ΔE. Typically, the propagation process is numerically stable only for time-steps which do not exceed $\Delta t \sim \frac{1}{10}\hbar/\Delta E$. The final accuracy of a Taylor propagation method scales as $O(\Delta t^n)$ where n is the order of the method.

A breakthrough in solving the TDSE was the development of the global Chebyshev propagator. The global Chebyshev propagator solves Eq. (1) for a *time-independent Hamiltonian*, $\hat{\mathbf{H}}(t) \equiv \hat{\mathbf{H}}$, without the necessity of a time-step scheme. The whole propagation interval is treated *globally* in a single step. The development of the method was led by the insight that a local time-step integration scheme is unnecessary when integration can be performed analytically. With a time-independent Hamiltonian, Eq. (1) can be directly integrated to yield

$$\psi(t) = \exp\left(-\frac{i}{\hbar}\hat{\mathbf{H}}t\right)\psi(0) \tag{2}$$

The direct computation of this expression becomes highly demanding for large-scale problems (see Sec. 2.3.1); the computation in the Chebyshev propagator is based on a polynomial Chebyshev expansion of the evolution operator, $\hat{\mathbf{U}} = \exp(-\frac{i}{\hbar}\hat{\mathbf{H}}t)$ [48]. The method has uniform convergence and does not accumulate errors. The computational effort scales as $\sim \frac{\Delta E\, t}{2\hbar}$. The Chebyshev scheme outperforms all other methods for problems with time independent Hamiltonian operators [27]. More than thirty years after it was introduced it is still the method of choice for efficient and high accuracy solution to large scale problems with a time-independent Hamiltonian [53].

The case of a non-Hermitian Hamiltonian requires a special care. The global Chebyshev propagator was developed for a Hermitian Hamiltonian. Without modification it is not suitable for non-Hermitian operators. A generalization of Chebyshev are Faber polynomials which enable to enter the complex plane [16,20,21]. A different approach for non-Hermitian operators led to the development of Newtonian propagators. The Chebyshev approximation on the real axis is replaced by a Newton interpolation in the complex plane. The first application was the solution of the Liouville–von-Neumann equation [5]. The Newtonian scheme was later implemented also to the Schrödinger equation with absorbing boundary conditions [1]. A comparison of the different schemes and the relation to other propagators [9,37,51,54] has been reviewed [15,24].

The global schemes mentioned above assume a previous knowledge on the eigenvalue domain of the Hamiltonian. Commonly, such a knowledge is missing, in particular in non-Hermitian problems. In such a case, the global approach can be implemented by the Arnoldi approach, using a restarted Arnoldi algorithm (see, for example, [47]). A paper on this topic has not been published to date.

The focus of the present study is solving the Schrödinger equation for problems in which the Hamiltonian is explicitly time-dependent. Such problems are common in ultrafast spectroscopy, coherent control and high harmonic generation. Another typical complication arises when the Hamiltonian becomes nonlinear, i.e. explicitly depends on the state $\psi(t)$. Mean field approximation typically lead to such equations. Examples are the time-dependent Gross–Pitaevskii approximation [4], time-dependent Hartree [26,29] and time-dependent DFT [14]. In general, Eq. (1) cannot be integrated analytically for a Hamiltonian with time-dependence or nonlinearity. A less common complication arises when a source term is added to the Schrödinger equation. Such equations can be found in scattering theory [34] and in particular problems in coherent control [35,41,42].

The common practice to overcome the explicit time dependence, or nonlinearity, is to resort again to a time-step scheme, which relies on Eq. (2). The propagation interval is divided into small time-steps, where the Hamiltonian is assumed to be stationary within the time-step. This becomes equivalent to a first order method in time. The result is either a significant increase in computational cost or low accuracy. In our opinion, this is a misuse of the global Chebyshev propagator, which is intended to overcome this very problem. A better scheme is based on the Magnus expansion [28] to correct for time ordering, and a low order polynomial approximation for the exponent of an operator [2,6,10,49]. This leads again to a local scheme, since the radius of convergence of the Magnus expansion is limited [6]. However, the scaling of the error with Δt is improved over the common naive practice. Another local approach with improved scaling is to use a high order splitting method [44].

Attempts have been made to implement the global approach in problems with time-dependent or nonlinear Hamiltonians. A very accurate scheme was developed, based on embedding in a larger space, which includes also time. In the extended space, the problem can be formulated by global means [38]. The drawback was the very high computational cost

which scaled as $\Delta E\,\Delta t\,\Delta\omega$ where ΔE is the eigenvalue range of the Hamiltonian, Δt is the time step and $\Delta\omega$ is the bandwidth of the explicit time-dependent function. In addition, the method was not applicable to nonlinear problems.

Another attempt was proposed in [3]. Like in the global Chebyshev propagator, the propagation method is based on the idea of global integration of Eq. (1). A direct integration leads to an integral equation:

$$\psi(t) = \psi(0) - \frac{i}{\hbar}\int_0^t \hat{\mathbf{H}}[\psi(\tau),\tau]\,\psi(\tau)\,d\tau \tag{3}$$

The integral is approximated by the expansion of the integrand in time in a truncated Chebyshev series, which can be integrated analytically. This results in a system of equations of $\psi(t)$ in multiple time-points. In the nonlinear case, the system of equations becomes also nonlinear. Seemingly, this replaces the problem of time-propagation with the even more difficult problem of optimization. However, the system can be solved by a relatively simple iterative scheme when the time-interval is sufficiently small. This leads again to a time-step scheme. The hope was that it will be possible to use larger time-steps than other propagation schemes, thus leading to reduction of the error accumulation during propagation. Later it was found that this approach led to extremely small time-steps and a large number of iterations, thus becoming highly inefficient (from our experience, and private communications with the author). The failure of the method clearly lies in the necessity of solving a system of equations, a task which was found to be much more demanding than a local propagation scheme.

The introduction of source terms was followed by the development of a global propagation scheme for inhomogeneous problems [22,32]. This led to new insight to the problem of the time-dependence or nonlinearity of the Hamiltonian. A new global scheme, based on integration in large time-steps, was first introduced in [33]. The method is based on another integrated version of Eq. (1), in which the $\psi(t)$ dependence in the integral expression is minimized in comparison to (3). Here again, an iterative scheme is used to solve the resulting system of equations. This scheme was a significant improvement to that introduced in [3]. However, we found that it gave inferior results in comparison to a 4'th order Runge–Kutta scheme (RK4). A drastic improvement was achieved by new insights on the propagation scheme [50]. The improved scheme was demonstrated to be significantly more efficient than the Taylor methods, particularly when high accuracy is required. Quite importantly, the scheme was generalized to solve an arbitrary set of ODE's.

The propagation approach of the new scheme, as well as Ref. [3], combines global and local elements. Hence, it can be classified as a *semi-global propagation approach*.

The original global Chebyshev propagator [48] was easy to program. This led to fast proliferation with many applications. The new algorithm for explicit time dependence became more involved with three user defined parameters which control the accuracy and efficiency. However, we believe that the vast increase in efficiency is worth the effort of learning and computing the algorithm.

The present paper consolidates the numerical scheme. In addition, the application of the algorithm is extended to non-Hermitian Hamiltonians. Our purpose is to give explicit description of all steps and considerations in the scheme to enable the potential user either to program from scratch or to be able to tailor an existing program to the problem of choice. Although the scheme is more involved than the basic Chebyshev scheme, we hope that the explicit description will lead to proliferation of the method.

2. Theory

2.1. Definition of the problem

Let us rewrite the time-dependent Schrödinger equation in a matrix–vector notation:

$$\frac{d\vec{\mathbf{u}}(t)}{dt} = -iH(t)\vec{\mathbf{u}}(t) \tag{4}$$

where $\vec{\mathbf{u}}(t)$ represents the state, and $H(t)$ is a matrix representing the time-dependent Hamiltonian of the system. (Atomic units are used throughout, so we set $\hbar = 1$.)

In our discussion, we shall consider a generalization of Eq. (4). First, we let H include a dependence on the state vector, $\vec{\mathbf{u}}(t)$, i.e. $H \equiv H(\vec{\mathbf{u}}(t), t)$. This results in a nonlinear equation of motion. In addition, we include an inhomogeneous *source term* $\vec{\mathbf{s}}(t)$. The time-dependent nonlinear inhomogeneous Schrödinger equation reads:

$$\frac{d\vec{\mathbf{u}}(t)}{dt} = -iH(\vec{\mathbf{u}}(t), t)\vec{\mathbf{u}}(t) + \vec{\mathbf{s}}(t) \tag{5}$$

Actually, Eq. (5) has the form of a much more general problem. A general set of ODE's is equivalent to an equation of the following form:

$$\frac{d\vec{\mathbf{u}}(t)}{dt} = \vec{\mathbf{g}}(\vec{\mathbf{u}}(t), t) \tag{6}$$

where $\vec{\mathbf{g}}(\vec{\mathbf{u}}(t),t)$ is an arbitrary vector function of $\vec{\mathbf{u}}(t)$ and t. This can be always rewritten as:

$$\frac{d\vec{\mathbf{u}}(t)}{dt} = G(\vec{\mathbf{u}}(t),t)\vec{\mathbf{u}}(t) + \vec{\mathbf{s}}(t) \tag{7}$$

where $G(\vec{\mathbf{u}}(t),t)$ is a matrix. Hence, the problem of solving Eq. (5) is equivalent to the problem of solving Eq. (7), by setting $G(\vec{\mathbf{u}}(t),t) = -iH(\vec{\mathbf{u}}(t),t)$. As a matter of convenience, we shall use the form of Eq. (7) in our discussion.

The initial condition for the vector state is:

$$\vec{\mathbf{u}}(0) = \vec{\mathbf{u}}_0 \tag{8}$$

We require the solution, $\vec{\mathbf{u}}(t)$, at an arbitrary time t.

2.2. Local approach—Taylor methods

The popular algorithms for solving a general set of ODE's are based on Taylor expansion considerations. In order to illustrate this approach, we will consider the *Euler method* which is the simplest Taylor method.

The Euler method is based on a first order Taylor expansion for approximation of the solution at a close point. The solution at $t = \Delta t$ is approximated by:

$$\vec{\mathbf{u}}(\Delta t) \approx \vec{\mathbf{u}}(0) + \Delta t \frac{d\vec{\mathbf{u}}(0)}{dt} \tag{9}$$

$\vec{\mathbf{u}}(0)$ is given by Eq. (8). $d\vec{\mathbf{u}}(0)/dt$ can be computed by plugging Eq. (8) into Eq. (7). Using a first order approximation, the solution will be of low accuracy, unless Δt is sufficiently small. In order to get an accurate solution far from $t = 0$, we have to march in small time-steps. The solution at $t = 2\Delta t$ is computed in the same way, using $\vec{\mathbf{u}}(\Delta t)$ from Eq. (9), and Eq. (7) for obtaining $d\vec{\mathbf{u}}(\Delta t)/dt$. We continue by repeating this propagation technique until we reach the solution at the final time, which will be denoted as $t = T$. If desired, the accuracy of the solution can be improved by choosing a smaller time-step Δt. Of course, this requires more computational effort.

The Euler method is rarely used, because of its slow convergence properties with the decrement of Δt. The error of the solution in T scales as $O(\Delta t)$. Other Taylor methods are based on higher order expansions. The error of a Taylor method of order n scales as $O(\Delta t^n)$.

The most popular Taylor methods are the *Runge–Kutta methods*. The idea underlying the Runge–Kutta methods is to approximate the Taylor expansion without a direct evaluation of high-order derivatives of $\vec{\mathbf{u}}(t)$. The Taylor expansion is approximated by first order derivative evaluations, using Eq. (7). This approximation preserves the scaling of the error with Δt. For instance, we consider the Runge–Kutta method of the 4'th order (RK4). The solution at $t = \Delta t$ is approximated by:

$$\begin{aligned}
&\vec{\mathbf{u}}(\Delta t) \approx \vec{\mathbf{u}}(0) + \frac{1}{6}(\vec{\mathbf{k}}_1 + 2\vec{\mathbf{k}}_2 + 2\vec{\mathbf{k}}_3 + \vec{\mathbf{k}}_4) \\
&\vec{\mathbf{k}}_1 = \Delta t\, \vec{\mathbf{g}}(\vec{\mathbf{u}}(0), 0) \\
&\vec{\mathbf{k}}_2 = \Delta t\, \vec{\mathbf{g}}\left(\vec{\mathbf{u}}(0) + \frac{\vec{\mathbf{k}}_1}{2}, \frac{\Delta t}{2}\right) \\
&\vec{\mathbf{k}}_3 = \Delta t\, \vec{\mathbf{g}}\left(\vec{\mathbf{u}}(0) + \frac{\vec{\mathbf{k}}_2}{2}, \frac{\Delta t}{2}\right) \\
&\vec{\mathbf{k}}_4 = \Delta t\, \vec{\mathbf{g}}(\vec{\mathbf{u}}(0) + \vec{\mathbf{k}}_3, \Delta t)
\end{aligned} \tag{10}$$

Eq. (10) approximates a fourth order Taylor expansion. In our numerical example (Sec. 4) we shall use RK4 as a reference method.

The Taylor approach is based on local considerations—in each time-step, the solution $\vec{\mathbf{u}}(t)$ is approximated using our knowledge on the local behavior of $\vec{\mathbf{u}}(t)$ at the previous time-point. The information on the behavior of $\vec{\mathbf{u}}(t)$ is deduced from its derivatives *at the time-point*. For this information to be accurate, it is essential that the time-point in which the solution is to be evaluated is close enough. Hence, it is necessary to propagate in small time-steps. The many time-step propagation scheme results in a large computational effort. Moreover, the error is accumulated during the propagation process. These drawbacks are direct consequences of the locality of the Taylor approach.

Another drawback of the Taylor approach lies in the slow convergence properties of the Taylor series. These reduce the efficiency of this approach when using high order Taylor expansions. The popular Runge–Kutta methods are based on 4'th or 5'th order expansions.

In what follows, we shall accommodate with these problems by developing a more global approach for the task of solving Eq. (7).

2.3. Global approach using closed integrated forms

In order to approach the problem in a global manner, we seek a way to treat the whole time interval of the problem in a single stage. Indeed, Eq. (7) can be solved in a single step in the special cases that it can be integrated analytically. The closed integrated forms in these special cases constitute the basis for the present approach.

2.3.1. Time-independent Hamiltonian

We start from the simplest case with a closed integrated form. The Hamiltonian is time-independent:

$$H(t) \equiv H_0$$

or, equivalently:

$$G(t) \equiv G_0$$

In addition, there is no inhomogeneous term:

$$\vec{\mathbf{s}}(t) \equiv 0$$

Eq. (7) becomes:

$$\frac{d\vec{\mathbf{u}}(t)}{dt} = G_0\vec{\mathbf{u}}(t) \tag{11}$$

This equation, with the initial condition (8), can be integrated directly to yield:

$$\vec{\mathbf{u}}(t) = \exp(G_0 t)\vec{\mathbf{u}}_0 \tag{12}$$

for an arbitrary t. In the special case of the Schrödinger equation, we have the well known result of the situation of stationary dynamics:

$$\vec{\mathbf{u}}(t) = \exp(-iH_0 t)\vec{\mathbf{u}}_0 \tag{13}$$

The problem that arises is that the exponent of the matrix $G_0 t$ cannot be computed directly (unless G_0 is diagonal). One immediate approach is to diagonalize G_0 and compute the function of the matrix in the basis of the eigenvectors of G_0. Then we can write:

$$\vec{\mathbf{u}}(t) = S\exp(Dt)S^{-1}\vec{\mathbf{u}}_0 \tag{14}$$

where D is the diagonalized G_0, and S is the transformation matrix from the eigenvector basis to the original basis. The problem with this approach is that when the dimension of G_0 is large, it becomes infeasible to diagonalize it, because of the high numerical cost of this operation—diagonalization scales as $O(N^3)$, where N is the dimension of the problem.

Usually, Eq. (13) is solved by another approach, which is less demanding numerically. We expand the RHS of Eq. (12) in a polynomial series in G_0. First, we define a function $f(x) = \exp(xt)$, where t is treated as a parameter. Then we approximate it by a truncated polynomial series:

$$f(x) \approx \sum_{n=0}^{K-1} a_n P_n(x) \tag{15}$$

where $P_n(x)$ is a polynomial of degree n, and a_n is the corresponding expansion coefficient. This requires the choice of the set of expansion polynomials $P_n(x)$, and the computation of the corresponding a_n's. Then, we approximate Eq. (12) as:

$$\vec{\mathbf{u}}(t) \approx \sum_{n=0}^{K-1} a_n P_n(G_0)\vec{\mathbf{u}}_0 \tag{16}$$

The expansion (15) has to be accurate in the eigenvalue domain of G_0 in order that the form (16) will be useful (see Appendix B.1).

The RHS of Eq. (16) can be computed by successive matrix–vector multiplications. Matrix–vector multiplications scale just as $O(N^2)$. In many cases, the computational effort can be reduced further. The direct multiplication of $\vec{\mathbf{u}}_0$ by G_0 can be replaced by the operation of an equivalent linear operator. The operation of the linear operator can be defined by a computational procedure, which may have a lower scaling with N. For instance, in the Fourier grid method (see [23]) the Hamiltonian operation scales as $O(N \ln N)$ only.

An immediate question that arises is how to choose the set of expansion polynomials $P_n(x)$. One might suggest to use the Taylor polynomials, i.e. $P_n(x) = x^n$, and expand $f(x)$ in a Taylor series. However, this would be a poor choice, because of the slow convergence properties of a Taylor series. The reason for the slow convergence lies in the low quality of the Taylor polynomials as expansion functions—as n increases, they are getting closer to be parallel in the function space. In order to

attain a fast convergence of the polynomial series with K, an orthogonal set of polynomials should be used. The expansion coefficients a_n are given by a scalar product of the $P_n(x)$'s with $f(x)$.

Usually, $f(x)$ is expanded in a *Chebyshev polynomial* series, or equivalently, by a *Newton interpolation polynomial* at the *Chebyshev points* of the eigenvalue domain. When the Hamiltonian is non-Hermitian, the eigenvalue domain becomes complex, and the Chebyshev approach is not appropriate anymore. Then, the *Arnoldi approach* should be used instead. In Appendix A we present the approximation methods of a function by a Newton interpolation polynomial or a Chebyshev polynomial series. In Appendix B we describe the different approximation methods for the multiplication of a vector by a function of matrix, by Chebyshev or Newton series, or by the Arnoldi approach.

In the Chebyshev or Newton methods, an approximation of degree $K-1$, with K expansion terms, requires $K-1$ matrix–vector multiplications. This is due to the recurrence relations between the expansion polynomials in both methods, as will be described in Appendix B. Similarly, in the Arnoldi approach, K matrix–vector multiplications are required for K expansion terms.

Note that using Eq. (13), the solution is given only at the chosen t, and not at intermediate time points. Usually, it is desirable to follow the whole physical process which leads to the result at the final time, and the intermediate times are also of interest. Actually, the solution at the intermediate time-points can be obtained with a negligible additional computational effort. Let us rewrite Eq. (15) for each of the time-points to be computed:

$$\begin{aligned} f_j(x) &= \exp(xt_j) \qquad j = 1, \ldots, N_{tp} \\ f_j(x) &= \sum_{n=0}^{K-1} a_{n,j} P_n(x) \end{aligned} \tag{17}$$

where N_{tp} is the number of time points, and t_j is the j'th time-point. The only difference between the t_j's is in the definition of the $f_j(x)$'s, and the corresponding expansion coefficients, $a_{n,j}$. The $P_n(x)$'s remain the same. Hence, it is sufficient to compute the $P_n(G_0)\vec{\mathbf{u}}_0$ just once for all the desired time points. The $a_{n,j}$'s are computed for each time-point t_j. The computational effort of the matrix–vector multiplications (or the equivalent linear operations) is much greater than that of the computation of the $a_{n,j}$'s, unless N is very small.

2.3.2. Addition of a source term with a polynomial time-dependence

Now let us add to Eq. (11) a source term:

$$\frac{d\vec{\mathbf{u}}(t)}{dt} = G_0\vec{\mathbf{u}}(t) + \vec{\mathbf{s}}(t) \tag{18}$$

A source term is not very common in quantum applications. Nevertheless, the treatment of the common cases of a time-dependent or nonlinear Hamiltonian relies on the results that will be derived in the present section.

Eq. (18) can be integrated using the *Duhamel principle* which relates the solution for the inhomogeneous equation to that of the corresponding homogeneous equation. Let us denote the evolution matrix for the homogeneous equation (11) by:

$$U_0(t) = \exp(G_0 t) \tag{19}$$

The Duhamel principle states that the solution of Eq. (11) can be written by the means of $U_0(t)$ in the following way:

$$\begin{aligned} \vec{\mathbf{u}}(t) &= U_0(t)\vec{\mathbf{u}}_0 + \int_0^t U_0(t-\tau)\vec{\mathbf{s}}(\tau)\,d\tau \\ &= \exp(G_0 t)\vec{\mathbf{u}}_0 + \int_0^t \exp[G_0(t-\tau)]\vec{\mathbf{s}}(\tau)\,d\tau \\ &= \exp(G_0 t)\vec{\mathbf{u}}_0 + \exp(G_0 t)\int_0^t \exp(-G_0\tau)\vec{\mathbf{s}}(\tau)\,d\tau \end{aligned} \tag{20}$$

Eq. (20) assumes a closed form when the integral in the RHS of Eq. (20),

$$\int_0^t \exp(-G_0\tau)\vec{\mathbf{s}}(\tau)\,d\tau \tag{21}$$

can be performed analytically.

We shall focus on a family of source terms for which (21) assumes a closed form—source terms with a polynomial time-dependence:

$$\vec{\mathbf{s}}(t) = \sum_{m=0}^{M-1} \frac{t^m}{m!} \vec{\mathbf{s}}_m \tag{22}$$

First, we need a closed expression for (21) in this particular case. For convenience, we will discuss the scalar version of (21), without loss of generality:

$$\int_0^t \exp(-z\tau) s(\tau)\, d\tau \tag{23}$$

where z is a complex variable, and $s(t)$ is a scalar function of the form

$$s(t) = \sum_{m=0}^{M-1} \frac{t^m}{m!} s_m \tag{24}$$

After we obtain a closed expression, we will be able to write the RHS of Eq. (20) by the means of multiplication of vectors by functions of the matrix G_0. Finally, we will show that the solution can be written in a modified form, in which the computational effort is much reduced.

First, we discuss a source term of the form:

$$s(t) = \frac{t^m}{m!} s_m \tag{25}$$

Let us define:

$$J_{m+1}(z,t) \equiv \int_0^t \exp(-z\tau)\tau^m\, d\tau, \qquad m = 0, 1, \ldots \tag{26}$$

which are the integrals that need to be evaluated. We start with the simplest situation, when $m = 0$, and $s(t)$ becomes a constant. In the case that $z \neq 0$ we obtain:

$$J_1(z,t) \equiv \int_0^t \exp(-z\tau)\, d\tau = \frac{1 - \exp(-zt)}{z} \tag{27}$$

When $z = 0$, we obtain:

$$J_1(0,t) = t \tag{28}$$

Now let us consider the case that $m > 0$. If $z \neq 0$, we can evaluate the integral using integration by parts. A simple calculation yields:

$$J_{m+1}(z,t) = -\frac{\exp(-zt)t^m}{z} + \frac{m}{z}\int_0^t \exp(-z\tau)\tau^{m-1}\, d\tau = -\frac{\exp(-zt)t^m}{z} + \frac{m}{z} J_m(z,t) \tag{29}$$

or, equivalently:

$$J_m(z,t) = -\frac{\exp(-zt)t^{m-1}}{z} + \frac{m-1}{z} J_{m-1}(z,t), \qquad m = 2, 3, \ldots \tag{30}$$

Successive operations of the resulting recursion formula lead to the following expression:

$$J_m(z,t) = \frac{(m-1)!}{z^m}\left[1 - \exp(-zt)\sum_{j=0}^{m-1}\frac{(zt)^j}{j!}\right], \qquad m = 1, 2, \ldots \tag{31}$$

Note that Eq. (31) applies also for $J_1(z,t)$. In the case that $z = 0$ we have:

$$J_m(0,t) = \frac{t^m}{m}, \qquad m = 1, 2, \ldots \tag{32}$$

We proceed to the evaluation of the scalar form of Eq. (20):

$$u(t) = \exp(zt)u_0 + \exp(zt)\int_0^t \exp(-z\tau)s(\tau)\,d\tau \tag{33}$$

for a source term of the form (24). We begin with the treatment of the second term in the RHS of Eq. (33). Plugging (24) into this term, we obtain:

$$\exp(zt)\sum_{m=0}^{M-1}\frac{1}{m!}\int_0^t \exp(-z\tau)t^m\,d\tau\,s_m = \exp(zt)\sum_{m=0}^{M-1}\frac{1}{m!}J_{m+1}(z,t)s_m = \sum_{m=0}^{M-1} f_{m+1}(z,t)s_m \tag{34}$$

where we defined:

$$f_m(z,t) \equiv \begin{cases} \frac{1}{z^m}\left[\exp(zt) - \sum_{j=0}^{m-1}\frac{(zt)^j}{j!}\right] & z \neq 0 \\ \frac{t^m}{m!} & z = 0 \end{cases} \qquad m = 1, 2, \ldots \tag{35}$$

The corresponding scalar form of Eq. (20) becomes:

$$u(t) = \exp(zt)u_0 + \sum_{m=0}^{M-1} f_{m+1}(z,t)s_m \tag{36}$$

Let us write Eq. (36) in a prettier way. First, we define the following set of constants:

$$w_m \equiv \begin{cases} u_0 & m = 0 \\ s_{m-1} & 0 < m \leq M \end{cases} \tag{37}$$

Second, we note that the definition (35) can be extended to the case of $m = 0$, using the convention that

$$\sum_{j=L}^{N} b_j = 0, \qquad N < L \tag{38}$$

for arbitrary b_j's. Using this extension of definition, we have:

$$f_0(z,t) = \exp(zt) \tag{39}$$

Then, Eq. (36) becomes:

$$u(t) = \sum_{m=0}^{M} f_m(z,t)w_m \tag{40}$$

We can use the form of Eq. (40) to write an analogous vector solution for Eq. (18):

$$\vec{\mathbf{u}}(t) = \sum_{m=0}^{M} f_m(G_0,t)\vec{\mathbf{w}}_m \tag{41}$$

where the $\vec{\mathbf{w}}_m$'s are defined in an analogous manner to the scalar w_m's.

When we compare Eq. (41) to Eq. (12), it seems that the addition of the source term is quite expensive numerically. In Eq. (12) it is necessary to evaluate just one multiplication of a vector by a function of a matrix. In Eq. (41), it is necessary to perform the same kind of operation $M + 1$ times. Actually, the computational effort can be much reduced, if we rewrite Eq. (41) in a modified form.

Let us return to the corresponding scalar equation, Eq. (40). We are going to show that it can be rewritten using just one of the $f_m(z,t)$ functions. Observing the definition (35), it can be easily seen that

$$f_m(z,t) = z f_{m+1}(z,t) + \frac{t^m}{m!} \tag{42}$$

Eq. (42) implies that a function $f_m(z,t)$ can be expressed using any of the other $f_k(z,t)$ functions. If $k > m$, we need $k - m$ successive operations of Eq. (42) in order to write $f_m(z,t)$ in the terms of $f_k(z,t)$. We obtain:

$$f_m(z,t) = z^{k-m} f_k(z,t) + \sum_{j=m}^{k-1}\frac{t^j}{j!}z^{j-m} \tag{43}$$

Eq. (43) can be applied also to the case of $k = m$, with the summation convention (38).

Using Eq. (43), we can express all the $f_m(z,t)$ functions in Eq. (40) by the function with the largest m, i.e. $f_M(z,t)$. Eq. (40) becomes:

$$\begin{aligned} u(t) &= \sum_{m=0}^{M}\left[z^{M-m} f_M(z,t) w_m + \sum_{j=m}^{M-1} \frac{t^j}{j!} z^{j-m} w_m\right] \\ &= f_M(z,t)\sum_{m=0}^{M} z^{M-m} w_m + \sum_{j=0}^{M-1}\frac{t^j}{j!}\sum_{m=0}^{j} z^{j-m} w_m \\ &= f_M(z,t) v_M + \sum_{j=0}^{M-1}\frac{t^j}{j!} v_j \end{aligned} \tag{44}$$

where we defined:

$$v_j \equiv \sum_{m=0}^{j} z^{j-m} w_m \tag{45}$$

Returning to the vector solution of Eq. (18), we can write an analogous expression:

$$\vec{\mathbf{u}}(t) = f_M(G_0,t)\vec{\mathbf{v}}_M + \sum_{j=0}^{M-1}\frac{t^j}{j!}\vec{\mathbf{v}}_j \tag{46}$$

where

$$\vec{\mathbf{v}}_j \equiv \sum_{m=0}^{j} G_0^{j-m}\vec{\mathbf{w}}_m, \qquad j=0,1,\ldots \tag{47}$$

Now, only one function of G_0 appears in the solution. However, the computation of the $\vec{\mathbf{v}}_j$ vectors is still an expensive operation—the computation of the $M+1$ vectors involves M sums, which require $O(M^2)$ matrix–vector multiplications. The computational effort can be much reduced when we notice that the $\vec{\mathbf{v}}_j$'s satisfy a recursion relation:

$$\vec{\mathbf{v}}_j = G_0\vec{\mathbf{v}}_{j-1} + \vec{\mathbf{w}}_j, \qquad j=1,2,\ldots \tag{48}$$

Using Eq. (48), all the $\vec{\mathbf{v}}_j$'s can be computed by M matrix–vector multiplications only, starting from

$$\vec{\mathbf{v}}_0 = \vec{\mathbf{w}}_0 = \vec{\mathbf{u}}_0 \tag{49}$$

Taking into account also the evaluation of the first term in Eq. (46), we can conclude that the overall computational cost is reduced to $M+K-1$ matrix–vector multiplications for the Chebyshev or Newton series approximation methods (see Sec. 2.3.1). Similarly, $M+K$ matrix–vector multiplications are required for the Arnoldi approach.

2.4. Approximated solutions based on closed integrated forms

2.4.1. Source term with an arbitrary time-dependence

Let us consider the case of Eq. (18) with a source term $\vec{\mathbf{s}}(t)$ with an arbitrary time-dependence. In general, the integral (21) does not assume a closed form, so a closed solution for Eq. (18) cannot be obtained. In the present approach, we utilize the closed solution for the case of (22) in order to approximate the solution in the general case. The idea is to approximate the general source term by a truncated polynomial series of the form of (22). Then, the solution is approximated by a direct application of Eq. (46).

The approximation of $\vec{\mathbf{s}}(t)$ by the form of (22) requires the computation of the $\vec{\mathbf{s}}_m$ coefficients. The form of Eq. (22) might suggest that we should set $\vec{\mathbf{s}}_m = d^m\vec{\mathbf{s}}(0)/dt^m$, to yield a truncated Taylor series of $\vec{\mathbf{s}}(t)$. However, as has already been mentioned, a Taylor series is a poor tool for approximation purposes. Thus, this approach is not recommended.

A better approach is to approximate $\vec{\mathbf{s}}(t)$ by an orthogonal polynomial set at the first stage:

$$\vec{\mathbf{s}}(t) \approx \sum_{n=0}^{M-1}\vec{\mathbf{b}}_n P_n(t) \tag{50}$$

where the $\vec{\mathbf{b}}_n$'s are computed by a scalar product of the $P_n(t)$'s with $\vec{\mathbf{s}}(t)$. The orthogonal expansion polynomials can be expressed in the terms of the Taylor polynomials:

$$P_n(t) = \sum_{m=0}^{n} q_{n,m} \frac{t^m}{m!} \tag{51}$$

Plugging Eq. (51) into Eq. (50) we obtain:

$$\vec{\mathbf{s}}(t) \approx \sum_{n=0}^{M-1} \sum_{m=0}^{n} q_{n,m} \vec{\mathbf{b}}_n \frac{t^m}{m!} = \sum_{m=0}^{M-1} \left(\sum_{n=m}^{M-1} q_{n,m} \vec{\mathbf{b}}_n \right) \frac{t^m}{m!} \tag{52}$$

Then, the result is equated to the Taylor form,

$$\sum_{m=0}^{M-1} \left(\sum_{n=m}^{M-1} q_{n,m} \vec{\mathbf{b}}_n \right) \frac{t^m}{m!} = \sum_{m=0}^{M-1} \frac{t^m}{m!} \vec{\mathbf{s}}_m \tag{53}$$

to yield the $\vec{\mathbf{s}}_m$ Taylor polynomial coefficients:

$$\vec{\mathbf{s}}_m = \sum_{n=m}^{M-1} q_{n,m} \vec{\mathbf{b}}_n \tag{54}$$

These are in general different from the Taylor expansion coefficients. In this way, we preserve the Taylor polynomial form of Eq. (22), but with the advantage of the fast convergence of an orthogonal polynomial set.

It is recommended to use the Chebyshev polynomials as the $P_n(t)$ set. An equivalent option is to use a Newton interpolation expansion in the Chebyshev points.

We still need a procedure for a systematic computation of the $q_{n,m}$ coefficients. Recursive algorithms can be derived from recursive definitions of different polynomial sets. In Appendix C we develop recursive conversion algorithms from Chebyshev and Newton expansions to a Taylor form.

Of course, the expansion (50) should not be confused with the similar expansion (15). The first approximates the function $\vec{\mathbf{s}}(t)$ in *time*, within the time interval of the solution, while the second approximates a function of the matrix G_0 in the *eigenvalue domain of* G_0.

2.4.2. Time-dependent Hamiltonian

Now we shall consider the case of a time-dependent Hamiltonian, $H = H(t)$. For the sake of generality, a source term is included in the equation. We have:

$$\frac{d\vec{\mathbf{u}}(t)}{dt} = -iH(t)\vec{\mathbf{u}}(t) + \vec{\mathbf{s}}(t) \tag{55}$$

or,

$$\frac{d\vec{\mathbf{u}}(t)}{dt} = G(t)\vec{\mathbf{u}}(t) + \vec{\mathbf{s}}(t) \tag{56}$$

In this case, the Duhamel principle cannot be applied directly for obtaining a closed form solution, as in the previous cases (see Sec. 2.3.2). However, we shall see that the results from the previous cases can be utilized for obtaining a procedure which approximates the solution in the present case.

First, it is always possible to split $G(t)$ into a sum of time-dependent and time-independent parts:

$$G(t) = \widetilde{G} + \bar{G}(t) \tag{57}$$

where $\widetilde{G}$ is arbitrary, and

$$\bar{G}(t) \equiv G(t) - \widetilde{G} \tag{58}$$

Let us define:

$$\vec{\mathbf{s}}_{ext}(\vec{\mathbf{u}}(t), t) = \vec{\mathbf{s}}(t) + \bar{G}(t)\vec{\mathbf{u}}(t) \tag{59}$$

$\vec{\mathbf{s}}_{ext}(\vec{\mathbf{u}}(t), t)$ is a new, extended "source term". Now, Eq. (56) can be written as

$$\frac{d\vec{\mathbf{u}}(t)}{dt} = \widetilde{G}\vec{\mathbf{u}}(t) + \vec{\mathbf{s}}_{ext}(\vec{\mathbf{u}}(t), t) \tag{60}$$

which resembles the form of Eq. (18). The Duhamel principle can be applied to yield:

$$\vec{\mathbf{u}}(t) = \exp(\widetilde{G}t)\vec{\mathbf{u}}_0 + \exp(\widetilde{G}t) \int_0^t \exp(-\widetilde{G}\tau)\vec{\mathbf{s}}_{ext}(\vec{\mathbf{u}}(\tau), \tau)\, d\tau \tag{61}$$

As in the case of Sec. 2.4.1, we can write an approximation of this equality in the form of Eq. (46):

$$\vec{\mathbf{u}}(t) \approx f_M(\widetilde{G}, t)\vec{\mathbf{v}}_M + \sum_{j=0}^{M-1} \frac{t^j}{j!}\vec{\mathbf{v}}_j \tag{62}$$

The $\vec{\mathbf{v}}_j$ vectors are computed by expanding $\vec{\mathbf{s}}_{ext}$ in time in the form of (22), and using the resulting coefficients, as in Sec. 2.4.1.

Apparently, this gives nothing—Eq. (61) is an integral equation, and the RHS includes a dependence on $\vec{\mathbf{u}}(t)$ itself, which is still unknown. Consequently, the $\vec{\mathbf{v}}_j$'s also depend on $\vec{\mathbf{u}}(t)$. However, it is possible to utilize this form for obtaining a solution by an *iterative procedure*. First, we *guess* a solution $\vec{\mathbf{u}}_g(t)$ in the desired time interval. Then, we use $\vec{\mathbf{u}}_g(t)$ for the computation of the RHS of the equation. We obtain for the LHS a new approximated solution. It seems reasonable that it should be closer to the actual solution than $\vec{\mathbf{u}}_g(t)$. We can use the improved solution for obtaining a better one by inserting it into the RHS, and so on. This procedure can be continued until the solution converges with a desired accuracy.

This iterative scheme sounds reasonable. However, we have not given a rigorous justification to it. Thus, one might suspect if it should actually work. Experience shows that this iterative process does converge to the solution, given that *the time-interval is sufficiently small*. Thus, the iterative procedure has a *convergence radius*. The size of the convergence radius is problem dependent. When the time interval is larger than the convergence radius, the solution diverges, i.e. it tends to infinity with the number of iterations.

A more rigorous justification to the iterative procedure can be obtained by a convergence analysis. This topic is left for a future paper.

In the case that the time of the desired solution is outside the convergence interval of the algorithm, this procedure cannot be used directly. Instead, we can use a *time-step algorithm*, in a similar manner to the Taylor approach. We divide the time-interval into smaller time-steps, in which the iterative procedure converges. In each time-step we solve the sub-problem of obtaining the solution within the time-step. We use the solution obtained in order to compute $\vec{\mathbf{u}}_g(t)$ for the next time-step, as will be described later.

We see that at the end of the day we still need a time-step propagation, as in the Taylor approach. The advantage of the present approach is that we can use much larger time-steps, which means that the accumulation of errors and the computational effort can be much reduced. The approach for the computation of each time-step is global, replacing the local considerations of the Taylor approach. However, the algorithm still contains an obvious local element, in the sense that the solution is computed separately in each local time-step. Hence, we can call this approach a *"semi-global approach"*.

Note that the definition of the $\vec{\mathbf{s}}_j$'s, the $\vec{\mathbf{w}}_j$'s and corresponding $\vec{\mathbf{v}}_j$'s is different for each time-step. Let us denote the k'th time-point by t_k. The time-interval of the k'th time-step is $[t_k, t_{k+1}]$. The $\vec{\mathbf{s}}_j$'s, the $\vec{\mathbf{w}}_j$'s and the $\vec{\mathbf{v}}_j$'s in the k'th time-step will be denoted by $\vec{\mathbf{s}}_{k,j}$, $\vec{\mathbf{w}}_{k,j}$ and $\vec{\mathbf{v}}_{k,j}$, accordingly. The $\vec{\mathbf{s}}_{k,j}$'s are computed using the expansion (50) of $\vec{\mathbf{s}}_{ext}(t)$ *within the k'th time-interval*. The $\vec{\mathbf{w}}_{k,j}$'s are defined as:

$$\vec{\mathbf{w}}_{k,j} \equiv \begin{cases} \vec{\mathbf{u}}(t_k) & j = 0 \\ \vec{\mathbf{s}}_{k,j-1} & 0 < j \le M \end{cases} \tag{63}$$

The $\vec{\mathbf{v}}_{k,j}$'s are computed accordingly by the recursion

$$\begin{aligned} &\vec{\mathbf{v}}_{k,0} = \vec{\mathbf{u}}(t_k) \\ &\vec{\mathbf{v}}_{k,j} = \widetilde{G}\vec{\mathbf{v}}_{k,j-1} + \vec{\mathbf{w}}_{k,j}, \qquad j = 1, 2, \ldots \end{aligned} \tag{64}$$

The solution within the k'th time-step is:

$$\vec{\mathbf{u}}(t) = f_M(\widetilde{G}, t - t_k)\vec{\mathbf{v}}_{k,M} + \sum_{j=0}^{M-1} \frac{(t - t_k)^j}{j!}\vec{\mathbf{v}}_{k,j}, \qquad t \in [t_k, t_{k+1}] \tag{65}$$

Two questions remained open: How $G(t)$ should be split (see Eq. (57)), and how the guess solution $\vec{\mathbf{u}}_g(t)$ should be chosen. We begin with the first question. From a physical point of view, it seems that a natural choice in many problems is to split the Hamiltonian in the following way:

$$H(t) = H_0 + V(t) \tag{66}$$

where H_0 is the unperturbed Hamiltonian and $V(t)$ is a time-dependent perturbation. If, in addition, $\vec{\mathbf{s}}(t) \equiv 0$, Eq. (61) becomes

$$\vec{\mathbf{u}}(t) = \exp(-iH_0 t)\vec{\mathbf{u}}_0 - i\int_0^t \exp[-iH_0(t-\tau)]V(\tau)\vec{\mathbf{u}}(\tau)\,d\tau \tag{67}$$

which has a striking resemblance to the well-known expression of the first-order time-dependent perturbation theory (see, for example, [11, Chapter XIII]). Indeed, the expressions become identical by the replacement of $\vec{\mathbf{u}}(\tau)$ in the integral by $\vec{\mathbf{u}}_0$. However, although this option is appealing in the sense of the directness of the physical interpretation, it needn't be the best option from a numerical point of view. The result may converge faster with other choices of splitting.

A more educated choice of splitting comes to us when we realize that the weak point of the algorithm lies in the point where we "cheat". This point is the treatment of the $\vec{\mathbf{u}}(t)$ dependent "source term", $\vec{\mathbf{s}}_{ext}(\vec{\mathbf{u}}(t),t)$, as an inhomogeneous, $\vec{\mathbf{u}}(t)$ independent term. We should choose the splitting in a way that minimizes the size of the $\vec{\mathbf{u}}(t)$ dependence in $\vec{\mathbf{s}}_{ext}(\vec{\mathbf{u}}(t),t)$. Hence, $\bar{G}(t)$ should be as small as possible. Consider the k'th time-step, in the time-interval $[t_k,t_{k+1}]$. For the sake of generality, we consider also a non-equidistant time-grid. Let us denote: $\Delta t_k = t_{k+1}-t_k$. Usually, Δt_k can be assumed to be small, by the requirements of the convergence radius of the algorithm. Hence, we can assume that $G(t)$ does not change much during the time-interval. Then it becomes reasonable to choose the following splitting:

$$\widetilde{G} = G\left(t_k+\frac{\Delta t_k}{2}\right) \tag{68}$$

$$\bar{G}(t) \equiv G(t) - G\left(t_k+\frac{\Delta t_k}{2}\right) \qquad t\in[t_k,t_{k+1}] \tag{69}$$

Obviously, the splitting of Eqs. (68)–(69) is time-step dependent, unlike the splitting of Eq. (66).

The choice of the guess solution $\vec{\mathbf{u}}_g(t)$ is lead by two contradicting considerations: On one hand, we need a sufficiently accurate starting point for the iterative process. On the other hand, we require that it can be obtained with minimal amount of extra numerical effort. One obvious choice, which requires no extra numerical effort, is the zero'th order approximation—within the k'th time-step interval, $[t_k,t_{k+1}]$, the guess solution is

$$\vec{\mathbf{u}}_g(t) \equiv \vec{\mathbf{u}}(t_k) \qquad t\in[t_k,t_{k+1}] \tag{70}$$

However, this approximation is of low accuracy. Actually, a very accurate approximation can be obtained from an *extrapolation* of the solution in the previous time-step. All we need to do is to use Eq. (65) for the previous time-step to compute the solution in the current time-step. We obtain the following approximated guess solution:

$$\vec{\mathbf{u}}_g(t) = f_M(\widetilde{G},t-t_{k-1})\vec{\mathbf{v}}_{k-1,M} + \sum_{j=0}^{M-1}\frac{(t-t_{k-1})^j}{j!}\vec{\mathbf{v}}_{k-1,j}, \qquad t\in[t_k,t_{k+1}] \tag{71}$$

This solution approximates the solution in the interval $[t_k,t_{k+1}]$, using information from the previous interval $[t_{k-1},t_k]$. Note that the second argument of the function $f_M(z,t)$ in Eq. (71) represents a different time interval from that of Eq. (65). As in Eq. (17), the function is expanded in z, and t serves as a parameter. Hence, the functions to be computed are different in the two equations, and new expansion coefficients have to be computed in the new interval. The numerical effort of this operation is negligible in comparison to the matrix–vector multiplications (or the equivalent linear operations), unless the dimension N of the problem is very small. Thus, by Eq. (71) we obtain an accurate guess with a relatively low computational cost.

In the first time-step, $\vec{\mathbf{u}}_g(t)$ can be computed by Eq. (70). More iterations will be required in comparison with the other time-steps, but the overall additional computational effort in a many time-point grid is negligible.

2.4.3. Nonlinear Hamiltonian

Let the Hamiltonian include also a dependence on $\vec{\mathbf{u}}(t)$, i.e. $H\equiv H(\vec{\mathbf{u}}(t),t)$. Now we get to the general case of Eq. (5), or equivalently, Eq. (7). The treatment of this case is completely analogous to that of the linear time-dependent Hamiltonian.

Let us split $G(\vec{\mathbf{u}}(t),t)$ in the following way:

$$G(\vec{\mathbf{u}}(t),t) = \widetilde{G} + \bar{G}(\vec{\mathbf{u}}(t),t) \tag{72}$$

where $\widetilde{G}$ is linear and time-independent. Let us define:

$$\vec{\mathbf{s}}_{ext}(\vec{\mathbf{u}}(t),t) = \vec{\mathbf{s}}(t) + \bar{G}(\vec{\mathbf{u}}(t),t)\vec{\mathbf{u}}(t) \tag{73}$$

The rest of the algorithm is identical to that of Sec. 2.4.2, for the same considerations.

In an analogous manner to the time-dependent linear case, it is recommended to choose the following splitting in the time-step algorithm:

$$\widetilde{G} = G\left[\vec{\mathbf{u}}\left(t_k+\frac{\Delta t_k}{2}\right),t_k+\frac{\Delta t_k}{2}\right] \tag{74}$$

$$\bar{G}(\vec{\mathbf{u}}(t),t) \equiv G(\vec{\mathbf{u}}(t),t) - G\left[\vec{\mathbf{u}}\left(t_k+\frac{\Delta t_k}{2}\right),t_k+\frac{\Delta t_k}{2}\right] \qquad t\in[t_k,t_{k+1}] \tag{75}$$

3. Implementation

3.1. The propagation time-grid

In order to obtain the $\vec{\mathbf{s}}_m$ Taylor like coefficients, we have to know the total time-dependence of $\vec{\mathbf{s}}_{ext}(\vec{\mathbf{u}}(t),t)$ (see Sec. 2.4.1). Of course, we have no explicit expression for the total time-dependence of $\vec{\mathbf{s}}_{ext}(\vec{\mathbf{u}}(t),t)$, because of its dependence on $\vec{\mathbf{u}}(t)$. Hence, the time-dependence of $\vec{\mathbf{s}}_{ext}(\vec{\mathbf{u}}(t),t)$ has to be approximated from several *sampling points* within the time interval. In each sampling point t_l, we need the values of $\vec{\mathbf{u}}(t_l)$, $G(\vec{\mathbf{u}}(t_l),t_l)$ and $\vec{\mathbf{s}}(t_l)$ in order to compute $\vec{\mathbf{s}}_{ext}(\vec{\mathbf{u}}(t_l),t_l)$. The $\vec{\mathbf{b}}_n$'s from Eq. (50) are obtained from the samplings of $\vec{\mathbf{s}}_{ext}(\vec{\mathbf{u}}(t),t)$ within the time-interval. Then, the $\vec{\mathbf{s}}_m$'s can be computed from the $\vec{\mathbf{b}}_n$'s as described in Sec. 2.4.1.

It is recommended to choose the *Chebyshev points* within the time-interval as the sampling points. Then, $\vec{\mathbf{s}}_{ext}(\vec{\mathbf{u}}(t),t)$ can be expanded in time either by a Chebyshev polynomial expansion, or by a Newton interpolation at the Chebyshev points (see Appendix A). When a Chebyshev polynomial expansion is used, the $\vec{\mathbf{b}}_n$'s from Eq. (50) correspond to the Chebyshev coefficients, that will be denoted by $\vec{\mathbf{c}}_n$, and the polynomials are the Chebyshev polynomials. When a Newton interpolation is used, the $\vec{\mathbf{b}}_n$'s are the divided differences, that will be denoted by $\vec{\mathbf{a}}_n$, and the polynomials are the Newton basis polynomials. In Appendix A we describe how the $\vec{\mathbf{a}}_n$'s or the $\vec{\mathbf{c}}_n$'s can be obtained from the samplings at the Chebyshev points.

In the time-step algorithm, the time interval of each time-step is sampled at the Chebyshev points of the interval. Thus, the structure of the time-grid necessary for the propagation is complex; it consists of adjacent time-intervals, each with an internal Chebyshev grid. In order to refer also to the internal grid of each interval, we shall replace the single index notation of the time-grid from Sec. 2.4.2, t_k, by a double index notation, $t_{k,l}$. The first index k refers to the k'th time-interval, where $k=1,2,\ldots,N_t$. The second index l indexes the points in the internal Chebyshev grid of each interval, as will be readily seen.

The length of the k'th time-interval is denoted by Δt_k, as in Sec. 2.4.2. For the sake of generality, we will consider also the possibility that the number of Chebyshev points is different for each time-step. The number of Chebyshev points in the k'th interval is denoted by M_k. M_k is also the number of expansion terms for the approximation of $\vec{\mathbf{s}}_{ext}(\vec{\mathbf{u}}(t),t)$ (see Eq. (50)).

We use the set of boundary including Chebyshev points. In the Chebyshev domain, $[-1,1]$, the Chebyshev grid is defined as follows (cf. Appendix A.1.2, Eq. (117); note that the equations of Appendix A are formulated in the terms of the order of the polynomial approximation, which is equivalent to M_k-1):

$$y_{k,l} \equiv -\cos\left(\frac{l\pi}{M_k-1}\right), \qquad l=0,1,\ldots,M_k-1 \tag{76}$$

In the k'th time-step domain, the Chebyshev points become (cf. Eq. (118)):

$$t_{k,l} = t_{k,0} + \frac{\Delta t_k}{2}(1+y_{k,l}) \tag{77}$$

Note that we have:

$$t_{k,M_k-1} = t_{k,0} + \Delta t_k = t_{k+1,0} \tag{78}$$

Eqs. (77), (78) define together the entire time grid, where $t_{1,0}$ and the Δt_k's are given.

3.2. Algorithm

We assume that the initial condition, $\vec{\mathbf{u}}(t_{1,0})$, is given. In addition, it is assumed that the structure of the propagation time-grid (i.e. Δt_k and M_k for each time-step) is known in advance. Alternatively, it is possible to choose it adaptively during the propagation, by an internal procedure. The number of expansion terms for the approximation of $f_{M_k}(\widetilde{G},t-t_{k,0})\vec{\mathbf{v}}_{M_k}$ (see Eq. (65)) is also supplied by the user. It may depend on k.

The accuracy of the solution is determined by a tolerance parameter, which will be denoted by ϵ. It represents the order of the accepted relative error of the solution. The tolerance parameter is supplied by the user.

The scheme of propagation goes as follows:

1. Set the guess solution of the first time-step in the internal Chebyshev grid (cf. Eq. (70)):

   $$\vec{\mathbf{u}}(t_{1,l}) = \vec{\mathbf{u}}(t_{1,0}), \qquad l=0,1,\ldots,M_1-1$$

2. for $k=1$ to N_t
   (a) Set the middle point of the internal grid, $t_{mid}=t_{k,M_k\backslash 2}$, where $\backslash$ denotes integer division.
   (b) $(l=0,1,\ldots,M_k-1, \qquad n=0,1,\ldots,M_k-1, \qquad j=0,1,\ldots,M_k)$
   (c) do
      i. Set $\vec{\mathbf{s}}^l_{ext} = \vec{\mathbf{s}}(t_{k,l}) + [G(\vec{\mathbf{u}}(t_{k,l}),t_{k,l}) - G(\vec{\mathbf{u}}(t_{mid}),t_{mid})]\vec{\mathbf{u}}(t_{k,l})$ (cf. Eqs. (73), (75)).
      ii. Use the $\vec{\mathbf{s}}^l_{ext}$'s to compute the expansion coefficients of $\vec{\mathbf{s}}_{ext}(\vec{\mathbf{u}}(t),t)$ in the time-step.

- *For a Newton interpolation*: Compute the divided differences $\vec{\mathbf{a}}_n$ recursively, as described in Appendix A.1 (relevant equations: (109), (110), (113)). Use $4t_{k,l}/\Delta t_k$ as the sampling points (see Sec. A.1.3), and the corresponding $\vec{\mathbf{s}}^l_{ext}$'s as the function values.
- *For a Chebyshev expansion*: Compute the Chebyshev coefficients $\vec{\mathbf{c}}_n$, as described in Appendix A.2 (relevant equation: (148)). Use the $\vec{\mathbf{s}}^l_{ext}$'s as the function values.

iii. Compute the $\vec{\mathbf{s}}_n$ Taylor-like coefficients recursively from the $\vec{\mathbf{a}}_n$'s or the $\vec{\mathbf{c}}_n$'s, using the conversion schemes described in Appendix C (relevant equations: (223)–(225), (215) for the $\vec{\mathbf{a}}_n$'s, (246)–(250) for the $\vec{\mathbf{c}}_n$'s).

iv. Compute the $\vec{\mathbf{v}}_j$ vectors recursively from $\vec{\mathbf{u}}_{k,0}$ and the $\vec{\mathbf{s}}_n$'s (see Eqs. (64), (63)), where $\widetilde{G} = G(\vec{\mathbf{u}}(t_{mid}), t_{mid})$ (cf. Eq. (74)).

v Store the current solution at the time-step edge for convergence check, $\vec{\mathbf{u}}_{old} = \vec{\mathbf{u}}(t_{k,M_k-1})$.

vi. Compute a new solution from the $\vec{\mathbf{v}}_j$'s by the expression (cf. Eq. (65)):

$$\vec{\mathbf{u}}(t_{k,l}) = f_{M_k}(\widetilde{G}, t_{k,l} - t_{k,0})\vec{\mathbf{v}}_{M_k} + \sum_{j=0}^{M_k-1} \frac{(t_{k,l} - t_{k,0})^j}{j!}\vec{\mathbf{v}}_j$$

$f_{M_k}(\widetilde{G}, t_{k,l} - t_{k,0})\vec{\mathbf{v}}_{M_k}$ is approximated by one of the methods described in Appendix B (note that the expansion vectors required for the approximation are computed just once for all time points—see Sec. 2.3.1).

vii. Repeat from step 2(c)i while

$$\frac{\|\vec{\mathbf{u}}(t_{k,M_k-1}) - \vec{\mathbf{u}}_{old}\|}{\|\vec{\mathbf{u}}_{old}\|} > \epsilon$$

(d) end do

(e) Compute the solution at any desired point $t_p \in [t_{k,0}, t_{k,M_k-1}]$ by

$$\vec{\mathbf{u}}(t_p) = f_{M_k}(\widetilde{G}, t_p - t_{k,0})\vec{\mathbf{v}}_{M_k} + \sum_{j=0}^{M_k-1} \frac{(t_p - t_{k,0})^j}{j!}\vec{\mathbf{v}}_j$$

(f) Set the guess solution for the next time-step; by definition: $\vec{\mathbf{u}}(t_{k+1,0}) = \vec{\mathbf{u}}(t_{k,M_k-1})$ (see Eq. (78)). The guess solution at the rest of the Chebyshev internal points is computed by (cf. Eq. (71)):

$$\vec{\mathbf{u}}(t_{k+1,m}) = f_{M_k}(\widetilde{G}, t_{k+1,m} - t_{k,0})\vec{\mathbf{v}}_{M_k} + \sum_{j=0}^{M_k-1} \frac{(t_{k+1,m} - t_{k,0})^j}{j!}\vec{\mathbf{v}}_j,$$

$$m = 1, \ldots, M_{k+1} - 1$$

3. end for

The algorithm is sketched schematically in Fig. 1.

3.3. Programming

3.3.1. Numerical stability of the time polynomial expansion

In Eqs. (22), (51), the coefficients of the t polynomials are defined as the coefficients of $t^m/m!$, in analogy to the Taylor expansion form. Accordingly, we obtained in Eq. (46) the $\vec{\mathbf{v}}_m$'s as the coefficients of $t^m/m!$. The $1/m!$ factor decreases very fast as m grows. Consequently, the $\vec{\mathbf{s}}_m$'s, the $q_{n,m}$'s and the $\vec{\mathbf{v}}_m$'s tend to attain huge values. This may lead to numerical instability.

The problem can be solved by the definition of alternative polynomial expansions, in which the coefficients absorb the $1/m!$ factor. The alternative expansions will lead to expressions which are more stable numerically. The source term $\vec{\mathbf{s}}(t)$ is expanded as

$$\vec{\mathbf{s}}(t) \approx \sum_{m=0}^{M-1} t^m \vec{\tilde{\mathbf{s}}}_m \tag{79}$$

where $\vec{\tilde{\mathbf{s}}}_m = \vec{\mathbf{s}}_m/m!$. Accordingly, Eq. (51) is replaced by

$$P_n(t) = \sum_{m=0}^{n} \tilde{q}_{n,m} t^m \tag{80}$$

where $\tilde{q}_{n,m} = q_{n,m}/m!$.

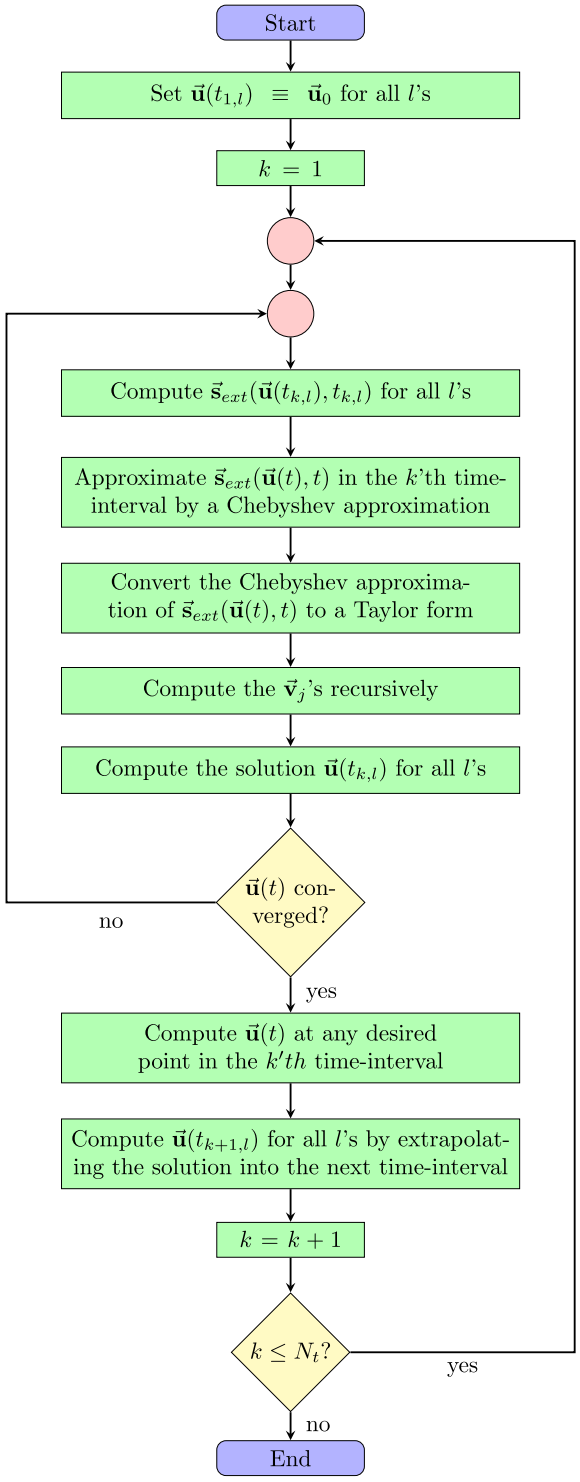

Fig. 1. The semi-global propagator algorithm.

First we derive the solution equation for a time-independent Hamiltonian. We plug the expansion (79) into Eq. (20). The derivation of the solution is similar to that of Sec. 2.3.2, but less appealing from an aesthetic point of view. We end with the following equation:

$$\vec{\mathbf{u}}(t) = \tilde{f}_M(G_0, t)\vec{\tilde{\mathbf{v}}}_M + \sum_{j=0}^{M-1} t^j \vec{\tilde{\mathbf{v}}}_j \tag{81}$$

where we defined

$$\tilde{f}_m(z,t) \equiv \begin{cases} \frac{m!}{z^m}\left[\exp(zt) - \sum_{j=0}^{m-1} \frac{(zt)^j}{j!}\right] & z \neq 0 \\ t^m & z = 0 \end{cases} \qquad m = 0, 1, \ldots \tag{82}$$

The $\vec{\tilde{\mathbf{v}}}_j$'s are defined recursively in the following way:

$$\vec{\tilde{\mathbf{v}}}_0 = \vec{\mathbf{u}}_0$$
$$\vec{\tilde{\mathbf{v}}}_j = \frac{G_0 \vec{\tilde{\mathbf{v}}}_{j-1} + \vec{\tilde{\mathbf{s}}}_{j-1}}{j}, \qquad j = 1, 2, \ldots \tag{83}$$

Note that we have:

$$\tilde{f}_m(z,t) = m! f_m(z,t) \tag{84}$$
$$\vec{\tilde{\mathbf{v}}}_j = \frac{\vec{\mathbf{v}}_j}{j!} \tag{85}$$

Thus, it can be easily verified that Eq. (81) is equivalent to Eq. (46).

In the case of a time-dependent nonlinear Hamiltonian, we expand $\vec{\mathbf{s}}_{ext}(t)$ as in Eq. (79), and replace G_0 in Eqs. (81), (83), by $\widetilde{G}$.

The computation of the $\tilde{q}_{n,m}$'s is completely analogous to that of the $q_{n,m}$'s. The procedure is given in Appendix C.

In summary, in order to improve the stability of the program, the following changes should be made in the algorithm:

- In step 2(c)iii, the $\vec{\tilde{\mathbf{s}}}_n$'s are computed instead of the $\vec{\mathbf{s}}_n$'s, via the computation of the $\tilde{q}_{n,m}$'s (relevant equations: (226)–(228), (221) for the $\vec{\mathbf{a}}_n$'s, (254)–(258) for the $\vec{\mathbf{c}}_n$'s).
- In step 2(c)iv, the $\vec{\tilde{\mathbf{v}}}_j$'s are computed instead of the $\vec{\mathbf{v}}_j$'s, by the recursion

$$\vec{\tilde{\mathbf{v}}}_0 = \vec{\mathbf{u}}_0$$
$$\vec{\tilde{\mathbf{v}}}_j = \frac{\widetilde{G}\vec{\tilde{\mathbf{v}}}_{j-1} + \vec{\tilde{\mathbf{s}}}_{j-1}}{j}, \qquad j = 1, 2, \ldots$$

  where $\widetilde{G} = G(\vec{\mathbf{u}}(t_{mid}), t_{mid})$.
- In steps 2(c)vi, 2e and 2f, the solution at the relevant time-points is computed by

$$\vec{\mathbf{u}}(t) = \tilde{f}_M(\widetilde{G}, t - t_{k,0})\vec{\tilde{\mathbf{v}}}_M + \sum_{j=0}^{M-1} (t - t_{k,0})^j \vec{\tilde{\mathbf{v}}}_j \tag{86}$$

3.3.2. The computation of $\tilde{f}_m(z,t)$

The function $f_m(z,t)$, and its variant, $\tilde{f}_m(z,t)$, include the following expression (see Eqs. (35), (82)):

$$\exp(zt) - \sum_{j=0}^{m-1} \frac{(zt)^j}{j!} \tag{87}$$

The sum is just a truncated Taylor expansion of $\exp(zt)$. If zt is small, the difference between $\exp(zt)$ and its truncated expansion becomes extremely small. Often, this results in roundoff errors.

The problem can be solved by an alternative computation of $\tilde{f}_m(z,t)$ for small zt values. Let us expand $\exp(zt)$ from (87) by a Taylor expansion. (87) can be expressed as a "tail" of the Taylor expansion in the following way:

$$\exp(zt) - \sum_{j=0}^{m-1} \frac{(zt)^j}{j!} = \sum_{j=0}^{\infty} \frac{(zt)^j}{j!} - \sum_{j=0}^{m-1} \frac{(zt)^j}{j!} = \sum_{j=m}^{\infty} \frac{(zt)^j}{j!} \tag{88}$$

The expression for $\tilde{f}_m(z,t)$ becomes:

$$\tilde{f}_m(z,t) = m! t^m \sum_{j=0}^{\infty} \frac{(zt)^j}{(j+m)!} \tag{89}$$

$\tilde{f}_m(z,t)$ can be computed by truncating the sum in (89). The Taylor expansion converges very slowly. Hence, the expansion should be truncated only after achieving the machine accuracy in the summation procedure. The form (89) should not be used when zt is large enough to be computed directly.

3.3.3. Efficiency of the computation of $\vec{\mathbf{s}}_{ext}(\vec{\mathbf{u}}(t),t)$

The number of required matrix–vector multiplications for the computation of the integrated form (46) is $M+K-1$ for a polynomial approximation of the function of matrix, and $M+K$ for the Arnoldi approach, cf. Sec. 2.3.2. The computation of $\vec{\mathbf{s}}_{ext}(\vec{\mathbf{u}}(t_{k,l}),t_{k,l})$ in the general case (step 2(c)i in the algorithm) seems to cost considerable amount of additional computational effort. It can be readily seen from step 2(c)i that a direct computation of each of the $\vec{\mathbf{s}}^{l}_{ext}$'s requires a subtraction of two matrices, and a matrix–vector multiplication. Subtraction of matrices has the same scaling as a matrix–vector multiplication, $O(N^2)$. An alternative, which is less time-consuming, is to perform the computation as $\vec{\mathbf{s}}^{l}_{ext}=\vec{\mathbf{s}}(t_{k,l})+G(\vec{\mathbf{u}}(t_{k,l}),t_{k,l})\vec{\mathbf{u}}(t_{k,l})-G(\vec{\mathbf{u}}(t_{mid}),t_{mid})\vec{\mathbf{u}}(t_{k,l})$. This requires two matrix–vector multiplications for each l. This is with the exception of $l=M_k\backslash 2$, which indexes the middle internal time-point, t_{mid}; the extended part of the inhomogeneous term vanishes, and we are left with $\vec{\mathbf{s}}^{M_k\backslash 2}_{ext}=\vec{\mathbf{s}}(t_{mid})$. Thus, the computation in the middle point does not involve additional expensive operations. The overall additional cost of step 2(c)i is $2(M_k-1)$ matrix–vector multiplications. This roughly doubles the computational effort.

Most frequently, the computational effort can be considerably reduced by a proper formulation of the calculation. First, in many problems, the operator represented by $G(\vec{\mathbf{u}}(t_{k,l}),t_{k,l})-G(\vec{\mathbf{u}}(t_{mid}),t_{mid})$ is diagonal in the basis of representation. Thus, the scaling of its operation on $\vec{\mathbf{u}}(t_{k,l})$ becomes linear, $O(N)$. The computational cost of this operation is negligible. A common example is a Hamiltonian which is composed from a stationary part and a time-dependent nonlinear potential,

$$H(\vec{\mathbf{u}}(t),t)=H_0+V(\vec{\mathbf{u}}(t),t) \tag{90}$$

In this case we have:

$$[G(\vec{\mathbf{u}}(t_{k,l}),t_{k,l})-G(\vec{\mathbf{u}}(t_{mid}),t_{mid})]\vec{\mathbf{u}}(t_{k,l})=-i[V(\vec{\mathbf{u}}(t_{k,l}),t_{k,l})-V(\vec{\mathbf{u}}(t_{mid}),t_{mid})]\vec{\mathbf{u}}(t_{k,l}) \tag{91}$$

When the problem is represented in the spatial basis, the potential becomes diagonal. Thus, the computation of $\vec{\mathbf{s}}_{ext}(\vec{\mathbf{u}}(t_{k,l}),t_{k,l})$ does not require any additional expensive operation.

Moreover, in many situations, the dependence of $G(\vec{\mathbf{u}}(t),t)$ on $\vec{\mathbf{u}}(t)$ and t is determined by a small number of parameters. This may reduce the required computational cost. For example, let us consider a Hamiltonian of the form of (90) with a potential

$$V(\vec{\mathbf{u}}(t),t)=\zeta(\vec{\mathbf{u}}(t),t)\mu \tag{92}$$

where $\zeta(\vec{\mathbf{u}}(t),t)$ is a scalar parameter and μ is a matrix. $\zeta(\vec{\mathbf{u}}(t),t)$ may represent an electric or magnetic field (up to a sign), and μ may represent the electric or magnetic moment operator, respectively (state dependent electric or magnetic fields occur, e.g., in coherent control problems, when the Krotov algorithm is employed; see [52] for a review). We have:

$$[G(\vec{\mathbf{u}}(t_{k,l}),t_{k,l})-G(\vec{\mathbf{u}}(t_{mid}),t_{mid})]\vec{\mathbf{u}}(t_{k,l})=-i[\zeta(\vec{\mathbf{u}}(t_{k,l}),t_{k,l})-\zeta(\vec{\mathbf{u}}(t_{mid}),t_{mid})]\,\mu\,\vec{\mathbf{u}}(t_{k,l}) \tag{93}$$

We see that the computation of $\vec{\mathbf{s}}_{ext}(\vec{\mathbf{u}}(t_{k,l}),t_{k,l})$ requires just one matrix–vector multiplication (when μ is a non-diagonal matrix). Thus, the additional computational cost of the $\vec{\mathbf{s}}^{l}_{ext}$'s is just M_k-1 matrix–vector multiplications.

More generally, let us consider $G(\vec{\mathbf{u}}(t),t)$ which can be represented in the following form:

$$G(\vec{\mathbf{u}}(t),t)=G_0+\sum_{j=1}^{L}\xi_j(\vec{\mathbf{u}}(t),t)G_j \tag{94}$$

where the $\xi_j(\vec{\mathbf{u}}(t),t)$'s are scalar parameters, the G_j's are matrices, and $L\ll N^2$. We have:

$$[G(\vec{\mathbf{u}}(t_{k,l}),t_{k,l})-G(\vec{\mathbf{u}}(t_{mid}),t_{mid})]\vec{\mathbf{u}}(t_{k,l})=\sum_{j=1}^{L}[\xi_j(\vec{\mathbf{u}}(t_{k,l}),t_{k,l})-\xi_j(\vec{\mathbf{u}}(t_{mid}),t_{mid})]G_j\vec{\mathbf{u}}(t_{k,l}) \tag{95}$$

The cost of this operation is less than a single multiplication of a vector by $G(\vec{\mathbf{u}}(t),t)$. Since $L\ll N^2$, the computational cost of the expressions $\xi_j(\vec{\mathbf{u}}(t_{k,l}),t_{k,l})-\xi_j(\vec{\mathbf{u}}(t_{mid}),t_{mid})$ becomes negligible in comparison to a matrix–vector multiplication (note that the condition for L becomes different when the multiplication by the matrix is represented by an equivalent linear operation procedure with lower scaling than $O(N^2)$).

3.4. Parameter choice

Several free parameters are involved in the propagation algorithm. They need to be supplied by the user in each problem. The choice of the parameters determines the efficiency and the accuracy of the algorithm. With an inappropriate choice, the algorithm might become inefficient, inaccurate, or even completely fail. Of course, the parameters which yield good results are problem dependent.

There are two main criteria for a successful choice of the parameters:

1. The accuracy of the results;

2. The efficiency of the algorithm.

The treatment of the first criterion is closely related to the ability to estimate the error of the different approximations involved in the algorithm. This important topic is left to Appendix D, due to its length. The present discussion mainly focuses on the second criterion.

In general, a successful choice of parameters can be achieved by trial and error. The choice may improve with experience with the algorithm, and after the treatment of similar problems. Here we give several recommendations which are based on our experience with the algorithm.

The choice of the ϵ tolerance parameter (see step 2(c)vii) is obvious: It should be determined by the desired accuracy of the solution. Note that the tolerance parameter is defined for a single time-step. The error of the final solution is expected to accumulate during the propagation, roughly as the sum of the errors of each time-step.

In addition, there are three free parameters which need to be specified in each time-step:

1. The length of the time-step interval, Δt_k;
2. The number of expansion terms for the approximation of $\vec{\mathbf{s}}_{ext}(\vec{\mathbf{u}}(t), t)$, M_k;
3. The number of expansion terms for the approximation of $\tilde{f}_{M_k}(\widetilde{G}, t - t_{k,0})\vec{\tilde{\mathbf{v}}}_{M_k}$, K_k.

Our experience shows that the parameters should be chosen such that *a single iteration is required in each time-step*. In other words, the steps of the loop in stage 2c of the algorithm are performed just once. This is with the exception of the first time-step, in which the guess solution is of low accuracy (see Eq. (70)), and usually two or more iterations are required for a sufficient accuracy. The observation that the algorithm becomes most efficient with a minimal number of iterations is consistent with the reasoning that lead us to the choice of the $G(\vec{\mathbf{u}}(t), t)$ splitting (see Sec. 2.4.2)—*the weak point of the algorithm lies in the iterative process*. Hence, the part that the iterative process takes in the computation of an accurate solution should be minimized.

Typically, the values of M_k and K_k should be chosen to lie in the range 5–13. For higher values, even though Δt_k can be increased, the algorithm usually becomes less efficient. The guess solution for the next time-step begins to become less accurate for large M. This is due to the high sensitivity of a high order extrapolation to roundoff errors. A high order K is usually simply unnecessary. It should be noted that for high M orders, the algorithm may become numerically unstable. The source of the instability lies in the fact that both $(t - t_{k,0})^j$ and $\tilde{f}_j(\widetilde{G}, t - t_{k,0})$ in Eq. (86) typically become exceedingly small for high j's. Accordingly, the $\vec{\tilde{\mathbf{v}}}_j$'s attain very large values, and the computational process for obtaining them becomes unstable.

In the future, we plan to develop a version of the algorithm which is parameter free. In such an algorithm, the parameters are specified adaptively by the procedure during the propagation process, according to the accuracy requirements.

4. Numerical example

In the present section we test the efficiency of the semi-global propagator in solving a numerical problem of physical importance. Several numerical examples of relatively simple problems are already presented in Ref. [50]. In this paper, we test the performance of the propagator in a more realistic physical model problem. Moreover, for the first time, we demonstrate the capability of the algorithm to solve a problem with a *complex spectrum*, i.e. a problem in which the eigenvalue spectrum of G is distributed on the *complex plane*. This requires the use of the Arnoldi approach (see Sec. B.2) for the computation of $f_M(\widetilde{G}, t - t_k)\vec{\mathbf{v}}_{k,M}$ (see the procedure in Sec. 3.2, step 2(c)vi).

We choose a physical problem which is known to be challenging numerically—an atom subject to an intense laser field. This physical situation is characterized by extreme conditions for which an accurate numerical calculation becomes difficult. Under the influence of the intense field, a partial ionization of the atom occurs. The ionized part of the electron wave-function has the characteristics of an unbound particle, thus is spread to large spatial distances from the parent atom. Its dynamics is characterized by a strongly accelerated motion under the influence of the intense field. The central potential of the parent atom has a Coulomb character, which is steep in nature. These characteristics of the problem contribute to the difficulty of an accurate computation of the dynamics.

In order to give a reliable description of the problem, *absorbing boundary conditions* have to be employed. These are implemented here by a *complex absorbing potential*. This is the origin of the complex spectrum in our problem. The topic will be discussed in more detail in Sec. 4.2.

In Sec. 4.1 we present the physical details of the model problem. In Sec. 4.2 we present the details of the numerical implementation of the physical problem. In Sec. 4.3 the results are presented, and compared to a reference method.

4.1. The details of the physical problem

Remark: Atomic units are used throughout.

We use a one-dimensional model for the problem. The central potential of the parent atom is represented by a truncated Coulomb potential (see Fig. 2):

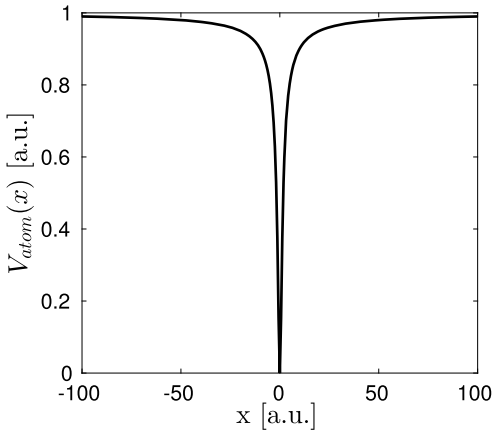

Fig. 2. The central atom model potential.

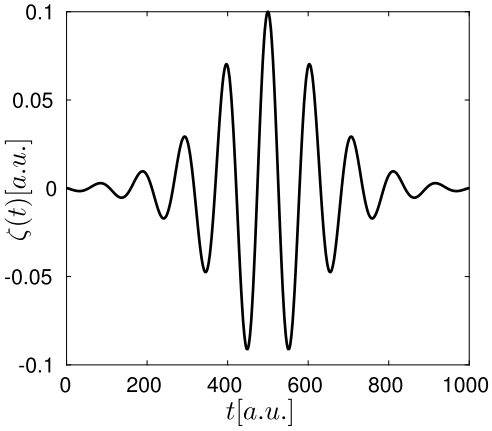

Fig. 3. The electric field.

$$V_{atom}(x) = 1 - \frac{1}{\sqrt{x^2+1}} \tag{96}$$

In this model, the singularity of the Coulomb potential at $x = 0$ is removed. The model has been extensively studied in the context of intense laser atomic physics. The fundamental energy difference, $\Delta E_1 = 0.395$ a.u., is similar to that of the hydrogen atom (0.375 a.u.).

The laser pulse electric field has the following form (see Fig. 3):

$$\zeta(t) = 0.1\,\mathrm{sech}^2\left(\frac{t-500}{170}\right)\cos[0.06(t-500)] \tag{97}$$

The central frequency of the pulse is $\omega = 0.06$ a.u., which corresponds to a wavelength $\lambda = 760_{nm}$—similar to the central wavelength of the common Titanium-Sapphire laser. The envelope sech^2 form is known to be similar to the actual form of laser pulses. The peak amplitude of the field, $\zeta_{max} = 0.1$ a.u., corresponds to an intensity of $I_{max} = 3.52 \times 10^{14}$ W/cm^2. The final time is $T = 1000$ a.u., which corresponds to a pulse duration of 24.2 fs.

The dipole approximation is employed for the potential induced by the laser field. The total potential of the *physical model* is

$$V_{phys}(x,t) = V_{atom}(x) + V_{field}(x,t) = 1 - \frac{1}{\sqrt{x^2+1}} - x\zeta(t) \tag{98}$$

However, the potential of the *numerical problem* must be modified in order to obtain a reliable description of the physical situation, as will be explained in Sec. 4.2.

The mass of the electron is $m = 1$ a.u., and the kinetic energy becomes $p^2/2$. The total time-dependent *physical Hamiltonian* is

$$H(t) = \frac{p^2}{2} + 1 - \frac{1}{\sqrt{x^2+1}} - x\zeta(t) \tag{99}$$

The dynamics is governed by the time-dependent Schrödinger equation, Eq. (4).

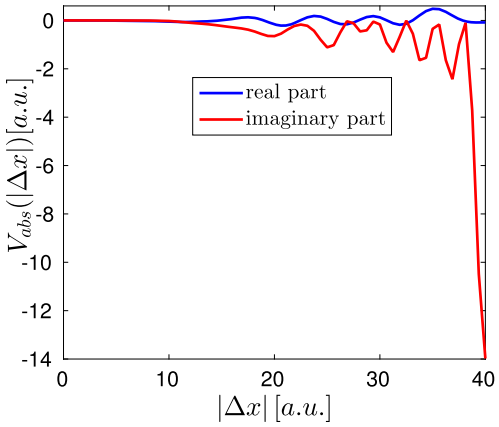

Fig. 4. The real and imaginary parts of the absorbing potential, as a function of the absolute distance $|\Delta x|$ from the beginning of the absorbing boundary at $x = \pm 200$ a.u.

.

4.2. Numerical implementation of the problem

The Fourier grid method [23] is employed for the Hamiltonian operation.

The x domain is $[-240, 240)$. We use an equidistant grid, with 768 points. The distance between adjacent grid points becomes 0.625 a.u..

The present physical situation, which involves a partial ionization of the electron, requires a special numerical treatment, in order to prevent the appearance of spurious effects. The reason is that the ionized part of the electron behaves as a free particle, and is spread to very large spatial distances from the parent atom. Hence, the problem cannot be described as is in a finite spatial grid of a reasonable length. The description of the problem by a finite grid involves spurious effects of wraparound or reflection (depending on the computational method) of the wave function at the boundaries of the grid.

Usually, this problem is overcome by the employment of absorbing boundary conditions. The absorbing boundaries are implemented here by the addition of a complex absorbing potential at both boundaries of the grid (for a thorough review see [31]). The part of the wave-function which incomes into the complex absorbing potential decays gradually under the influence of the potential, until it becomes practically zero at the edge of the grid. Thus, the spurious effects are prevented. With the addition of the complex potential, the Hamiltonian becomes non-Hermitian, and consequently, the eigenvalue spectrum becomes complex.

Different absorbing potentials vary in their absorption capabilities. The part of the amplitude which is not absorbed by the potential is either reflected by the potential or transmitted. Thus, the efficiency of the absorbing boundaries in the prevention of spurious effects depends on the choice of the absorbing potential. The question of the choice of the absorbing potential becomes important when there is an interest in small amplitude effects. One of the major applications of the present physical situation is in the generation of high-harmonic spectrum, which is a small amplitude effect. It has already been recognized in the former high-harmonic generation simulations (see [25]) that reflection from the absorbing boundaries is responsible to large spurious effects in the calculation of the high-harmonic spectrum. An appropriate absorbing potential for this calculation could not be found by inspection.

In our simulation, we use an absorbing potential which is optimized numerically to maximize the absorption. The procedure basically relies on the principles presented in [36], but with several necessary modifications. The real part of the absorbing potential is constructed from a finite cosine series. The imaginary part is constructed from another finite cosine series, where the imaginary potential is given by squaring the cosine series and adding a minus sign. The optimization parameters are the cosine coefficients. This topic will be hopefully presented elsewhere. We choose the length of the absorbing boundary to be 40 a.u.. The obtained potential has a large imaginary part, which induces a significant shift of the Hamiltonian eigenvalues from the real axis into the fourth quarter of complex plane. The real and imaginary parts of the potential are plotted in Fig. 4 as a function of the absolute distance from the beginning of the absorbing boundary at $x = \pm 200$ a.u.. The absorbing potential added at the left boundary is the mirror image of that added at the right boundary. The values of the absorbing potential are available in the complementary material.

It was verified that the results do not change significantly if the grid length is doubled, and the form of the high-harmonic spectrum is very well preserved (a test which failed in Ref. [25] for high intensity field, even for very large grid). The peak error of $|u_i|^2$ from the doubled grid does not exceed the order of 10^{-5}, where u_i is the i'th component of $\vec{\mathbf{u}}$ (of course, in the physical region, $|x| \leq 200$). Thus, the boundary effects are reduced to a reasonable magnitude.

Since the absorbing potential is optimized for an ideal absorption, the influence of the physical potential at the boundaries should be "turned off", in order that the absorption will not be damaged. This is achieved by a modification of the physical potential to a potential which is constant at the absorbing boundaries. In order to avoid discontinuities in the potential derivatives, it is desirable to "turn off" the physical potential in a continuous manner. For that purpose, we use the following practice. Let us define a *soft rectangular function*:

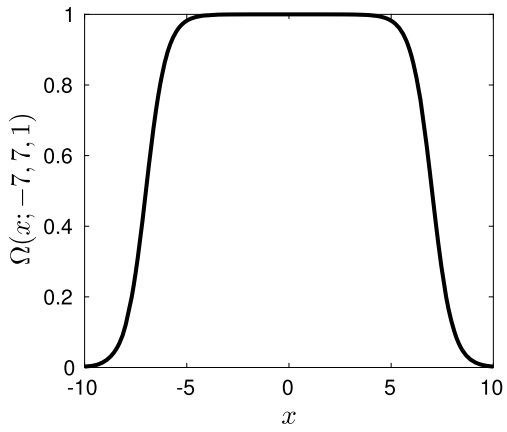

Fig. 5. The soft rectangular function, with parameters $a=-7$, $b=7$, $\alpha=1$.

$$\Omega(x;a,b,\alpha)=\frac{1}{2}\left\{\tanh\left[\alpha(x-a)\right]-\tanh\left[\alpha(x-b)\right]\right\} \tag{100}$$

The function is plotted in Fig. 5 for arbitrary parameters. The modified potential, which is constant at the absorbing boundaries, will be denoted as $V_{mod}(x)$. It is defined to satisfy the following conditions:

$$V_{mod}(0)=V_{phys}(0) \tag{101}$$

$$V'_{mod}(x)=V'_{phys}(x)\Omega(x;a,b,\alpha) \tag{102}$$

$V_{mod}(x)$ is obtained by integration in the following way:

$$\begin{aligned}V_{mod}(x)&=V_{phys}(0)+\int_0^x V'_{phys}(\xi)\Omega(\xi;a,b,\alpha)\,d\xi\\&=V_{phys}(0)+V_{phys}(x)\Omega(x;a,b,\alpha)-V_{phys}(0)\Omega(0;a,b,\alpha)-\int_0^x V_{phys}(\xi)\Omega'(\xi;a,b,\alpha)\,d\xi\\&\approx V_{phys}(x)\Omega(x;a,b,\alpha)-\int_0^x V_{phys}(\xi)\Omega'(\xi;a,b,\alpha)\,d\xi\end{aligned} \tag{103}$$

where we utilized the fact that $V_{phys}(0)\Omega(0;a,b,\alpha)\approx V_{phys}(0)$. We have:

$$\Omega'(x;a,b,\alpha)=\frac{1}{2}\alpha\left\{\operatorname{sech}^2\left[\alpha(x-a)\right]-\operatorname{sech}^2\left[\alpha(x-b)\right]\right\} \tag{104}$$

The integration in Eq. (103) is performed numerically.

In our problem, we choose the following parameters: $a=-197.5$ a.u., $b=197.5$ a.u., $\alpha=1$.

The numerical potential is given by

$$V_{num}(x)\equiv V_{mod}(x)+\begin{cases}0 & |x|<200\\ V_{abs}(|x|-200) & |x|\geq 200\end{cases} \tag{105}$$

In Fig. 6, we plot both $V_{mod}(x)$ and the real part of $V_{num}(x)$ for a relatively small field, $\zeta=0.005$ a.u..

4.3. Results

The problem was solved by the semi-global propagator for different choices of M and K values. For each M and K choice, the problem was solved several times with different values of Δt (the time-step is constant throughout the propagation, as well as M and K). We compute the magnitude of the relative error of the final solution for each parameter choice. The efficiency of the propagator is demonstrated by a comparison of the resulting errors with those obtained by Runge–Kutta of the 4'th order (RK4, see Sec. 2.2). We compare also between the results of the semi-global propagator for the different M and K values.

In order to compare between different methods and parameter choices, we should compare the computational effort required for a similar accuracy. This may be done by choosing several specific examples. However, a fuller and a more

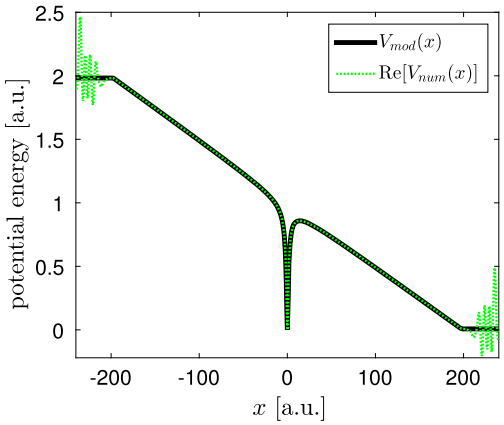

Fig. 6. $V_{mod}(x)$ and the real part of $V_{num}(x)$ for $V_{phys}(x) = V_{atom}(x) - 0.005x$.

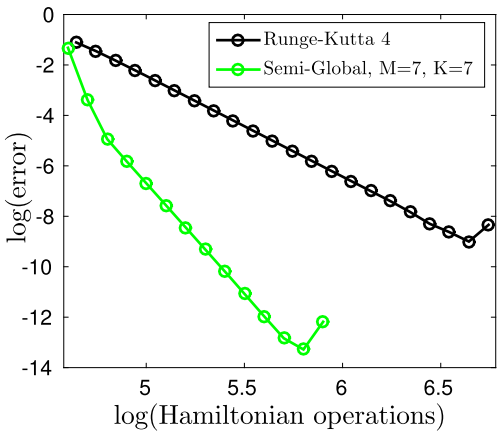

Fig. 7. The error decay curves of the RK4 method, and the semi-global propagator with parameters $M = K = 7$. The $\log_{10}$ of the relative error is plotted vs. the $\log_{10}$ of the number of Hamiltonian operations. The last sampling point in each curve represents the limit in which the effects of roundoff errors become important, and the error ceases to decay. The RK4 curve shows a linear behavior with a slope very close to -4. There is also a seemingly linear region in the semi-global curve, with a slope close to -9.

reliable comparison is obtained by investigating the *behavior* of the error decay with the computational effort. This is done by plotting the *error decay curve* for each method and parameter choice. In the error decay curve, the error is plotted Vs. the computational effort for several choices of Δt. A log–log plot is used. The computational effort is measured here by the number of Hamiltonian operations, which constitute the majority of the computational effort.

In order to obtain a consistent behavior of the error decay for the semi-global propagator, we use a slightly different version of the algorithm from that presented in Sec. 3.2; we restrict the number of iterations to a single iteration, i.e. the steps of the loop in stage 2c of the algorithm are performed just once. This is with the exception of the first time-step, in which the solution is computed without a limitation on the number of iterations, where the parameter ϵ (see Sec. 3.2) represents the machine accuracy of the double-precision. This version of the algorithm restricts the inner freedom in the algorithm, thus ensures the consistency of the error decay curve.

The relative error should be computed from a highly accurate solution of the problem. A highly accurate solution cannot be obtained from RK4 with double-precision, even with an extremely small time-step. This is due to accumulation of the machine errors. Hence, we use a reference solution obtained by the semi-global propagator, with the following parameters: $M = 9$, $K = 13$, $\Delta t = 1/30$. No limitation is imposed on the number of iterations, and ϵ represents the machine accuracy of the double precision. It was verified that the estimated errors, computed by the tests for the different sources of the error in Appendix D, do not exceed the order of the machine accuracy. The high accuracy of the obtained solution is evident from the shapes of the error decay curves.

First, we shall compare the results of the semi-global propagator with the parameters $M = K = 7$, with those of RK4. The error decay curves are plotted in Fig. 7. The sampling points represent gradually decreasing values of Δt, for which the computational effort gradually increases. In each curve, Δt is decreased until the error stops to decay, due to the effects of roundoff errors.

The linear behavior of the RK4 curve is apparent. This is in consistence with theory—the error of the RK4 method is of $O(\Delta t^4)$ (see Sec. 2.2). Since $\Delta t = T/N_t$, the error decays as N_t^{-4}. The number of Hamiltonian operations is linear with N_t. Thus, the log–log plot yields a -4 slope. The slope obtained by a linear fit of the linear region of the curve agrees very well with theory (the obtained slope is -3.99). The semi-global curve has also a seemingly linear region. The slope obtained

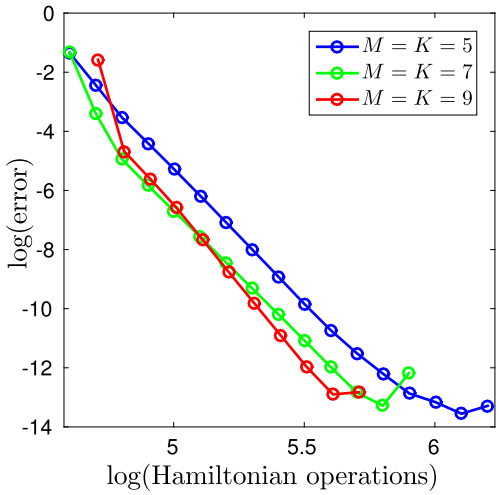

Fig. 8. The error decay curves of the semi-global propagator with different choices of M and K. The $\log_{10}$ of the relative error is plotted vs. the $\log_{10}$ of the number of Hamiltonian operations. All curves include a region with a seemingly linear behavior.

by a linear fit of the linear region is close to -9 (the precise value obtained is -8.77). The advantage of the semi-global propagator can be clearly seen.

Another advantage of the semi-global propagator is the maximal accuracy which can be obtained with the same machine accuracy. The minimal relative error which was obtained by RK4 is 9.96×10^{-10}. Thus, in this problem, the RK4 method with double-precision is limited to an accuracy of about 10^{-9}. The minimal error obtained for the semi-global propagator with this choice of parameters is 5.25×10^{-14}.

Relying on the linear fit of both curves, one can estimate the number of Hamiltonian operations needed for a requested accuracy for each method. Regular accuracy requirements for most physical applications are of the order of 10^{-5}. One finds that for an accuracy of 10^{-5}, the RK4 method requires 6.8 times the number of Hamiltonian operations required for the semi-global propagator. The advantage of the semi-global propagator becomes more apparent for applications which require a high accuracy solution. For an accuracy of 10^{-9}, the RK4 method requires 24 times the number of Hamiltonian operations required for the semi-global propagator.

We proceed with a comparison between different choices of K and M. We shall compare between the following choices: $M = K = 5$, $M = K = 7$, $M = K = 9$ (the results of $M = K = 7$ have already been presented in Fig. 7). The error decay curves are shown in Fig. 8. The choice of $M = K = 5$ is shown to give inferior results in comparison to the higher orders. The $M = K = 7$ choice is slightly advantageous over the $M = K = 9$ for regular accuracy requirements. The $M = K = 9$ choice becomes superior for high accuracy requirements. These findings are by no means general; the ideal parameter choice for a required accuracy is problem dependent.

All curves include a region with a seemingly linear behavior. The slope of the linear region for $M = K = 5$ is very close to -9 (the precise value is -8.98). The slope for $M = K = 9$ is -10.6. As has already been mentioned, the slope of $M = K = 7$ is -8.77.

The explanation of these results requires a detailed error analysis. One can show that the error resulting from each of the three error sources, mentioned in Appendix D, behaves polynomially with Δt, under certain approximation assumptions. This is the origin of the seemingly linear behavior in the error decay curves. However, the order of each error source with Δt is different. Since the overall error depends on several error sources, its behavior with Δt is considerably more complicated than that of RK4, in which there is only one error source. A detailed error analysis is beyond the scope of the present paper, and is left for a future publication.

Nevertheless, the error decay rate in each curve is shown to be much advantageous over the common Taylor methods. The error decay rate is higher for each parameter choice than a Taylor method of the same order (the order of approximation for each of the three parameter choices is $M - 1$, and the slope predicted for a corresponding Taylor method is $-(M - 1)$). Particularly, the $M = K = 5$ choice has the same approximation order as RK4, and the decay rate is much higher.

We can summarize that the semi-global propagator has two advantages over the Taylor approach, which lead to a higher error decay rate:

1. In general, approximation by higher orders leads to higher error decay rate. The use of high order expansion is not recommended in Taylor methods, because of the inefficiency of a Taylor series as an approximation tool in higher orders (see Sec. 2.2). The semi-global approach, being free of Taylor considerations, allows to use higher order expansions than the Taylor approach;
2. The error decay rate of the semi-global propagator is higher even for the same expansion order as the Taylor method.

In this context, it is interesting to compare between the current approach and a class of propagators, known as *exponential integrators* (see, e.g., [8,17–19,30,43]). The propagation technique of the exponential integrators is also based on the $f_m(z,t)$ functions (defined in Eq. (35)). Certain elements of the current approach can be found to have been employed in exponential integrators (see, in particular, [13]). However, there is a fundamental difference between this class of propagators and the current approach: The exponential integrator studies always employ local Taylorian considerations for constructing the propagation, while the propagation technique of the present approach is Taylor free. Hence, the error decay rate of the exponential integrators is limited by the order of the Taylor approximation employed, with the slope predicted for a simple Taylor method of the same order (see, for example, [43]). In contrast, the error decay rates in the present algorithm significantly exceed those of a Taylor method with the same expansion order. This fundamental difference between the approaches also allows to use higher expansion orders and larger time steps in the present approach in comparison to the exponential integrators.

5. Conclusion

The solution of the time-dependent Schrödinger equation is one of the most important tasks in quantum physics. It can be solved by global means when the Hamiltonian is stationary. This approach leads to vast improvement in accuracy and efficiency. In the present paper we presented a generalization of the global approach to the general case of a time-dependent, nonlinear Hamiltonian, with the additional inclusion of an inhomogeneous source term. The global approach can be implemented for the inhomogeneous Schrödinger equation with a stationary Hamiltonian. The solution method in this case constitutes the basis for the present approach for the general case of time-dependence or nonlinearity of the Hamiltonian. The general case can be treated by a semi-global approach, which combines global and local elements. The semi-global approach is characterized by propagation in relatively large time-steps, each of which is treated by global means.

The semi-global approach was shown to be significantly more efficient than the common local approach, which is based on Taylor considerations. Since the propagation method is Taylor free, it has the advantage of being able to use higher order approximations. The error decay rates were shown to be much higher in the new approach, thus enabling to achieve highly accurate solutions with a vast decrease in computational effort.

The semi-global algorithm applies also to the solution of a general set of ODE's, a fundamental problem in numerical analysis. The solution of the Schrödinger equation is an application in a special case of the general problem.

It was demonstrated that the semi-global algorithm is applicable also to non-Hermitian operators by the use of the Arnoldi approach. Thus, it applies also to problems in non-Hermitian quantum mechanics or to the solution of the Liouville–von-Neumann equation.

We asserted that the success of the present approach is mainly attributed to the global element of the method, in which large intervals are treated as a whole in a unified process. This significantly reduces the main problem in the regular local schemes, in which the step-by-step propagation leads to a large numerical effort and error accumulation.

This advantage of the global approach over the local one reveals a more fundamental difference. The local approach relies (explicitly or implicitly) on a Taylor approximation. The Taylor considerations are based on the derivative concept, which is local in nature. The ability to deduce global information from local information is limited. Hence, it is not surprising that the Taylor expansion has poor convergence properties, thus becomes an inefficient tool for approximation purposes. In our opinion, the extreme importance of the Taylor expansion for analysis led, unjustly, to its wide spread as an approximation tool. In contrary, the present approach relies on orthogonal polynomial expansions, which have fast convergence properties. The approximation by an orthogonal set is intimately related to the fundamental concept of *interpolation* (see Appendix A.2). The interpolation concept is global in nature, where the information is deduced from samplings which are distributed all over the approximation interval. Thus, it becomes significantly advantageous over the Taylor expansion as an approximation tool.

It should be noted that even the local element in the semi-global approach is advantageous over the local Taylor considerations. The initial information for the propagation into the next time-step is obtained by *extrapolation*, rather than local derivative information. The extrapolation concept is just an extension of the interpolation concept. A global interpolation approximation inside the interpolation interval can supply relatively accurate information for extrapolation outside the interval.

The main disadvantage of the present version of the algorithm is the necessity of specifying three parameters by the user. A successful choice of the parameters requires a trial and error process and experience. In order to accommodate with this problem, a parameter free version of the algorithm should be developed. The parameters will be determined adaptively by the procedure during the propagation process, to achieve maximal efficiency for the required accuracy. Such a version of the algorithm is expected also to significantly enhance the efficiency. The Chebyshev approximation should be replaced by a Leja approximation [39] for the flexibility of the parameter determination process. The efficiency of the procedure requires an accurate error estimation. One of the advantages of interpolation approximations is the relative ease of the error analysis and estimation. Thus, the adaptive semi-global scheme is expected to be more successful than the available adaptive schemes for Taylor propagators.

We believe that a proliferation of the semi-global algorithm will lead to a significant improvement of accuracy and efficiency in quantum applications, as well as in the vast variety of problems which require the solution of a large set of ODE's.

Acknowledgements

We want to thank Christiane Koch, Lutz Marder and Erik Torrontegui for their interest and help in this project. Calculations were carried out on high performance computers purchased with the help of the Wolfson Foundation. Work supported by Army Research Office (ARO) contract number W911NF-15-10250.

Appendix A. Polynomial approximations

A.1. Approximation by a Newton interpolation

A.1.1. The Newton interpolation

The Newton interpolation is a polynomial interpolation of a function $f(x)$. The function is sampled at specific points distributed in the domain of approximation. The Newton interpolation polynomial approximates $f(x)$ within the domain from the sampling points. The approximation becomes exact at the sampling points.

Let us denote the sampling points by

$$x_j, \qquad j = 0, 1, \ldots, N \tag{106}$$

The value of $f(x_j)$ is given for all sampling points. The Newton interpolation approximates $f(x)$ using the following form:

$$\begin{aligned} f(x) &\approx a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1) + \ldots + a_N(x - x_0)(x - x_1)\cdots(x - x_{N-1}) \\ &= \sum_{n=0}^{N} a_n R_n(x) \end{aligned} \tag{107}$$

where the a_n's are coefficients, and the $R_n(x)$'s are defined by

$$\begin{aligned} R_0(x) &= 1 \\ R_n(x) &= \prod_{j=0}^{n-1}(x - x_j) \qquad n > 0 \end{aligned} \tag{108}$$

The $R_n(x)$'s are called "Newton basis polynomials". In order to find the a_n's, we first have to become familiar with the concept of *divided difference*, which will be presented below.

Let us consider a function $f(x)$, and a set of points, as in Eq. (106). The divided differences have a recursive definition. We will give the definition, and examples will follow immediately. The divided difference for a single point x_0 is defined as

$$f[x_0] \equiv f(x_0) \tag{109}$$

The divided difference of more than a single point is defined as

$$f[x_0, x_1, \ldots, x_n] = \frac{f[x_1, x_2, \ldots, x_n] - f[x_0, x_1, \ldots, x_{n-1}]}{x_n - x_0} \tag{110}$$

For instance, the divided difference of two points is

$$f[x_0, x_1] \equiv \frac{f[x_1] - f[x_0]}{x_1 - x_0} = \frac{f(x_1) - f(x_0)}{x_1 - x_0} \tag{111}$$

The divided difference of three points is

$$f[x_0, x_1, x_2] \equiv \frac{f[x_1, x_2] - f[x_0, x_1]}{x_2 - x_0} = \frac{\frac{f(x_2) - f(x_1)}{x_2 - x_1} - \frac{f(x_1) - f(x_0)}{x_1 - x_0}}{x_2 - x_0} \tag{112}$$

The a_n coefficients are given by the divided differences as follows:

$$a_n = f[x_0, x_1, \ldots, x_n] \tag{113}$$

A.1.2. Interpolation at Chebyshev points

When using a high order polynomial interpolation at equally spaced points, we may encounter a phenomenon which prevents it from being useful. The interpolation polynomial does not converge to the function $f(x)$ at the edges of the domain of approximation (even though it may converge very well at the middle of the domain). Instead, we might observe very large oscillations at the edges. The problem becomes more severe as N grows. This phenomenon is known as *Runge phenomenon*.

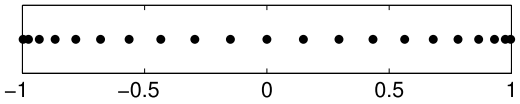

Fig. 9. The Chebyshev points (Eq. (114)) for $N = 20$.

The Runge phenomenon disappears when we choose the sampling points of $f(x)$ appropriately. There is more than one appropriate choice of a set of sampling points. The most commonly used set is given by the so-called *Chebyshev points* of the domain of approximation. The Chebyshev points become denser at the edges of the domain (see Fig. 9; this property is common to all sets of points which solve the Runge phenomenon). The Chebyshev points are originally defined for the domain $[-1, 1]$, but they can be easily transformed to another domain by a simple linear transformation, as will be described later.

The Chebyshev points for the domain $[-1, 1]$ are:

$$y_j \equiv \cos\left[\frac{(2j+1)\pi}{2(N+1)}\right], \qquad j = 0, 1, \ldots, N \tag{114}$$

Note that $y_0 > y_1 > \ldots > y_N$. Frequently, it is more convenient to index the points in an increasing order. We can reverse the order of the points, by defining them in the following way:

$$y_j \equiv -\cos\left[\frac{(2j+1)\pi}{2(N+1)}\right], \qquad j = 0, 1, \ldots, N \tag{115}$$

There is an alternative set of Chebyshev points, with similar characteristics:

$$y_j \equiv \cos\left(\frac{j\pi}{N}\right), \qquad j = 0, 1, \ldots, N \tag{116}$$

or, with a reversed order:

$$y_j \equiv -\cos\left(\frac{j\pi}{N}\right), \qquad j = 0, 1, \ldots, N \tag{117}$$

In this set of points, the function is sampled at the boundaries of the domain, unlike in the set (114). This can be advantageous in certain circumstances (for example, in the context of the semi-global propagation scheme, in which adjacent time-steps share a common point; see Sec. 3.1).

The Chebyshev points can be transformed to an arbitrary domain on the real axis, $[x_{min}, x_{max}]$. First, they are stretched or compressed to match the size $\Delta x = x_{max} - x_{min}$ of the domain:

$$y_j \longrightarrow \frac{\Delta x}{2} y_j$$

Then, they are shifted to the middle of the domain by adding the middle point,

$$\frac{x_{min} + x_{max}}{2}$$

The sampling points are finally obtained by the following linear transformation:

$$x_j \equiv \frac{1}{2}(y_j\,\Delta x + x_{min} + x_{max}) \tag{118}$$

There are other sets of points which solve the Runge phenomenon. The set of *Leja points* [39] can be advantageous when the required degree of approximation N is difficult to be estimated in advance. This topic is beyond the scope of this paper.

A.1.3. Numerical stability of the Newton interpolation

The Newton interpolation usually becomes numerically unstable when N grows. The main problem is that the a_n's of high n tend to become very large, and the corresponding adjacent polynomials become very small, or vice versa. This problem does not exist when the domain of approximation, $[x_{min}, x_{max}]$, is of length 4 [46]. Hence, for a domain defined on the real axis, the problem can be overcome by transforming the problem to a length 4 domain.

First, we transform the sampling points to a length 4 domain. The points in the new domain are defined as

$$\bar{x}_j \equiv \frac{4}{\Delta x} x_j \tag{119}$$

In general, we can define a transformed *variable*:

$$\bar{x} = \frac{4}{\Delta x} x \tag{120}$$

We also define a new function $\bar{f}(x)$, such that:

$$\bar{f}(\bar{x}) = f(x) \tag{121}$$

Then, we can approximate $f(x)$ at an arbitrary x value in the original domain, by using the Newton interpolation for the function $\bar{f}(x)$ at the sampling points $\bar{x}_j$. The approximation to $f(x)$ at an arbitrary x is given by the interpolation polynomial value at $\bar{x}$.

A.2. Chebyshev approximation

In the Chebyshev approximation, a function $f(x)$ is approximated by a truncated series of orthogonal polynomials, in the form of Eq. (15). The basis of expansion consists of the *Chebyshev polynomials*. They are defined as follows:

$$T_n(x) = \cos(n \cos^{-1} x), \qquad x \in [-1, 1], \qquad n = 0, 1, \ldots \tag{122}$$

In order to clarify the meaning of this weird definition, let us define a variable θ such that

$$x = \cos\theta \tag{123}$$

Then we obtain the following equivalent definition of the Chebyshev polynomials:

$$T_n(\cos\theta) = \cos(n\theta), \qquad \theta \in [0, \pi], \qquad n = 0, 1, \ldots \tag{124}$$

For instance,

$$T_0(x) = \cos(0) = 1 \tag{125}$$

$$T_1(x) = \cos\theta = x \tag{126}$$

$$T_2(x) = \cos(2\theta) = 2\cos^2\theta - 1 = 2x^2 - 1 \tag{127}$$

Note that the functions $\cos(n\theta)$ from the RHS of Eq. (124) span the function space in the domain $\theta \in [0, \pi]$. In addition, they are *orthogonal* in this domain:

$$\int_0^\pi \cos(m\theta)\cos(n\theta)\, d\theta = \frac{\pi \alpha_n}{2} \delta_{mn}, \qquad \alpha_n \equiv \begin{cases} 2 & n = 0 \\ 1 & n > 0 \end{cases} \tag{128}$$

The Chebyshev polynomials are obtained from the cosine basis by mapping θ into x with the nonlinear transformation (123). The resulting functions are also orthogonal in the x space, but with respect to a *weight function* in the new space. This can be seen by changing the integration variable of Eq. (128) from θ to x. We obtain:

$$\int_0^\pi \cos(m\theta)\cos(n\theta)\, d\theta = -\int_1^{-1} \frac{T_m(x) T_n(x)}{\sin\theta}\, dx = \int_{-1}^{1} \frac{T_m(x) T_n(x)}{\sqrt{1 - x^2}}\, dx = \frac{\pi \alpha_n}{2} \delta_{mn} \tag{129}$$

We have obtained the orthogonality relation of the Chebyshev polynomials under the weight function

$$w(x) = \frac{1}{\sqrt{1 - x^2}} \tag{130}$$

The Chebyshev polynomials satisfy the following recurrence relation:

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x), \qquad n \geq 1 \tag{131}$$

All Chebyshev polynomials can be obtained recursively from $T_0(x)$ and $T_1(x)$ (see Eqs. (125), (126)), using Eq. (131).

Now suppose we want to approximate a function $f(x)$ in a given domain $[x_{min}, x_{max}]$, from several samplings of the function in the domain. Suppose we can choose the sampling points as we wish. The function can be approximated from the sampling points by a Chebyshev series, as will be described below. For simplicity, let us first assume that the domain of approximation is $[-1, 1]$.

$f(x)$ can be spanned in the following form:

$$f(x) \approx \sum_{n=0}^{N} c_n T_n(x) \tag{132}$$

The c_n's are called the *Chebyshev coefficients* of $f(x)$. They can be obtained by projecting $f(x)$ onto each of the $T_n(x)$ basis functions. Using the orthogonality relation (129), the Chebyshev coefficients are given by the following scalar product expression:

$$c_n = \frac{2}{\pi \alpha_n} \int_{-1}^{1} f(x) T_n(x) w(x)\, dx \tag{133}$$

Equivalently, we can perform the scalar product in the θ space:

$$c_n = \frac{2}{\pi \alpha_n} \int_{0}^{\pi} f(\cos\theta) \cos(n\theta)\, d\theta \tag{134}$$

Suppose that the form of $f(x)$ is unknown, or that the integral cannot be performed analytically. We have to compute the Chebyshev coefficients *numerically* from a finite number of samplings of $f(x)$ within the domain. Fortunately, the problem of computing the Chebyshev coefficients can be reformulated by discrete means, without reduction of the quality of the approximation.

We utilize the fact that there exist discrete versions of the orthogonality relations between the cosine basis functions. The orthogonality relations are given by a finite sum over samplings of the cosine functions in equally spaced points in θ. For instance, we have the following orthogonality relation:

$$\sum_{j=0}^{K} \cos(k\theta_j) \cos(l\theta_j) = \frac{(K+1)\alpha_k}{2} \delta_{kl}, \qquad \theta_j \equiv \frac{(2j+1)\pi}{2(K+1)} \tag{135}$$

where α_k is defined in Eq. (128), and $K \geq 0$. Note that in the x space, the points $x_j = \cos\theta_j$ are just the Chebyshev points defined in (114). The orthogonality relation (135) can be utilized in order to find the Chebyshev coefficients from the sampling of $f(x)$ at the Chebyshev points, as we shall see. Let us define:

$$g(\theta) \equiv f(\cos\theta) \tag{136}$$

Eq. (132) can be rewritten as

$$g(\theta) \approx \sum_{m=0}^{N} c_m \cos(m\theta) \tag{137}$$

Let us multiply $g(\theta)$ by the basis function $\cos(n\theta)$, and sum over the set of $N+1$ Chebyshev points:

$$\sum_{j=0}^{N} g(\theta_j) \cos(n\theta_j), \qquad \theta_j = \frac{(2j+1)\pi}{2(N+1)}$$

We substitute $g(\theta)$ with the approximation (137), and obtain:

$$\sum_{j=0}^{N} g(\theta_j) \cos(n\theta_j) \approx \sum_{j=0}^{N} \sum_{m=0}^{N} c_m \cos(m\theta_j) \cos(n\theta_j) = \frac{(N+1)\alpha_n}{2} c_n \tag{138}$$

where we applied the orthogonality relation (135). The Chebyshev coefficients are finally given by

$$c_n = \frac{2}{(N+1)\alpha_n} \sum_{j=0}^{N} g(\theta_j) \cos(n\theta_j) = \frac{2}{(N+1)\alpha_n} \sum_{j=0}^{N} f\left[\cos\left(\frac{(2j+1)\pi}{2(N+1)}\right)\right] \cos\left[\frac{n(2j+1)\pi}{2(N+1)}\right] \tag{139}$$

Note that the Chebyshev coefficients defined by Eq. (139) are not identical with those defined by the integral version of Eq. (134). However, the inaccuracy in (139) is originated in the truncation error of Eq. (132) itself (see Eq. (138)). Thus, the total error will be of the same order of magnitude as the truncation error, and the quality of the approximation will be similar.

Actually, the set of $N+1$ equations defined by (139) is a *linear transformation* of the function value vector, $[g(\theta_0), g(\theta_1), \ldots, g(\theta_N)]^T$, into the coefficient vector, $[c_0, c_1, \ldots, c_N]^T$. This transformation is called a *discrete cosine transform* (DCT). It has an apparent similarity to the discrete Fourier transform (DFT). There are several kinds of DCT's. The transformation defined in Eq. (139) is sometimes referred as a *DCT of the second kind*.

The DCT's are reversible transformations. The inverse transformation of Eq. (139) is actually defined by Eq. (137) for the set of θ_j's. Hence, the approximation is *exact* for the Chebyshev sampling points. In that sense, Eq. (132) with the c_n's

computed by Eq. (139) defines an *interpolation* of $f(x)$ in the set of $N+1$ Chebyshev points. A fundamental theorem of interpolation theory states that the interpolation polynomial for a given set of sampling points is *unique*. Thus, the approximation presented here is equivalent to the approximation by a Newton interpolation at the Chebyshev points, described in Sec. A.1.2.

The DCT transformations can be computed very efficiently by algorithms derived from the fast Fourier transform (FFT) algorithm. Hence, the scaling of the computational effort is of $O(N \ln N)$. There are available programs for computation of the DCT. When using them, a care should be taken on the exact definition of the transformation—usually, there are slight differences in the definition of the transformation coefficients.

It is also possible to approximate the Chebyshev coefficients by sampling $f(x)$ at the set of boundary including Chebyshev points (see Eq. (116)). We use the following discrete orthogonality relation:

$$\sum_{j=0}^{K} \frac{1}{\beta_j} \cos(k\theta_j)\cos(l\theta_j) = \frac{K\beta_k}{2}\delta_{kl}, \qquad \theta_j \equiv \frac{j\pi}{K}, \qquad \beta_j \equiv \begin{cases} 2 & j = 0, K \\ 1 & 1 \le j \le K-1 \end{cases} \tag{140}$$

where $K \ge 1$. Following similar steps as above, we obtain:

$$c_n = \frac{2}{N\beta_n} \sum_{j=0}^{N} \frac{1}{\beta_j} g(\theta_j)\cos(n\theta_j) = \frac{2}{N\beta_n} \sum_{j=0}^{N} \frac{1}{\beta_j} f\left[\cos\left(\frac{j\pi}{N}\right)\right] \cos\left(\frac{nj\pi}{N}\right) \tag{141}$$

Here again, the set of $N+1$ equations defined by Eq. (141) is a linear transformation of the function value vector into the coefficient vector. It is another kind of a DCT, sometimes referred as a *DCT of the first kind*. Its inverse is defined by Eq. (137) for this set of points. The boundary including points are preferable when the exact value of $f(x)$ at the boundaries is important.

In the case that the domain of approximation of $f(x)$ is different from the Chebyshev domain, $[-1, 1]$, it is required to shift the problem to the Chebyshev domain. The treatment of the problem is similar to that of Sec. A.1.2. For instance, let us discuss the approximation by the boundary including Chebyshev points. We denote the domain of approximation by $x \in [x_{min}, x_{max}]$. Let us denote the set of Chebyshev points by y_j:

$$y_j \equiv \cos\left(\frac{j\pi}{N}\right), \qquad j = 0, 1, \ldots, N \tag{142}$$

The Chebyshev points in the domain $[x_{min}, x_{max}]$ are defined by the linear transformation (118). We can also define a *variable* $y \in [-1, 1]$, which is given by the inverse linear transformation of x:

$$y \equiv \frac{2x - x_{min} - x_{max}}{\Delta x} \tag{143}$$

We define a function $\bar{f}(x)$ such that

$$\bar{f}(y) = f(x) \tag{144}$$

The approximation to $f(x)$ is given by

$$f(x) = \bar{f}(y) \approx \sum_{n=0}^{N} c_n T_n(y) \tag{145}$$

where

$$c_n = \frac{2}{N\beta_n} \sum_{j=0}^{N} \frac{1}{\beta_j} \bar{f}(y_j)\cos(n\theta_j) = \frac{2}{N\beta_n} \sum_{j=0}^{N} \frac{1}{\beta_j} f(x_j)\cos\left(\frac{nj\pi}{N}\right) \tag{146}$$

We see that the Chebyshev coefficients are simply given by a discrete cosine transform of $f(x)$, sampled at the Chebyshev points of the domain $[x_{min}, x_{max}]$.

The treatment in the case of the points of (114) is completely identical. We obtain:

$$c_n = \frac{2}{(N+1)\alpha_n} \sum_{j=0}^{N} f(x_j)\cos\left[\frac{n(2j+1)\pi}{2(N+1)}\right] \tag{147}$$

where the x_j's are given by Eq. (118), with the y_j's of Eq. (114).

As was mentioned in Sec. A.1.2, it is often more convenient to reverse the order of the Chebyshev points, in order to obtain increasing values of x with the point index. This can be done by defining them as in Eqs. (115), (117). However, care

should be taken to preserve the original form of Eqs. (146), (147), in which each of the $f(x_j)$'s is multiplied by the cosine of the corresponding angle. Hence, the order of the angles should also be reversed. This is equivalent to the addition of a minus sign to the RHS of the two equations. Eq. (146) is replaced by

$$c_n = -\frac{2}{N\beta_n}\sum_{j=0}^{N}\frac{1}{\beta_j}f(x_j)\cos\left(\frac{nj\pi}{N}\right) \tag{148}$$

and Eq. (147) is replaced by

$$c_n = -\frac{2}{(N+1)\alpha_n}\sum_{j=0}^{N}f(x_j)\cos\left[\frac{n(2j+1)\pi}{2(N+1)}\right] \tag{149}$$

Appendix B. Approximation methods for the multiplication of a vector by a function of matrix

Here we discuss several methods for the computation of the following vector:

$$\vec{\mathbf{u}} = f(A)\vec{\mathbf{v}} \tag{150}$$

where $\vec{\mathbf{v}}$ is an arbitrary vector, A is a matrix, and $f(x)$ is a function. All approximation methods are based on the following realization: When the multiplication of $f(A)$ with $\vec{\mathbf{v}}$ is all what required, we can avoid the direct computation of $f(A)$, which is highly demanding numerically. Instead, we use successive multiplications of vectors by the matrix A. As has already been mentioned in Sec. 2.3.1, in certain cases we can replace the direct multiplication of the vector by the matrix A, by a computational procedure which is less demanding numerically.

B.1. Polynomial series approximations

The first two methods presented here are based on approximation of $f(x)$ by a polynomial $Q_L(x)$ of degree L. $Q_L(x)$ is a truncated polynomial series of $f(x)$ (cf. Eq. (15)):

$$f(x) \approx Q_L(x) \equiv \sum_{n=0}^{L} b_n P_n(x) \tag{151}$$

where the $P_n(x)$'s are polynomials of degree n, and the b_n's are the corresponding expansion coefficients. We can approximate $\vec{\mathbf{u}}$ by the following expression (cf. Eq. (16)):

$$\vec{\mathbf{u}} \approx Q_L(A)\vec{\mathbf{v}} = \sum_{n=0}^{L} b_n P_n(A)\vec{\mathbf{v}} \tag{152}$$

The idea is to compute the expressions $P_n(A)\vec{\mathbf{v}}$ by successive multiplications of vectors by A, instead of computing $P_n(A)$ and multiplying $\vec{\mathbf{v}}$ by the resulting matrix.

As we have seen in Appendix A, when using a polynomial expansion for the approximation of a function, we first have to define the approximation domain, $[x_{min}, x_{max}]$. The approximation is expected to be accurate only inside the approximation domain. In our problem, the approximation should be accurate in the *eigenvalue domain* of A. This can be readily seen by considering the decomposition of $\vec{\mathbf{v}}$ into the eigenvectors of A:

$$\vec{\mathbf{v}} = \sum_{j=0}^{N-1} v_j \vec{\boldsymbol{\varphi}}_j \tag{153}$$

where the $\vec{\boldsymbol{\varphi}}_j$'s are the eigenvectors of A, the v_j's are the components of $\vec{\mathbf{v}}$ in the eigenvector basis, and N is the dimension of A. Plugging (153) into (150), we obtain:

$$\vec{\mathbf{u}} = \sum_{j=0}^{N-1} v_j f(\lambda_j)\vec{\boldsymbol{\varphi}}_j \tag{154}$$

where λ_j is the eigenvalue of $\vec{\boldsymbol{\varphi}}_j$. When using the approximation of Eq. (151), we actually replace the accurate expression of Eq. (154) by the following approximated expression:

$$\vec{\mathbf{u}} \approx \sum_{j=0}^{N-1} v_j Q_L(\lambda_j)\vec{\boldsymbol{\varphi}}_j \tag{155}$$

It is clear that $Q_L(\lambda_j)$ should be accurate for each of the λ_j's. Hence, the approximation domain has to cover the whole eigenvalue domain of A.

Frequently, the eigenvalue domain of A is unknown, and we have to estimate it. Note that an overestimation of the eigenvalue domain size costs additional numerical effort, because more terms are required to approximate $f(x)$. In cases that the eigenvalue domain cannot be estimated, or when the eigenvalue domain is complex, the Arnoldi approach (see Sec. B.2) should be used instead of the polynomial expansion methods.

B.1.1. Newton interpolation

One approach for approximation of $\vec{\mathbf{u}}$ by a polynomial series is by using a Newton interpolation of $f(x)$ at the Chebyshev points of the eigenvalue domain, defined by Eq. (118) (see Appendix A.1.2) [46]. The Newton interpolation polynomial is (cf. Eq. (107)):

$$Q_L(x) = \sum_{n=0}^{L} a_n R_n(x) \tag{156}$$

The x_j sampling points which define the a_n's and the $R_n(x)$'s are given by Eq. (118). The $R_n(x)$'s satisfy the following recurrence relation:

$$R_{n+1}(x) = (x - x_n) R_n(x) \tag{157}$$

$\vec{\mathbf{u}}$ is approximated by

$$\vec{\mathbf{u}} \approx \sum_{n=0}^{L} a_n R_n(A)\vec{\mathbf{v}} \tag{158}$$

The recurrence relation (157) can be utilized in order to compute the expressions $R_n(A)\vec{\mathbf{v}}$ successively.

Let us index the sampling points by $j = 0, 1, \ldots, L$. The algorithm for the computation of $\vec{\mathbf{u}}$ is described below:

1. Compute the Chebyshev points y_j by Eq. (115) or by Eq. (117).
2. Compute the Chebyshev points x_j in the eigenvalue domain $[x_{min}, x_{max}]$ from the y_j's by Eq. (118).
3. Compute the function values $f_j = f(x_j)$.
4. Compute the divided differences a_n, $n = 0, 1, \ldots, L$, recursively from f_j and x_j, using Eqs. (113), (109), (110).
5. $\vec{\mathbf{w}} = \vec{\mathbf{v}}$
6. $\vec{\mathbf{u}} = a_0\vec{\mathbf{w}}$
7. for $i = 1$ to L
   (a) $\vec{\mathbf{w}} = A\vec{\mathbf{w}} - x_{i-1}\vec{\mathbf{w}}$
   (b) $\vec{\mathbf{u}} = \vec{\mathbf{u}} + a_i\vec{\mathbf{w}}$
8. end for

In practice, the Newton interpolation problem should be transferred to a domain of length 4, for numerical stability (see Appendix A.1.3). This amounts of two slight changes. We have to add to the recursion relation of Eq. (157) an additional conversion factor, in order to transform x to the domain of length 4:

$$R_{n+1}(x) = \frac{4}{\Delta x}(x - x_n) R_n(x) \tag{159}$$

where $\Delta x = x_{max} - x_{min}$. In addition, the sampling points x_j that appear in the denominator of the divided difference formula (110) are replaced by $\bar{x}_j = 4x_j/\Delta x$. Accordingly, we insert the following changes into the algorithm:

1. The x_j's in stage 4 are replaced by $4x_j/\Delta x$ (note that the f_j's from stage 3 remain the same).
2. Stage 7a is replaced by $\vec{\mathbf{w}} = (A\vec{\mathbf{w}} - x_{i-1}\vec{\mathbf{w}})4/\Delta x$

$\vec{\mathbf{u}}$ can be computed with other sets of sampling points which solve the Runge phenomenon (for instance, the Leja points; see Appendix A.1.2). We use the same recursion formula and algorithm, where the x_j's are the desired set of points.

B.1.2. Chebyshev expansion

Another approach for approximating $\vec{\mathbf{u}}$ is by the expansion of $f(A)$ in a Chebyshev series [45]. The approximation polynomial $Q_L(x)$ is given by

$$Q_L(x) = \sum_{n=0}^{L} c_n T_n(y) \tag{160}$$

where

$$y \equiv \frac{2x - x_{min} - x_{max}}{\Delta x} \tag{161}$$

The c_n's are given by Eq. (147) or Eq. (146), where the x_j's are given by Eq. (118), together with Eq. (114) or Eq. (116), respectively (see Appendix A.2).

$\vec{\mathbf{u}}$ is approximated by

$$\vec{\mathbf{u}} \approx Q_L(A)\vec{\mathbf{v}} = \sum_{n=0}^{L} c_n T_n(\bar{A})\vec{\mathbf{v}} \tag{162}$$

where

$$\bar{A} = \frac{2A - x_{min} - x_{max}}{\Delta x} \tag{163}$$

The Chebyshev polynomials satisfy the recurrence relation

$$T_{n+1}(y) = 2yT_n(y) - T_{n-1}(y), \qquad n \geq 1 \tag{164}$$

where

$$T_0(y) = 1 \tag{165}$$

$$T_1(y) = y \tag{166}$$

We can compute the expressions $T_n(\bar{A})\vec{\mathbf{v}}$ successively by utilizing the recurrence relation, where y in Eqs. (164), (166), is substituted by $\bar{A}$.

The algorithm for the computation of $\vec{\mathbf{u}}$ is described below:

1. Compute the Chebyshev points y_j by Eq. (114) or by Eq. (116).
2. Compute the Chebyshev points x_j in the eigenvalue domain $[x_{min}, x_{max}]$ from the y_j's by Eq. (118).
3. Compute the function values $f_j = f(x_j)$.
4. Compute the Chebyshev coefficients c_n from the f_j's by Eq. (147) or by Eq. (146).
5. $\vec{\mathbf{w}}_1 = \vec{\mathbf{v}}$
6. $\vec{\mathbf{w}}_2 = [2A\vec{\mathbf{v}} - (x_{min} + x_{max})\vec{\mathbf{v}}]/\Delta x$
7. $\vec{\mathbf{u}} = c_0\vec{\mathbf{w}}_1 + c_1\vec{\mathbf{w}}_2$
8. for $i = 2$ to L
   (a) $\vec{\mathbf{w}}_3 = 2[2A\vec{\mathbf{w}}_2 - (x_{min} + x_{max})\vec{\mathbf{w}}_2]/\Delta x - \vec{\mathbf{w}}_1$
   (b) $\vec{\mathbf{u}} = \vec{\mathbf{u}} + c_i\vec{\mathbf{w}}_3$
   (c) $\vec{\mathbf{w}}_1 = \vec{\mathbf{w}}_2$
   (d) $\vec{\mathbf{w}}_2 = \vec{\mathbf{w}}_3$
9. end for

B.2. Arnoldi approach

The approximations of Sec. B.1 are based on the assumption that the eigenvalue domain is known, or can be estimated. They cannot be applied when it is impossible to estimate the eigenvalue domain. The difficulty in the estimation of the eigenvalue domain becomes severe when the eigenvalues of A are distributed on the complex plane, and not only on the real or on the imaginary axis.

Moreover, the concept of Chebyshev sampling is defined for a one dimensional *axis*, which may be the real or imaginary axis. Hence, a Chebyshev approximation can be applied for functions of a real variable, or a purely imaginary variable. However, a Chebyshev approximation is not suitable for functions of *complex* variables, which are distributed on the two dimensional *complex plane*. Thus, the methods of Sec. B.1 are not applicable when the eigenvalue spectrum of A is complex.

In the present section, we introduce the Arnoldi approach for approximation of (150) [47]. In the Arnoldi approach, the eigenvalue domain needn't be known, and the sampling is chosen by different considerations. Thus, it becomes suitable also for the treatment of matrices with a complex spectrum.

The Arnoldi approximation is intimately related to the polynomial approximations, but the approach to the problem is different. We can view our basic problem as the problem of reduction of large-scale matrix calculations into simplified approximations. The polynomial approximations reduce the *calculation* of the matrix into a *simplified calculation* of the *same matrix*, in which a function is reduced into a low degree polynomial. In the Arnoldi approach, the *matrix itself* is reduced into a small-scale matrix. This is done by the choice of a "good" set of a small number of vectors, which can be representative in the framework of the specific problem. The matrix A is represented in the reduced subspace spanned by the vectors.

The approximation is based on a reduction of the problem to the subspace spanned by the following vectors:

$$\vec{\mathbf{v}}, A\vec{\mathbf{v}}, A^2\vec{\mathbf{v}}, \ldots, A^L\vec{\mathbf{v}} \tag{167}$$

The subspace is spanned by multiplications of the vector $\vec{\mathbf{v}}$ by powers of A, from degree 0 to L. A vector space of this kind is called a *Krylov space*. The space spanned by (167) is a *Krylov space of dimension* $L+1$. The Krylov space is a "good" subspace in our problem for two reasons, which are interrelated:

1. Any polynomial approximation of $\vec{\mathbf{u}}$ of degree L in the form of (152) can be spanned by the Krylov space. This is a direct consequence of the fact that any polynomial can be expressed in the terms of the Taylor polynomials. Thus, the Krylov space is actually the characteristic subspace of polynomial approximations of degree L.
2. Suppose $\vec{\mathbf{v}}$ can be spanned by a subspace in the eigenvector spectrum of A. The subspace is invariant to multiplication by A or $f(A)$, which are diagonal in the eigenvector basis (this can be readily seen by expanding $\vec{\mathbf{v}}$ in the eigenvector space, as in Eq. (154)). Thus, $\vec{\mathbf{u}}$ remains in the same eigenvector subspace as $\vec{\mathbf{v}}$. The Krylov space also remains in the eigenvector subspace, for the same reason. Hence, it is expected to be effective for the representation of $\vec{\mathbf{u}}$. Note that if $\vec{\mathbf{v}}$ is spanned by an eigenvector space of dimension up to $L+1$, $\vec{\mathbf{u}}$ can be fully represented by the Krylov space. Even when $\vec{\mathbf{v}}$ is spanned by the whole eigenvector space, frequently a small number of eigenvectors dominate its composition. Thus, the Krylov space may still be effective in approximating it.

The vectors of Eq. (167) are in general non-orthogonal. Moreover, the vectors are getting closer to be parallel with the degree of A. When using them as a basis set, this might be a source of numerical instability. We should work with an orthonormal basis set for spanning the Krylov space, for the sake of numerical stability and the simplicity of the calculations.

In order to obtain an orthonormal basis, we use the *Gram–Schmidt process*. The idea underlying the Gram–Schmidt orthonormalization goes as follows: The orthonormal basis vectors are computed successively from the original non-orthogonal set of vectors. We subtract from each vector from the original set its projection on the subspace spanned by the already computed orthonormal vectors. We are left with a vector which is orthogonal to the subspace spanned by the previous vectors. Then, we simply normalize it, and obtain an orthonormal set which is enlarged by one dimension. Then, we continue to the next vector from the original set, and so on.

In practice we use the *Modified-Gram–Schmidt* (MGS) algorithm which is equivalent, mathematically, to Gram–Schmidt but less sensitive to roundoff errors. The orthonormalization in the context of the Krylov space can be implemented by an iterative process, as will be seen. The iterative algorithm is referred as *Arnoldi iteration*.

Let us denote the orthonormal set by

$$\vec{\boldsymbol{\upsilon}}_0, \vec{\boldsymbol{\upsilon}}_1, \ldots, \vec{\boldsymbol{\upsilon}}_L$$

As will be seen, the vector $\vec{\boldsymbol{\upsilon}}_n$ belongs to the Krylov space of dimension $n+1$. In the algorithm, we compute an additional orthonormal vector, $\vec{\boldsymbol{\upsilon}}_{L+1}$, which belongs to the Krylov space of dimension $L+2$. The necessity of its computation will be clarified in what follows. We denote the scalar product of two vectors, $\vec{\mathbf{r}}$ and $\vec{\mathbf{s}}$, by $\langle \vec{\mathbf{r}}, \vec{\mathbf{s}} \rangle$.

The algorithm is described below. During the algorithm, we also store in the memory a set of constants, which participate in the computation. The necessity of this will be clarified in what follows.

1. Compute the first vector in the orthonormal set as the normalized $\vec{\mathbf{v}}$:

$$\vec{\boldsymbol{\upsilon}}_0 = \frac{\vec{\mathbf{v}}}{\|\vec{\mathbf{v}}\|}$$

2. for $j = 0$ to L
   (a) Compute a non-orthonormalized new vector by setting: $\vec{\boldsymbol{\upsilon}}_{j+1} = A\vec{\boldsymbol{\upsilon}}_j$.
   (b) for $i = 0$ to j
      i. Set: $\Gamma_{i,j} = \langle \vec{\boldsymbol{\upsilon}}_i, \vec{\boldsymbol{\upsilon}}_{j+1} \rangle$.
      ii. Subtract from $\vec{\boldsymbol{\upsilon}}_{j+1}$ its projection on $\vec{\boldsymbol{\upsilon}}_i$: $\vec{\boldsymbol{\upsilon}}_{j+1} = \vec{\boldsymbol{\upsilon}}_{j+1} - \Gamma_{i,j}\vec{\boldsymbol{\upsilon}}_i$.
   (c) end for
   (d) Set: $\Gamma_{j+1,j} = \|\vec{\boldsymbol{\upsilon}}_{j+1}\|$.
   (e) Normalize $\vec{\boldsymbol{\upsilon}}_{j+1}$ by setting:

$$\vec{\boldsymbol{\upsilon}}_{j+1} = \frac{\vec{\boldsymbol{\upsilon}}_{j+1}}{\Gamma_{j+1,j}}$$

3. end for

Clearly, the $\vec{\boldsymbol{\upsilon}}_{j+1}$ computed in stage 2a belongs to the Krylov space of dimension $j+2$, since it is composed from some linear combination of the first $j+2$ Krylov vectors. Thus, in each iteration, the Krylov space is enlarged by one dimension.

At the end of the algorithm, we are left with an orthonormal set of dimension $L+2$. We also computed, seemingly by the way, the $\Gamma_{i,j}$'s. These have a special significance. Actually, $\Gamma_{i,j}$ is the *matrix element of A in the orthonormal basis*, i.e., it is equivalent to $A_{i,j} = \langle \vec{\boldsymbol{v}}_i, A\vec{\boldsymbol{v}}_j \rangle$. This can be readily seen from the algorithm. In stage 2(b)i, the vector $\vec{\boldsymbol{v}}_{j+1}$ is given by

$$A\vec{\boldsymbol{v}}_j - \sum_{k=0}^{i-1} \Gamma_{k,j}\vec{\boldsymbol{v}}_k$$

with the summation convention of Eq. (38). Using the orthonormality relations between the vectors, we obtain immediately:

$$\Gamma_{i,j} = \left\langle \vec{\boldsymbol{v}}_i, A\vec{\boldsymbol{v}}_j - \sum_{k=0}^{i-1} \Gamma_{k,j}\vec{\boldsymbol{v}}_k \right\rangle = \langle \vec{\boldsymbol{v}}_i, A\vec{\boldsymbol{v}}_j \rangle = A_{i,j}, \qquad i \le j \tag{168}$$

It can also be shown that $A_{j+1,j}$ is equivalent to $\Gamma_{j+1,j}$, computed in stage 2(d). In analogy to stage 2(b)i, the matrix element $A_{j+1,j}$ is given by the projection of $\vec{\boldsymbol{v}}_{j+1}$ in stage 2(d), on the final $\vec{\boldsymbol{v}}_{j+1}$, obtained in stage 2(e). In stage 2d, $\vec{\boldsymbol{v}}_{j+1}$ is in its final direction, but is still unnormalized. The projection of a vector on a unit vector in the same direction gives its norm. This clarifies the definition of $\Gamma_{j+1,j}$ in stage 2(b)i.

The matrix elements $\Gamma_{i,j}$ with $i > j+1$ are not computed in the algorithm. It can be easily seen that they vanish. The expression $A\vec{\boldsymbol{v}}_j$ belongs to the Krylov space of dimension $j+2$. The vectors $\vec{\boldsymbol{v}}_i$ with $i > j+1$ are orthogonal to this space, and hence, the scalar product vanishes.

Now we are able to construct the matrix representation of A in the reduced orthonormalized Krylov basis. It will be denoted by Γ. The matrix is a square matrix of dimension $L+1$, with the following structure:

$$\Gamma \equiv \begin{bmatrix} \Gamma_{0,0} & \Gamma_{0,1} & \Gamma_{0,2} & \cdots & \Gamma_{0,L-1} & \Gamma_{0,L} \\ \Gamma_{1,0} & \Gamma_{1,1} & \Gamma_{1,2} & \cdots & \Gamma_{1,L-1} & \Gamma_{1,L} \\ & \Gamma_{2,1} & \Gamma_{2,2} & \cdots & \Gamma_{2,L-1} & \Gamma_{2,L} \\ & & \Gamma_{3,2} & \cdots & \Gamma_{3,L-1} & \Gamma_{3,L} \\ & \mathbf{0} & & \ddots & \vdots & \vdots \\ & & & & \Gamma_{L,L-1} & \Gamma_{L,L} \end{bmatrix} \tag{169}$$

The matrix structure is similar to an upper triangular matrix, with zero elements below the first subdiagonal. A matrix of this structure is called an *upper Hessenberg matrix*. In the context of the Arnoldi process, Γ is referred as the *Hessenberg matrix of A*. Note that in the algorithm, we compute also $\Gamma_{L+1,L}$, which is not included in the definition of Γ. It will be useful to define also an extended matrix $\bar{\Gamma}$, of dimension $(L+2)\times(L+1)$. $\bar{\Gamma}$ is given by the extension of Γ by one row, in the following way:

$$\bar{\Gamma} \equiv \begin{bmatrix} & & \Gamma & & \\ & & & & \\ 0 & 0 & \cdots & 0 & \Gamma_{L+1,L} \end{bmatrix} \tag{170}$$

Note that the last column of $\bar{\Gamma}$ is computed in the algorithm during the process of obtaining $\vec{\boldsymbol{v}}_{L+1}$. We see that this process is necessary for the computation of the Hessenberg matrix, even though $\vec{\boldsymbol{v}}_{L+1}$ does not participate in the approximation itself, which takes place in a Krylov space of dimension $L+1$ only. The computation of an additional vector costs an additional matrix–vector multiplication. Thus, in the Arnoldi process, we need $L+1$ matrix–vector multiplications, in comparison to L for a polynomial approximation of the same order. Nevertheless, the extension of the Krylov space by one dimension is useful for the estimation of the error of the approximation, as will be seen.

The Arnoldi process can be summarized by $L+1$ vector equations in the following way:

$$\vec{\boldsymbol{v}}_{n+1} = \frac{A\vec{\boldsymbol{v}}_n - \sum_{k=0}^{n} \Gamma_{k,n}\vec{\boldsymbol{v}}_k}{\Gamma_{n+1,n}}, \qquad 0 \le n \le L \tag{171}$$

or,

$$A\vec{\boldsymbol{v}}_n = \sum_{k=0}^{n+1} \Gamma_{k,n}\vec{\boldsymbol{v}}_k, \qquad 0 \le n \le L \tag{172}$$

These equations can be written compactly in a matrix form:

$$A\Upsilon = \bar{\Upsilon}\bar{\Gamma} \tag{173}$$

where Υ is an $N \times (L+1)$ matrix (N denotes the dimension of A) which its columns are the $\vec{\upsilon}_n$'s:

$$\Upsilon \equiv \left[\vec{\upsilon}_0, \quad \vec{\upsilon}_1, \quad \cdots \quad \vec{\upsilon}_L \right] \tag{174}$$

and $\bar{\Upsilon}$ is an extension of Υ by one column, which contains $\vec{\upsilon}_{L+1}$:

$$\bar{\Upsilon} \equiv \left[\vec{\upsilon}_0, \quad \vec{\upsilon}_1, \quad \cdots \quad \vec{\upsilon}_L, \quad \vec{\upsilon}_{L+1} \right] \tag{175}$$

We can write Eq. (173) in an alternative form, which does not involve the extended matrices. We utilize the fact that the last row of $\bar{\Gamma}$ contains only one non-zero element, which affects only the last column of the resulting $N \times (L+1)$ matrix. It can be easily seen that we have:

$$A\Upsilon = \Upsilon\Gamma + \Gamma_{L+1,L}\, \vec{\upsilon}_{L+1}\, \vec{\mathbf{e}}_{L+1}^{T} \tag{176}$$

where $\vec{\mathbf{e}}_n$ denotes a unit vector of dimension $L+1$, which its j'th element is given by $\delta_{j,n}$.

Now we are about to use the reduced basis, obtained by the Arnoldi iteration, for writing an approximation of $\vec{\mathbf{u}}$. The $\vec{\upsilon}_n$'s in the orthonormalized Krylov basis representation are given by $\vec{\mathbf{e}}_{n+1}$. Thus, we can define the vector which represents $\vec{\mathbf{v}}$ in the reduced basis in the following way:

$$\vec{\omega} \equiv \|\vec{\mathbf{v}}\|\, \vec{\mathbf{e}}_1 \tag{177}$$

Υ has an important significance—it is the transformation matrix from the reduced Krylov basis representation to the original N dimensional representation of A and $\vec{\mathbf{v}}$:

$$\vec{\upsilon}_n = \Upsilon \vec{\mathbf{e}}_{n+1} \tag{178}$$

Following the reasoning that was introduced in the beginning of this section, we can approximate $\vec{\mathbf{u}}$ by performing the calculation in the reduced basis representation, and transforming the result to the original basis. Let us define the corresponding vector of $f(A)\vec{\mathbf{v}}$ in the reduced representation:

$$\vec{\eta} \equiv f(\Gamma)\vec{\omega} \tag{179}$$

$\vec{\eta}$ is transformed to the original representation, to yield the following approximation of $\vec{\mathbf{u}}$:

$$\vec{\mathbf{u}} \approx \Upsilon \vec{\eta} \tag{180}$$

As was mentioned in Sec. 2.3.1, the direct computation of a function of matrix via diagonalization, in the form of Eq. (14), is very expensive numerically for a large scale matrix like A. However, it is not an expensive operation to diagonalize Γ, which is a small scale matrix, in order to calculate $\vec{\eta}$. Thus, $\vec{\eta}$ can be computed as

$$\vec{\eta} = S f(D) S^{-1} \vec{\omega} \tag{181}$$

where D is the diagonal matrix which represents Γ in the basis of the Γ's eigenvectors, and S is the transformation matrix from the diagonalized basis representation to the $\vec{\upsilon}_n$ basis representation.

In the approximation obtained, we do not need any previous knowledge about the eigenvalue domain of A. Moreover, we do not rely on any assumption on the form of the eigenvalue domain. Hence, it can be applied also for A with a complex spectrum.

The justification that was given to the approximation in the form of Eq. (180) is intuitive rather than rigorous. Hence, it becomes unclear what is the quality of the approximation, and how to estimate the resulting error. In what follows, we shall give a fuller justification to the Arnoldi approach approximation. It will be shown that Eq. (180) is actually equivalent to a *polynomial approximation* of $\vec{\mathbf{u}}$. This will enable us to estimate the error of the approximation.

Let us return, for the moment, to the polynomial approximation methods. We shall present a new approach for the choice of the approximation polynomial $Q_L(z)$ (see Eqs. (151), (152)), which is led by different considerations from the polynomial approximations of Sec. B.1.

Until now, we used the concept of interpolation as an approximation tool for a function. Now we shall see that there exists a more intimate relation between a function of matrix and the interpolation concept. Let $g(z)$ be a function which interpolates $f(z)$ in the eigenvalues of A, i.e.:

$$g(\lambda_j) = f(\lambda_j), \qquad j = 0, 1, \ldots, N-1 \tag{182}$$

It can be easily shown that when the set of eigenvectors spans the N dimensional space, then $g(A)$ is completely equivalent to $f(A)$:

$$g(A) = f(A). \tag{183}$$

This can be proved by showing that

$$[f(A) - g(A)]\vec{\mathbf{w}} = \vec{\mathbf{0}}$$

for an arbitrary vector $\vec{\mathbf{w}}$. Let us expand $\vec{\mathbf{w}}$ in the eigenvectors of A:

$$\vec{\mathbf{w}} = \sum_{j=0}^{N-1} w_j \vec{\boldsymbol{\varphi}}_j \tag{184}$$

This yields:

$$[f(A) - g(A)]\vec{\mathbf{w}} = \sum_{j=0}^{N-1} w_j [f(\lambda_j) - g(\lambda_j)]\vec{\boldsymbol{\varphi}}_j = \vec{\mathbf{0}} \tag{185}$$

The condition (182) on $g(z)$ is necessary for the equivalence of the *matrices* (Eq. (183)). Note that the condition for the *vector* equivalence $g(A)\vec{\mathbf{v}} = f(A)\vec{\mathbf{v}}$ may be even weaker. This happens when $\vec{\mathbf{v}}$ is spanned by a subspace of several eigenvectors. Then, it is sufficient to choose $g(z)$ which interpolates $f(z)$ in the eigenvalues of these eigenvectors only, as is apparent from (185).

We see that in order to obtain $f(A)$, we needn't know the behavior of $f(z)$ everywhere, but only its values at the λ_j's. The origin of this somewhat surprising finding lies in the fact that the spectrum of A is *discrete*. The discrete spectrum originates in the fact that A itself is a discrete entity. Thus, $f(A)$ itself is also a discrete entity, and its approximation requires discrete knowledge on $f(z)$.

This leads to a different approach for the choice of $Q_L(z)$: Instead of trying to approximate the full behavior of $f(z)$ by the ideal approximation polynomial of $f(z)$ in the domain, we can choose a polynomial which well represents the behavior of $f(z)$ in several representative eigenvalues of A. This feature is satisfied by an *interpolation polynomial of* $f(z)$ in these eigenvalues. The question that remains is: How can we know the eigenvalues of A, without diagonalizing it?

Here the Arnoldi approach enters: We state that the eigenvalues of the reduced matrix Γ provide estimation of several representative eigenvalues of A itself. These can be used as interpolation points of $f(z)$. The eigenvalues of Γ tend to be distributed in the eigenvalue domain of A in the same way as the whole eigenvalue spectrum of A. In the case that several eigenvectors dominate the composition of $\vec{\mathbf{v}}$, the spectrum of Γ usually contains estimation of most of them. This is because of the second feature of the Krylov space mentioned in the beginning of this section—the Krylov space remains in the same eigenvector subspace as $\vec{\mathbf{v}}$, and reflects its eigenvector composition, at least to some extent. This characteristic of the Γ spectrum is particularly useful for the specific approximation of $f(A)\vec{\mathbf{v}}$, since the interpolation polynomial becomes adjusted to represent the behavior of $f(z)$ in the dominant eigenvectors in $\vec{\mathbf{v}}$.

Let us denote the eigenvalues of Γ by $\tilde{\lambda}_j$. Our approximation polynomial is a Newton interpolation polynomial,

$$Q_L(z) = \sum_{n=0}^{L} a_n R_n(z) \tag{186}$$

with

$$\begin{aligned} R_0(z) &\equiv 1 \\ R_n(z) &\equiv \prod_{j=0}^{n-1} (z - \tilde{\lambda}_j), \qquad n > 0 \end{aligned} \tag{187}$$

$\vec{\mathbf{u}}$ is approximated in the form of Eq. (158). If we follow the scheme of Sec. B.1.1, we need L matrix–vector multiplications for this operation. This is in addition to the $L+1$ matrix–vector multiplications required for the Arnoldi iteration process. It seems that the overall cost of this approximation is more than twice the cost of the approximations of Sec. B.1. Actually, we can avoid any additional large-scale matrix–vector multiplication, as will be seen. Thus, the overall number of the required large-scale matrix–vector multiplications will remain $L+1$.

In order to show that, we shall use the first feature of the Krylov space, mentioned in the beginning of this section. Let us consider again Eq. (152), with an arbitrary $Q_L(z)$. Both $\vec{\mathbf{v}}$ and the approximated $\vec{\mathbf{u}}$ lie in the Krylov space of dimension $L+1$, since $Q_L(A)\vec{\mathbf{v}}$ can be decomposed into the Taylor polynomial vectors (167). Thus, the operation of the matrix $Q_L(A)$ on $\vec{\mathbf{v}}$ takes place in the reduced Krylov space, as well as the operation of any of the $Q_L(A)$'s polynomial components. Hence, the whole calculation can take place in the reduced space, where A is replaced by its reduced representation, Γ, and $\vec{\mathbf{v}}$ is replaced by $\vec{\boldsymbol{\omega}}$. The result is transferred back to the full N dimensional space, to yield the same $\vec{\mathbf{u}}$ as in Eq. (152), without any further approximation. Thus, the following equality is *exact* for *any approximation polynomial* $Q_L(z)$:

$$Q_L(A)\vec{\mathbf{v}} = \Upsilon\, Q_L(\Gamma)\vec{\boldsymbol{\omega}} \tag{188}$$

In our particular case, $\vec{\mathbf{u}}$ is approximated as

$$\vec{\mathbf{u}} \approx \Upsilon \sum_{n=0}^{L} a_n R_n(\Gamma)\vec{\boldsymbol{\omega}} \tag{189}$$

Now we note that our approximation polynomial, which interpolates $f(z)$ in the entire spectrum of Γ, is actually equivalent to $f(\Gamma)$, by Eq. (183). Thus, we have (cf. (179)):

$$\vec{\eta} = \sum_{n=0}^{L} a_n R_n(\Gamma)\vec{\omega} \tag{190}$$

Then, Eq. (189) becomes equivalent to Eq. (180). This concludes the justification to the approximation of Eq. (180), which is shown to be equivalent to a specific polynomial approximation.

The new interpretation of the Arnoldi approximation enables us to estimate the resulting error. Assuming fast convergence of the polynomial expansion with n, we can estimate the truncation error of Eq. (189) by the magnitude of the next term in the sum,

$$E = |a_{L+1}| \, \|R_{L+1}(A)\vec{\mathbf{v}}\| \tag{191}$$

In order to specify the next term, we first have to choose an additional sampling point. The additional point should be in the eigenvalue domain of A, which is unknown. It is reasonable to choose the average point of the estimated eigenvalues:

$$z_{L+1} = \frac{\sum_{j=0}^{L} \tilde{\lambda}_j}{L+1} \tag{192}$$

It should be verified that z_{L+1} is not equal to any of the $\tilde{\lambda}_j$'s. $R_{L+1}(z)$ is independent of z_{L+1}, and is given by Eq. (187) with $n = L+1$. z_{L+1} determines a_{L+1} only.

The error should be computed without additional large-scale matrix–vector multiplications. Hence, it is desirable to express $R_{L+1}(A)\vec{\mathbf{v}}$ in the terms of the reduced Krylov space, which involves only small-scale calculations. Here we encounter the problem that the expression $R_{L+1}(A)\vec{\mathbf{v}}$ belongs to a Krylov space of dimension $L+2$. Hence, A cannot be simply replaced by Γ, which contains information on the Krylov space of dimension $L+1$ only. However, the computation of the reduced basis vectors involved also the additional vector, $\vec{\boldsymbol{v}}_{L+1}$. Thus, it is still possible to obtain a reduced calculation of (191) by the information we already have. First, we write:

$$R_{L+1}(A)\vec{\mathbf{v}} = (A - \tilde{\lambda}_L)R_L(A)\vec{\mathbf{v}} \tag{193}$$

$R_L(A)\vec{\mathbf{v}}$ lies in the Krylov space of dimension $L+1$. Hence, we can express it in the terms of the $L+1$ dimensional space:

$$R_L(A)\vec{\mathbf{v}} = \Upsilon R_L(\Gamma)\vec{\omega} = \Upsilon\vec{\mu} \tag{194}$$

where we defined:

$$\vec{\mu} \equiv R_L(\Gamma)\vec{\omega} \tag{195}$$

Eq. (193) becomes:

$$R_{L+1}(A)\vec{\mathbf{v}} = A\Upsilon\vec{\mu} - \tilde{\lambda}_k\Upsilon\vec{\mu} \tag{196}$$

Then, we apply Eq. (176) to yield:

$$\begin{aligned} R_{L+1}(A)\vec{\mathbf{v}} &= \left(\Upsilon\Gamma + \Gamma_{L+1,L}\,\vec{\boldsymbol{v}}_{L+1}\,\vec{\mathbf{e}}_{L+1}^{T}\right)\vec{\mu} - \tilde{\lambda}_k\Upsilon\vec{\mu} \\ &= \Upsilon\left(\Gamma - \tilde{\lambda}_k\right)\vec{\mu} + \Gamma_{L+1,L}\mu_{L+1}\,\vec{\boldsymbol{v}}_{L+1} \\ &= \Upsilon R_{L+1}(\Gamma)\vec{\omega} + \Gamma_{L+1,L}\mu_{L+1}\,\vec{\boldsymbol{v}}_{L+1} \end{aligned} \tag{197}$$

where μ_{L+1} is the $(L+1)$'th element of $\vec{\mu}$. We see that the resulting expression is spanned by the extended orthonormalized Krylov set, which includes $\vec{\boldsymbol{v}}_{L+1}$. We can use an extended representation of the extended set, of dimension $L+2$, in order to represent $R_{L+1}(A)\vec{\mathbf{v}}$. In the extended representation, $R_{L+1}(A)\vec{\mathbf{v}}$ is defined by the following vector:

$$\vec{\tilde{\mu}} = \begin{bmatrix} R_{L+1}(\Gamma)\vec{\omega} \\ \Gamma_{L+1,L}\mu_{L+1} \end{bmatrix} \tag{198}$$

In principle, $\vec{\tilde{\mu}}$ can be transferred back to the original representation by the extended transformation matrix, $\bar{\Upsilon}$, in the following way:

$$R_{L+1}(A)\vec{\mathbf{v}} = \bar{\Upsilon}\vec{\tilde{\mu}} \tag{199}$$

However, this operation is unnecessary, as long as we are interested only in the *norm* of this expression (see (191)). The vectors in $\bar{\Upsilon}$ are *unit vectors*, and the norm remains unaltered by the change of representation:

$$\|\bar{\Upsilon}\vec{\tilde{\mu}}\| = \|\vec{\tilde{\mu}}\| \tag{200}$$

The error is finally given by

$$E = |a_{L+1}| \, \|\vec{\boldsymbol{\mu}}\| \tag{201}$$

The relative error is given by

$$E_{rel} = \frac{E}{\|\vec{\mathbf{u}}\|} \approx \frac{E}{\|\Upsilon \vec{\boldsymbol{\eta}}\|} = \frac{E}{\|\vec{\boldsymbol{\eta}}\|} \tag{202}$$

If we compute the error, Eq. (181) becomes unnecessary; $\vec{\boldsymbol{\eta}}$ should be computed by the Newton interpolation form, Eq. (190), utilizing the operations needed for the computation of the error. The $R_n(\Gamma)\vec{\boldsymbol{\omega}}$ terms and the a_n's are computed iteratively, as explained in Appendix A.1 and Sec. B.1.1. $R_{L+1}(\Gamma)\vec{\boldsymbol{\omega}}$ and a_{L+1} for the error estimation are obtained by continuing the iterative process to the next order.

In order to increase the numerical stability of the Newton interpolation, the size of the approximation domain should be changed, as in the case of interpolation on a one-dimensional axis (see Appendix A.1.3). In the case of a two-dimensional domain, defined on the complex plain, the problem should be transformed to a domain which its size is divided by the *capacity of the domain* [45]. We give only an expression for the estimation of the capacity, and not an exact definition. The capacity is estimated by choosing a point in the domain, z_p, and computing the following expression:

$$\rho = \left(\left|z_p - z_0\right| \left|z_p - z_1\right| \cdots \left|z_p - z_N\right|\right)^{\frac{1}{N+1}} \tag{203}$$

where the z_n's ($n = 0, 1, \ldots, N$) are the sampling points, and $N+1$ is the number of sampling points. z_p can be chosen as the average point of the sampling points, as in Eq. (192). In the case of interpolation on an axis, in a domain of length 4 (see Appendix A.1.3), the capacity is 1. This can be observed intuitively from (203), by choosing z_p in the middle of the domain. If the capacity of the domain is different from 1, we should perform a transformation similar to that described in Appendix A.1.3, where the conversion factor of $4/\Delta x$ is replaced by $1/\rho$. In practice, the instructions at the end of Sec. B.1.1 should be followed, with the appropriate conversion factor.

The interpolation polynomial interpretation of the Arnoldi approach reveals an important advantage of the Arnoldi algorithm over the polynomial approximations. This advantage exists in certain circumstances, even for A with eigenvalues distributed on a one-dimensional axis. If most of the eigenvalues are concentrated in a portion of the eigenvalue domain, with an additional small number of spread eigenvalues in the whole domain, the Arnoldi approach is expected to be more efficient. The reason is that the $\tilde{\lambda}$'s are distributed in the eigenvalue domain in a similar way to the λ's. Thus, the important region in which most of the eigenvalues are concentrated is better represented than the less important regions. In contrary, the Chebyshev sampling is uniformly distributed in the domain.

In the application of the time-dependent Hamiltonian propagator, we apply a time-step scheme. The short time intervals can be treated by a relatively small Krylov space, typically—up to $L = 15$. When the Hamiltonian is time-independent, the whole time-interval is treated in a single step, and the required approximation space is large. The Arnoldi approach usually becomes problematic in a large space. The main reasons are:

1. In the Arnoldi process, we store $L+2$ vectors in the memory. For very large N and large L, this might be impermissible.
2. During the Arnoldi process, we perform $(L+1)^2/2$ scalar products, each of which scales as N. This becomes quite demanding for large L.

When a large dimension approximation is required, a *restarted Arnoldi algorithm* should be used (see, for example, [47]). This topic is beyond the scope of this paper.

Appendix C. Conversion schemes of polynomial expansions to a Taylor form

C.1. Conversion scheme for a Newton expansion

C.1.1. Conversion scheme for the $q_{n,m}$ coefficients

Let us write the Newton expansion form for $\vec{\mathbf{s}}(t)$ (cf. Eq. (107)):

$$\vec{\mathbf{s}}(t) \approx \sum_{n=0}^{M-1} \vec{\mathbf{a}}_n R_n(t) \tag{204}$$

The $R_n(t)$'s satisfy the following recursion formula (see Eq. (108)):

$$R_{n+1}(t) = (t - t_n) R_n(t) \tag{205}$$

where

$$R_0(t) = 1 \tag{206}$$

and the t_n's are the sampling points.

We need the conversion coefficients of the $R_n(t)$'s to a Taylor form (cf. Eq. (51)):

$$R_n(t) = \sum_{m=0}^{n} q_{n,m} \frac{t^m}{m!} \tag{207}$$

It can be immediately observed from Eq. (206) that:

$$q_{0,0} = 1 \tag{208}$$

The rest of the $q_{n,m}$'s can be computed from Eq. (208) by the derivation of recurrence relations. Plugging Eq. (207) into (205) we obtain:

$$R_{n+1}(t) = \sum_{m=0}^{n} q_{n,m} \frac{t^{m+1}}{m!} - t_n \sum_{m=0}^{n} q_{n,m} \frac{t^m}{m!} = \sum_{m=1}^{n+1} q_{n,m-1} \frac{t^m}{(m-1)!} - t_n \sum_{m=0}^{n} q_{n,m} \frac{t^m}{m!} \tag{209}$$

On the other hand, we have from (207):

$$R_{n+1}(t) = \sum_{m=0}^{n+1} q_{n+1,m} \frac{t^m}{m!} \tag{210}$$

Equating the RHS of Eqs. (209) and (210) we obtain:

$$\sum_{m=0}^{n+1} q_{n+1,m} \frac{t^m}{m!} = \sum_{m=1}^{n+1} q_{n,m-1} \frac{t^m}{(m-1)!} - t_n \sum_{m=0}^{n} q_{n,m} \frac{t^m}{m!} \tag{211}$$

Equating the coefficients of similar powers of t, we obtain the following recurrence relations for the $q_{n,m}$'s:

$$q_{n+1,0} = -t_n q_{n,0} \tag{212}$$

$$q_{n+1,m} = m q_{n,m-1} - t_n q_{n,m} \qquad 1 \le m \le n \tag{213}$$

$$q_{n+1,n+1} = (n+1) q_{n,n} \tag{214}$$

Starting from Eq. (208), all $q_{m,n}$'s can be computed in a recursive manner, using Eqs. (212)–(214).

Once the $q_{n,m}$'s are obtained, the $\vec{\mathbf{s}}_m$ Taylor coefficients can be computed by (cf. Eq. (54)):

$$\vec{\mathbf{s}}_m = \sum_{n=m}^{M-1} q_{n,m} \vec{\mathbf{a}}_n \tag{215}$$

C.1.2. Conversion scheme for the $\tilde{q}_{n,m}$ coefficients

As was mentioned in Sec. 3.3, for numerical stability, it is recommended to absorb the $1/m!$ factor in the $\vec{\mathbf{s}}_m$ Taylor coefficients from Eq. (22). Accordingly, Eq. (207) is replaced by:

$$R_n(t) = \sum_{m=0}^{n} \tilde{q}_{n,m} t^m \tag{216}$$

The recurrence relations for the $\tilde{q}_{n,m}$'s are slightly different from those of the $q_{n,m}$'s. Following the same steps as above, we obtain the recursion formulas:

$$\tilde{q}_{n+1,0} = -t_n \tilde{q}_{n,0} \tag{217}$$

$$\tilde{q}_{n+1,m} = \tilde{q}_{n,m-1} - t_n \tilde{q}_{n,m} \qquad 1 \le m \le n \tag{218}$$

$$\tilde{q}_{n+1,n+1} = \tilde{q}_{n,n} \tag{219}$$

In addition, we have from (206):

$$\tilde{q}_{0,0} = 1 \tag{220}$$

which completes the required information for computing the $\tilde{q}_{n,m}$'s recursively.

The $\vec{\tilde{\mathbf{s}}}_m$'s are computed by:

$$\vec{\tilde{\mathbf{s}}}_m = \sum_{n=m}^{M-1} \tilde{q}_{n,m} \vec{\mathbf{a}}_n \tag{221}$$

C.1.3. Conversion scheme for a length 4 domain

We mentioned in Sec. A.1 that it is recommended to use a domain of length 4 in a Newton expansion, for numerical stability. When transferring the original t domain to a length 4 domain, the recursion formula for the $R_n(t)$'s gains an additional conversion factor (cf. Eq. (205)):

$$R_{n+1}(t) = \frac{4}{\Delta t}(t - t_n)R_n(t) \tag{222}$$

where Δt is the length of the original domain. Accordingly, the RHS of the recurrence relations (212)–(214) and (217)–(219) is multiplied by the same factor, to yield:

$$q_{n+1,0} = -\frac{4}{\Delta t}t_n q_{n,0} \tag{223}$$

$$q_{n+1,m} = \frac{4}{\Delta t}(m q_{n,m-1} - t_n q_{n,m}) \qquad 1 \le m \le n \tag{224}$$

$$q_{n+1,n+1} = \frac{4}{\Delta t}(n+1)q_{n,n} \tag{225}$$

and

$$\tilde{q}_{n+1,0} = -\frac{4}{\Delta t}t_n \tilde{q}_{n,0} \tag{226}$$

$$\tilde{q}_{n+1,m} = \frac{4}{\Delta t}(\tilde{q}_{n,m-1} - t_n \tilde{q}_{n,m}) \qquad 1 \le m \le n \tag{227}$$

$$\tilde{q}_{n+1,n+1} = \frac{4}{\Delta t}\tilde{q}_{n,n} \tag{228}$$

The $\vec{\mathbf{s}}_m$'s and the $\vec{\tilde{\mathbf{s}}}_m$'s are computed as in Eqs. (215), (221).

C.2. Conversion scheme for a Chebyshev expansion

The Chebyshev expansion is defined for the domain $[-1, 1]$. Suppose we want approximate $\vec{\mathbf{s}}(t)$ in an arbitrary domain $t \in [t_{min}, t_{max}]$ by a Chebyshev expansion. We define a variable $y \in [-1, 1]$ such that

$$y \equiv \frac{2t - t_{min} - t_{max}}{\Delta t} \tag{229}$$

where $\Delta t = t_{max} - t_{min}$ (see Sec. A.2). Then we expand $\vec{\mathbf{s}}(t)$ in a Chebyshev series:

$$\vec{\mathbf{s}}(t) \approx \sum_{n=0}^{M-1} \vec{\mathbf{c}}_n T_n(y) = \sum_{n=0}^{M-1} \vec{\mathbf{c}}_n T_n\left(\frac{2t - t_{min} - t_{max}}{\Delta t}\right) \tag{230}$$

Let us define the following set of polynomials:

$$\phi_n(t) \equiv T_n\left(\frac{2t - t_{min} - t_{max}}{\Delta t}\right) \tag{231}$$

Note that (229) is a linear transformation. Hence, $\phi_n(t)$ remains a polynomial of degree n, like $T_n(y)$. The Chebyshev expansion can be rewritten as a polynomial series in the terms of the $\phi_n(t)$'s:

$$\vec{\mathbf{s}}(t) \approx \sum_{n=0}^{M-1} \vec{\mathbf{c}}_n \phi_n(t) \tag{232}$$

Using this form, the coefficients of the Taylor like form, $\vec{\mathbf{s}}_m$, can be computed from the Chebyshev coefficients, $\vec{\mathbf{c}}_m$, via Eq. (54).

First, we expand the $\phi_n(t)$'s in a Taylor form (cf. Eq. (51)):

$$\phi_n(t) = \sum_{m=0}^{n} q_{n,m}\frac{t^m}{m!} \tag{233}$$

We can utilize the recurrence relation between the Chebyshev polynomials in order to find the $q_{n,m}$'s. The Chebyshev polynomials satisfy the following recursion formula:

$$T_{n+1}(y) = 2yT_n(y) - T_{n-1}(y) \tag{234}$$

where

$$T_0(y) = 1 \tag{235}$$

$$T_1(y) = y \tag{236}$$

The recursion formula can be rewritten in the terms of t and the $\phi_n(t)$'s:

$$\phi_{n+1}(t) = \frac{4t - 2(t_{min} + t_{max})}{\Delta t}\phi_n(t) - \phi_{n-1}(t) \tag{237}$$

where

$$\phi_0(t) = 1 \tag{238}$$

$$\phi_1(t) = \frac{2t - t_{min} - t_{max}}{\Delta t} \tag{239}$$

It can be observed from Eqs. (238), (239), that:

$$q_{0,0} = 1 \tag{240}$$

$$q_{1,0} = -\frac{t_{min} + t_{max}}{\Delta t} \tag{241}$$

$$q_{1,1} = \frac{2}{\Delta t} \tag{242}$$

Here, again, the rest of the $q_{n,m}$'s can be obtained from Eqs. (240)–(242) by the derivation of recurrence relations.

Plugging Eq. (233) into (237) we obtain:

$$\begin{aligned}\phi_{n+1}(t) &= \frac{4}{\Delta t}\sum_{m=0}^{n} q_{n,m}\frac{t^{m+1}}{m!} - \frac{2(t_{min} + t_{max})}{\Delta t}\sum_{m=0}^{n} q_{n,m}\frac{t^m}{m!} - \sum_{m=0}^{n-1} q_{n-1,m}\frac{t^m}{m!} \\ &= \frac{4}{\Delta t}\sum_{m=1}^{n+1} q_{n,m-1}\frac{t^m}{(m-1)!} - \frac{2(t_{min} + t_{max})}{\Delta t}\sum_{m=0}^{n} q_{n,m}\frac{t^m}{m!} - \sum_{m=0}^{n-1} q_{n-1,m}\frac{t^m}{m!}\end{aligned} \tag{243}$$

On the other hand, from (233) we have:

$$\phi_{n+1}(t) = \sum_{m=0}^{n+1} q_{n+1,m}\frac{t^m}{m!} \tag{244}$$

From Eqs. (243), (244), we obtain:

$$\sum_{m=0}^{n+1} q_{n+1,m}\frac{t^m}{m!} = \frac{4}{\Delta t}\sum_{m=1}^{n+1} q_{n,m-1}\frac{t^m}{(m-1)!} - \frac{2(t_{min} + t_{max})}{\Delta t}\sum_{m=0}^{n} q_{n,m}\frac{t^m}{m!} - \sum_{m=0}^{n-1} q_{n-1,m}\frac{t^m}{m!} \tag{245}$$

The recurrence relations are obtained by equating the coefficients of similar powers of t:

$$q_{n+1,0} = -\frac{2(t_{min} + t_{max})}{\Delta t}q_{n,0} - q_{n-1,0} \tag{246}$$

$$q_{n+1,m} = \frac{4}{\Delta t}mq_{n,m-1} - \frac{2(t_{min} + t_{max})}{\Delta t}q_{n,m} - q_{n-1,m} \qquad 1 \le m \le n-1 \tag{247}$$

$$q_{n+1,n} = \frac{4}{\Delta t}nq_{n,n-1} - \frac{2(t_{min} + t_{max})}{\Delta t}q_{n,n} \tag{248}$$

$$q_{n+1,n+1} = \frac{4}{\Delta t}(n+1)q_{n,n} \tag{249}$$

The Taylor like coefficients are given by (cf. Eq. (54)):

$$\vec{\mathbf{s}}_m = \sum_{n=m}^{M-1} q_{n,m}\vec{\mathbf{c}}_n \tag{250}$$

The $\tilde{q}_{n,m}$'s can be computed in an analogous manner to the $q_{n,m}$'s. From Eqs. (238), (239), we have:

$$\tilde{q}_{0,0} = 1 \tag{251}$$

$$\tilde{q}_{1,0} = -\frac{t_{min} + t_{max}}{\Delta t} \tag{252}$$

$$\tilde{q}_{1,1} = \frac{2}{\Delta t} \tag{253}$$

Using the same technique as for the $q_{n,m}$'s, we obtain the following recurrence relations:

$$\tilde{q}_{n+1,0} = -\frac{2(t_{min} + t_{max})}{\Delta t}\tilde{q}_{n,0} - \tilde{q}_{n-1,0} \tag{254}$$

$$\tilde{q}_{n+1,m} = \frac{4}{\Delta t}\tilde{q}_{n,m-1} - \frac{2(t_{min} + t_{max})}{\Delta t}\tilde{q}_{n,m} - \tilde{q}_{n-1,m} \qquad 1 \le m \le n-1 \tag{255}$$

$$\tilde{q}_{n+1,n} = \frac{4}{\Delta t}\tilde{q}_{n,n-1} - \frac{2(t_{min} + t_{max})}{\Delta t}\tilde{q}_{n,n} \tag{256}$$

$$\tilde{q}_{n+1,n+1} = \frac{4}{\Delta t}\tilde{q}_{n,n} \tag{257}$$

The $\vec{\tilde{\mathbf{s}}}_m$'s are given by

$$\vec{\tilde{\mathbf{s}}}_m = \sum_{n=m}^{M-1} \tilde{q}_{n,m}\vec{\mathbf{c}}_n \tag{258}$$

Appendix D. Error estimation and control

One of the most important issues in any numerical method is the ability to estimate the magnitude of the error of the method. When the error can be estimated, it can usually be controlled by altering the parameters of the approximation. Thus, we should point out the possible sources of inaccuracy in the propagation procedure, and provide estimations of the magnitude of the error.

First, we focus on the error of the local solution *in a given time-step and a given iteration*. Then, we discuss the relation between the local errors and the global error of the algorithm, i.e. the error of the whole propagation process.

D.1. Local error

The local solution is computed in step 2(c)vi of the algorithm in Sec. 3.2. There are three sources of inaccuracy in this computation:

1. *Convergence error*: The computation is based on the previous $\vec{\mathbf{u}}(t_{k,l})$, i.e. the guess solution or the solution from the previous iteration (step 2(c)i);
2. *Time-discretization error*: The time behavior of $\vec{\mathbf{s}}_{ext}(\vec{\mathbf{u}}(t), t)$ is approximated from sampling at the discrete Chebyshev points (steps 2(c)i, 2(c)ii);
3. *Function of matrix computation error*: $f_{M_k}(\widetilde{G}, t_{k,l} - t_{k,0})\vec{\mathbf{v}}_{M_k}$ is approximated by a polynomial expansion, or by the Arnoldi approach.

The different sources of inaccuracy result in an inadequate representation of Eq. (65) by the algorithm. The effects of the different inaccuracy sources on this representation vary. In any case, the algorithm does not represent Eq. (65) accurately, but something else. It is important to understand what the algorithm does represent, for the understanding of the behavior of the algorithm in each situation in which the algorithm fails to yield the required accuracy. The different situations can be classified as follows:

1. The algorithm represents an equation which differs from Eq. (65); we can distinguish between two situations, which correspond to different sources of inaccuracy:
   (a) *Time-discretization error*: The time-sampling is insufficient to represent $G(t)$ or $\vec{\mathbf{s}}(t)$ properly. We can view this situation in the following way: We actually solve an equation of the general form of Eq. (56), but for another problem, in which $G(t)$ or $\vec{\mathbf{s}}(t)$ are replaced by their truncated time-expansions;
   (b) *Function of matrix computation error*: The expansion of $f_{M_k}(\widetilde{G}, t_{k,l} - t_{k,0})\vec{\mathbf{v}}_{M_k}$ does not approximate the expression properly. In this case, the equation represented by the algorithm does not correspond to an equation of the form of Eq. (56).
2. The algorithm does not represent a continuous *equation* of time, but a discretized *vector problem* in time. The algorithm is based on samplings of $\vec{\mathbf{u}}(t)$ at several discrete time-points, which constitute a time-vector. When the time-sampling is insufficient to represent the time-behavior of $\vec{\mathbf{u}}(t)$ properly, a time-discretization error results. In such a case, the

algorithm does represent the requirement that Eq. (65) will be satisfied at the sampling Chebyshev time-points, but fails to represent the requirement that it will be satisfied at the intermediate points. One outcome is that the resulting $\vec{\mathbf{v}}_j$'s become inaccurate, and so is $\vec{\mathbf{u}}(t)$. Another outcome is that Eq. (65) is not satisfied at the intermediate points, even with the resulting $\vec{\mathbf{v}}_j$'s.

3. The algorithm does not represent an equality at all. Eq. (65) is an implicit equation, because of the dependence of the $\vec{\mathbf{v}}_j$'s on $\vec{\mathbf{u}}(t)$. Step 2(c)vi would have represented it accurately only if the $\vec{\mathbf{u}}(t_{k,l})$'s in step 2(c)i were exact. Because of the convergence error, step 2(c)vi may represent the equality only within the required extent of accuracy. If the convergence error is too large, the algorithm fails to represent the equality to the required accuracy.

In what follows, we give estimations to the magnitude of the error for the three inaccuracy sources, and discuss the ways to control the error for each source.

D.1.1. Convergence error

An estimation of the convergence error is already included in the algorithm in Sec. 3.2, as the convergence criterion in step 2(c)vii. The convergence rate of the iterative process can be assumed to be fast. Consequently, the error of $\vec{\mathbf{u}}_{old}$ is larger by orders of magnitude than the error of the new solution. Hence, the $\vec{\mathbf{u}}(t_{k,M_k-1})$ obtained in step 2(c)vi can practically represent the accurate solution in this context. Thus, the convergence criterion yields an excellent approximation to the relative error of the old solution at the edge of the time-step,

$$\frac{\left\|\vec{\mathbf{u}}_{old}-\vec{\mathbf{u}}(t_{k,M_k-1})\right\|}{\left\|\vec{\mathbf{u}}(t_{k,M_k-1})\right\|}$$

where $\vec{\mathbf{u}}(t)$ here is the *exact solution*. We assumed that $\|\vec{\mathbf{u}}_{old}\|$ in the denominator of step 2(c)vii is close enough to the converged solution, and can safely replace $\|\vec{\mathbf{u}}(t_{k,M_k-1})\|$. Assuming that the iterative process converges, the error of the old solution yields an upper limit to the error of the new one.

The problem with this estimation is that it greatly overestimates the convergence error, since the error of the old solution is assumed to be larger than the error of the new one by several orders of magnitude. Much better estimations to the convergence error of the new solution can be obtained with some extra numerical effort. This topic is left for a future publication.

In the algorithm in Sec. 3.2, the magnitude of the convergence error is controlled by the number of iterations. The convergence error can be controlled also by altering the other parameters of the algorithm. A decrement of the length of the time-step, Δt_k, is effective in the reduction of the convergence error. An increment of the number of expansion terms in the time-expansion *in the previous time-step*, M_{k-1}, may be also helpful, since the guess solution becomes more accurate (unless M_{k-1} becomes too large; see Sec. 3.4). The same is true for K_{k-1}, the number of expansion terms for the function of matrix in the previous time-step.

D.1.2. Time-discretization error

The estimation of the time-discretization error requires some additional insight into the origin of the error. The time-discretization error actually results from the replacement of the exact $\vec{\mathbf{s}}_{ext}(\vec{\mathbf{u}}(t),t)$, by another time-dependent inhomogeneous term, which approximates it by interpolation at the Chebyshev points. Let us denote the approximated $\vec{\mathbf{s}}_{ext}(\vec{\mathbf{u}}(t),t)$ by $\vec{\mathbf{s}}^{int}_{ext}(t)$. The absolute time-discretization error is obtained by the difference between the solutions of two different problems—the approximated problem and the actual one:

$$E^{int}(t)=\left\|\vec{\mathbf{u}}^{int}(t)-\vec{\mathbf{u}}(t)\right\| \tag{259}$$

where $\vec{\mathbf{u}}^{int}(t)$ denotes the solution of the problem with $\vec{\mathbf{s}}^{int}_{ext}(t)$ as the inhomogeneous term. By the Duhamel principle (Eqs. (20), (61)), we have:

$$\vec{\mathbf{u}}(t)=\exp\left[\widetilde{G}(t-t_{k,0})\right]\vec{\mathbf{u}}(t_{k,0})+\int_{t_{k,0}}^{t}\exp\left[\widetilde{G}(t-\tau)\right]\vec{\mathbf{s}}_{ext}(\vec{\mathbf{u}}(\tau),\tau)\,d\tau \tag{260}$$

$$\vec{\mathbf{u}}^{int}(t)=\exp\left[\widetilde{G}(t-t_{k,0})\right]\vec{\mathbf{u}}(t_{k,0})+\int_{t_{k,0}}^{t}\exp\left[\widetilde{G}(t-\tau)\right]\vec{\mathbf{s}}^{int}_{ext}(\tau)\,d\tau \tag{261}$$

and thus:

$$E^{int}(t)=\left\|\int_{t_{k,0}}^{t}\exp\left[\widetilde{G}(t-\tau)\right]\left[\vec{\mathbf{s}}^{int}_{ext}(\tau)-\vec{\mathbf{s}}_{ext}(\vec{\mathbf{u}}(\tau),\tau)\right]d\tau\right\|=\left\|\int_{t_{k,0}}^{t}\exp\left[\widetilde{G}(t-\tau)\right]\Delta\vec{\mathbf{s}}^{int}_{ext}(\tau)\,d\tau\right\| \tag{262}$$

where we defined:

$$\Delta\vec{\mathbf{s}}_{ext}^{int}(t) \equiv \vec{\mathbf{s}}_{ext}^{int}(t) - \vec{\mathbf{s}}_{ext}(\vec{\mathbf{u}}(t), t). \tag{263}$$

The integral expression can be readily used to yield an upper limit for the error:

$$\begin{aligned} E^{int}(t) &\leq \max_{\tau\in[t_{k,0},t]} \left\| \exp\left[\widetilde{G}(t-\tau)\right] \Delta\vec{\mathbf{s}}_{ext}^{int}(\tau) \right\| (t - t_{k,0}) \\ &\equiv \left\| \exp\left[\widetilde{G}(t-t_{max})\right] \Delta\vec{\mathbf{s}}_{ext}^{int}(t_{max}) \right\| (t - t_{k,0}) \end{aligned} \tag{264}$$

where t_{max} denotes the time-point in the interval $[t_{k,0}, t]$, which maximizes the magnitude of the expression.

Next, we utilize the fact that the time-step is assumed to be small, due to the stability requirements of the algorithm. Hence, we can assume that

$$\left\| \exp\left[\widetilde{G}(t-t_{max})\right] \Delta\vec{\mathbf{s}}_{ext}^{int}(t_{max}) \right\| \approx \left\| \Delta\vec{\mathbf{s}}_{ext}^{int}(t_{max}) \right\| \tag{265}$$

Note that for a Hermitian Hamiltonian, $\widetilde{G}$ becomes anti-Hermitian, and Eq. (265) becomes exact.

At the edge of the time-step, the expression for the absolute time-discretization error yields the following estimation:

$$E^{int}(t_{k,M_k-1}) \approx \left\| \Delta\vec{\mathbf{s}}_{ext}^{int}(t_{max}) \right\| \Delta t_k \tag{266}$$

The relative error can be estimated by

$$E_{rel}^{int}(t_{k,M_k-1}) \approx \frac{\left\| \Delta\vec{\mathbf{s}}_{ext}^{int}(t_{max}) \right\| \Delta t_k}{\left\| \vec{\mathbf{u}}^{int}(t_{k,M_k-1}) \right\|} \tag{267}$$

Observing Eqs. (266), (267), one encounters the problem that t_{max} is unknown. However, the precise knowledge of t_{max} is unnecessary, since we need only an estimation for the order of magnitude of the error. It is reasonable to use the point which is furthest from neighboring sampling points instead of the precise t_{max}. Hence, we can choose the middle point between t_{mid} and the next Chebyshev point.

$\Delta\vec{\mathbf{s}}_{ext}^{int}(t)$ at the estimated t_{max} can be computed directly, via Eq. (263). $\vec{\mathbf{s}}_{ext}(\vec{\mathbf{u}}(t), t)$ at the estimated t_{max} is computed in the same way as the samplings of $\vec{\mathbf{s}}_{ext}(\vec{\mathbf{u}}(t), t)$ at the Chebyshev points (step 2(c)i of the algorithm). $\vec{\mathbf{s}}_{ext}^{int}(t)$ at the estimated t_{max} is computed by the evaluation of the approximation polynomial of $\vec{\mathbf{s}}_{ext}(\vec{\mathbf{u}}(t), t)$ at the point, using the coefficients computed in step 2(c)ii of the algorithm.

This error estimation tends to overestimate the time-discretization error, typically by one or two orders of magnitude at the time-step edge. The reason lies in the oscillatory nature of $\Delta\vec{\mathbf{s}}_{ext}^{int}(t)$, which is responsible for cancellation of errors during the integration in Eq. (262). An accurate estimation of the time-discretization error requires a more detailed analysis. This topic is left for a future publication.

The magnitude of the time-discretization error can be primarily controlled by the number of Chebyshev time-points, M_k. However, M_k cannot be increased indefinitely in order to reduce the error, because of reduction in the efficiency of the algorithm in higher orders (see Sec. 3.4). An alternative option for error reduction is the decrement of Δt_k.

D.1.3. Function of matrix computation error

The estimation of the function of matrix computation error is more direct. It can be readily observed from the computation in step 2(c)vi that the absolute error of the computation of $f_{M_k}(\widetilde{G}, t_{k,l} - t_{k,0})\vec{\mathbf{v}}_{M_k}$ is just the same as the resulting absolute error of $\vec{\mathbf{u}}(t_{k,l})$ itself. Let us denote the absolute error as $E^{fm}(t)$; the relative error at the edge of the time-step can be estimated by

$$E_{rel}^{fm}(t_{k,M_k-1}) = \frac{E^{fm}(t_{k,M_k-1})}{\left\| \vec{\mathbf{u}}^{fm}(t_{k,M_k-1}) \right\|} \tag{268}$$

where $\vec{\mathbf{u}}^{fm}(t)$ denotes the solution resulting from the approximation of $f_{M_k}(\widetilde{G}, t - t_{k,0})\vec{\mathbf{v}}_{M_k}$. An estimation to $E^{fm}(t)$ is given by an estimation of the truncation error of the expansion of $f_{M_k}(\widetilde{G}, t - t_{k,0})\vec{\mathbf{v}}_{M_k}$. In the case that a polynomial expansion approximation is used, one simple way to estimate the truncation error is by computation of the error of the same approximation at several test points; the value of $f_{M_k}(z, \Delta t_k)$ is interpolated in z at several representative test points in the eigenvalue domain, and the relative error from the exact value is computed; the estimated $E^{fm}(t_{k,M_k-1})$ is given by the multiplication of the obtained relative error by $\| f_{M_k}(\widetilde{G}, \Delta t_k)\vec{\mathbf{v}}_{M_k} \|$. An estimation of the error for the Arnoldi approach is given in Appendix B.2.

The magnitude of the function of matrix computation error can be primarily controlled by the number of expansion terms for the approximation, K_k. A decrement of Δt_k also reduces the error.

D.2. Global error

Our main interest is in the estimation of the global error of the final solution obtained by the algorithm. One would have suggest that the global error of the algorithm is just an additive sum of the local errors in each time-step. This is indeed the case when the propagation process is numerically stable. However, it is important to be aware of the fact that the global behavior of the algorithm might be different; we may observe three distinguished behaviors of error accumulation in the algorithm:

1. *Additive accumulation*: The final error is the sum of the errors of the solutions in the last iteration of each time-step;
2. *Divergence with propagation*: The error accumulates in an explosive manner during the propagation. The magnitude of the solution tends to infinity *with the propagation*;
3. *Divergence of the iterative process*: The iterative process in a specific time-step fails to converge. The magnitude of the solution tends to infinity *with the number of iterations*.

In the last two behaviors, the estimation of the local error is useless for the estimation of the global error. The divergence of the iterative process can be always detected, since the algorithm fails to continue the propagation process. Most frequently, the divergence with propagation is also easily detected, by the occurrence of an overflow, or an unreasonably large magnitude of the solution. Seldom, it may occur that the divergent process has stopped at an early stage at the end of the propagation, and the magnitude of the solution is not unreasonable. In this case, it might not be easy to detect the divergent behavior of the error.

Of course, it is highly desirable to prevent an unstable behavior of the algorithm. First, we should attribute the different unstable behaviors to the responsible causes.

The divergence of the iterative process clearly originates in a too large time-step, which is outside the convergence radius of the algorithm. If a divergence of the iterative process occurs, the length of the time-step should be decreased.

The origin of the divergence with propagation is less obvious. Let us recall the classification into the different situations in which the algorithm fails to represent Eq. (65) adequately. When the algorithm represents an equation of the general form of Eq. (56) (situation 1a above), the behavior of the solution is expected to preserve the characteristic features of Eq. (56). For instance, in the homogeneous Schrödinger equation (with a Hermitian Hamiltonian) the norm of the solution is expected to be conserved. Hence, a divergent behavior with the propagation is not expected. In contrary, the behavior of the algorithm in the other situations is unexpected.

In practice, a divergent behavior was observed only in situation 1b above, i.e. when there is a function of matrix computation error. Experience shows that only a low accuracy expansion of $f_{M_k}(\widetilde{G}, t_{k,l} - t_{k,0})\vec{\mathbf{v}}_{M_k}$ leads to a divergent behavior. The instability disappears when more expansion terms are used. We can conclude that it is not recommended to expand $f_{M_k}(z, t_{k,l} - t_{k,0})$ by a low accuracy expansion, even when a low accuracy solution is sufficient. Further research is required to estimate the maximal allowed inaccuracy in the expansion for stability of the propagation. In the problems that were tested so far, the following criterion was found to be sufficient for stability:

$$\frac{E^{fm}(t_{k,M_k-1})}{\left\| f_{M_k}(\widetilde{G}, \Delta t_k)\vec{\mathbf{v}}_{M_k} \right\|} < 10^{-5}$$

It is noteworthy that a divergent behavior with the propagation might be observed also for the convergence error, in a different version of the algorithm than that presented in Sec. 3.2; if one restricts the allowed number of iterations in each time-step (like in the numerical example of Sec. 4), a divergence with propagation might occur. This phenomenon is typical for high order M values. The use of high order M is not recommended anyway, by efficiency considerations (see Sec. 3.4).

It should be noted that the phenomenon of divergence with propagation is common to other propagators, when the time-step is too large.

The inclusion of tests for the magnitude of the error in the algorithm increases the robustness of the algorithm. The tests may be also used for an adaptive choice of the parameters during the propagation. In the future, we plan to develop an improved version of the algorithm, based on this principle.

Appendix E. Supplementary material

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.jcp.2017.04.017.

References

[1] Guy Ashkenazi, Ronnie Kosloff, Sanford Ruhman, Hillel Tal-Ezer, Newtonian propagation methods applied to the photodissociation dynamics of I_3^-, J. Chem. Phys. 103 (23) (1995) 10005–10014.
[2] Philipp Bader, Arieh Iserles, Karolina Kropielnicka, Pranav Singh, Efficient methods for linear Schrödinger equation in the semiclassical regime with time-dependent potential, Proc. R. Soc. A 472 (2016) 20150733.
[3] Roi Baer, Accurate and efficient evolution of nonlinear Schrödinger equations, Phys. Rev. A 62 (2000) 063810.
[4] Weizhu Bao, Dieter Jaksch, Peter A. Markowich, Numerical solution of the Gross–Pitaevskii equation for Bose–Einstein condensation, J. Comput. Phys. 187 (1) (2003) 318–342.

[5] Michael Berman, Ronnie Kosloff, Hillel Tal-Ezer, Solution of the time-dependent Liouville–von Neumann equation: dissipative evolution, J. Phys. A, Math. Gen. 25 (5) (1992) 1283.
[6] Sergio Blanes, Fernando Casas, J.A. Oteo, José Ros, The Magnus expansion and some of its applications, Phys. Rep. 470 (5) (2009) 151–238.
[7] J.C. Butcher, Runge–Kutta methods, in: Numerical Methods for Ordinary Differential Equations, 2010, pp. 143–331.
[8] Marco Caliari, Alexander Ostermann, Implementation of exponential Rosenbrock-type integrators, Appl. Numer. Math. 59 (3–4) (2009) 568–581, Selected Papers from NUMDIFF-11.
[9] Rongqing Chen, Hua Guo, Discrete energy representation and generalized propagation of physical systems, J. Chem. Phys. 108 (15) (1998) 6068–6077.
[10] O. Chuluunbaatar, V.L. Derbov, A. Galtbayar, A.A. Gusev, M.S. Kaschiev, S.I. Vinitsky, T. Zhanlav, Explicit Magnus expansions for solving the time-dependent Schrödinger equation, J. Phys. A, Math. Theor. 41 (29) (2008) 295203.
[11] C. Cohen-Tannoudji, B. Diu, F. Laloë, Quantum Mechanics, Wiley-Interscience, 2005.
[12] M.D. Feit, J.A. Fleck, A. Steiger, Solution of the Schrödinger equation by a spectral method, J. Comput. Phys. 47 (3) (1982) 412–433.
[13] Richard A. Friesner, Laurette S. Tuckerman, Bright C. Dornblaser, Thomas V. Russo, A method for exponential propagation of large systems of stiff nonlinear differential equations, J. Sci. Comput. 4 (4) (1989) 327–354.
[14] E.K.U. Gross, W. Kohn, Time-dependent density functional theory, Adv. Quantum Chem. 21 (1990) 255–291.
[15] Hua Guo, Recursive solutions to large eigenproblems in molecular spectroscopy and reaction dynamics, Rev. Comput. Chem. 25 (2007) 285–347.
[16] Hua Guo, Rongqing Chen, Short-time Chebyshev propagator for the Liouville–von Neumann equation, J. Chem. Phys. 110 (14) (1999) 6626–6634.
[17] M. Hochbruck, A. Ostermann, J. Schweitzer, Exponential integrators of Rosenbrock-type, Oberwolfach Rep. 3 (2006) 1107–1110.
[18] Marlis Hochbruck, Christian Lubich, Exponential integrators for quantum-classical molecular dynamics, BIT Numer. Math. 39 (4) (1999) 620–645.
[19] Marlis Hochbruck, Alexander Ostermann, Explicit exponential Runge–Kutta methods for semilinear parabolic problems, SIAM J. Numer. Anal. 43 (3) (2005) 1069–1090.
[20] Youhong Huang, Donald J. Kouri, David K. Hoffman, General, energy-separable Faber polynomial representation of operator functions: theory and application in quantum scattering, J. Chem. Phys. 101 (12) (1994) 10493–10506.
[21] Wilhelm Huisinga, Lorenzo Pesce, Ronnie Kosloff, Peter Saalfrank, Faber and Newton polynomial integrators for open-system density matrix propagation, J. Chem. Phys. 110 (12) (1999) 5538–5547.
[22] Christiane P. Koch, Mamadou Ndong, Ronnie Kosloff, Two-photon coherent control of femtosecond photoassociation, Faraday Discuss. 142 (2009) 389–402.
[23] D. Kosloff, R. Kosloff, A Fourier method solution for the time dependent Schrödinger equation as a tool in molecular dynamics, J. Comput. Phys. 52 (1) (1983) 35–53.
[24] Ronnie Kosloff, Propagation methods for quantum molecular dynamics, Annu. Rev. Phys. Chem. 45 (1) (1994) 145–178.
[25] Jeffrey L. Krause, Kenneth J. Schafer, Kenneth C. Kulander, Calculation of photoemission from atoms subject to intense laser fields, Phys. Rev. A 45 (1992) 4998–5010.
[26] Kenneth C. Kulander, Time-dependent Hartree–Fock theory of multiphoton ionization: helium, Phys. Rev. A 36 (6) (1987) 2726.
[27] Claude Leforestier, R.H. Bisseling, Charly Cerjan, M.D. Feit, Rich Friesner, A. Guldberg, A. Hammerich, G. Jolicard, W. Karrlein, H.-D. Meyer, et al., A comparison of different propagation schemes for the time dependent Schrödinger equation, J. Comput. Phys. 94 (1) (1991) 59–80.
[28] Wilhelm Magnus, On the exponential solution of differential equations for a linear operator, Commun. Pure Appl. Math. 7 (4) (1954) 649–673.
[29] H-D. Meyer, Uwe Manthe, Lorenz S. Cederbaum, The multi-configurational time-dependent Hartree approach, Chem. Phys. Lett. 165 (1) (1990) 73–78.
[30] Borislav V. Minchev, Will M. Wright, A review of exponential integrators for first order semi-linear problems, 2005.
[31] J.G. Muga, J.P. Palao, B. Navarro, I.L. Egusquiza, Complex absorbing potentials, Phys. Rep. 395 (6) (2004) 357–426.
[32] Mamadou Ndong, Hillel Tal-Ezer, Ronnie Kosloff, Christiane P. Koch, A Chebychev propagator for inhomogeneous Schrödinger equations, J. Chem. Phys. 130 (12) (2009) 124108.
[33] Mamadou Ndong, Hillel Tal-Ezer, Ronnie Kosloff, Christiane P. Koch, A Chebychev propagator with iterative time ordering for explicitly time-dependent Hamiltonians, J. Chem. Phys. 132 (6) (2010) 064105.
[34] Daniel Neuhauser, Michael Baer, The application of wave packets to reactive atom–diatom systems: a new approach, J. Chem. Phys. 91 (8) (1989) 4651–4657.
[35] José P. Palao, Ronnie Kosloff, Christiane P. Koch, Protecting coherence in optimal control theory: state-dependent constraint approach, Phys. Rev. A 77 (6) (2008) 063412.
[36] J.P. Palao, J.G. Muga, A simple construction procedure of absorbing potentials, Chem. Phys. Lett. 292 (1–2) (1998) 1–6.
[37] Tae Jun Park, J.C. Light, Unitary quantum time evolution by iterative Lanczos reduction, J. Chem. Phys. 85 (10) (1986) 5870–5876.
[38] Uri Peskin, Ronnie Kosloff, Nimrod Moiseyev, The solution of the time dependent Schrödinger equation by the (t, t') method: the use of global polynomial propagators for time dependent hamiltonians, J. Chem. Phys. 100 (12) (1994) 8849–8855.
[39] Lothar Reichel, Newton interpolation at Leja points, BIT 30 (2) (1990) 332–346.
[40] J.M. Sanz-Serna, Methods for the numerical solution of the nonlinear Schrödinger equation, Math. Comput. 43 (167) (1984) 21–27.
[41] Ido Schaefer, Ronnie Kosloff, Optimal-control theory of harmonic generation, Phys. Rev. A 86 (2012) 063417.
[42] I. Serban, J. Werschnik, E.K.U. Gross, Optimal control of time-dependent targets, Phys. Rev. A 71 (5) (2005) 053810.
[43] A.Y. Suhov, An accurate polynomial approximation of exponential integrators, J. Sci. Comput. 60 (3) (2014) 684–698.
[44] Zhigang Sun, Weitao Yang, Dong H. Zhang, Higher-order split operator schemes for solving the Schrödinger equation in the time-dependent wave packet method: applications to triatomic reactive scattering calculations, Phys. Chem. Chem. Phys. 14 (6) (2012) 1827–1845.
[45] Hillel Tal-Ezer, Polynomial approximation of functions of matrices and applications, J. Sci. Comput. 4 (1) (1989) 25–60.
[46] Hillel Tal-Ezer, High degree polynomial interpolation in Newton form, SIAM J. Sci. Stat. Comput. 12 (3) (1991) 648–667.
[47] Hillel Tal-Ezer, On restart and error estimation for Krylov approximation of $w = f(A)v$, SIAM J. Sci. Comput. 29 (6) (2007) 2426–2441.
[48] Hillel Tal-Ezer, R. Kosloff, An accurate and efficient scheme for propagating the time dependent Schrödinger equation, J. Chem. Phys. 81 (9) (1984) 3967–3971.
[49] Hillel Tal-Ezer, Ronnie Kosloff, Charles Cerjan, Low-order polynomial approximation of propagators for the time-dependent Schrödinger equation, J. Comput. Phys. 100 (1) (1992) 179–187.
[50] Hillel Tal-Ezer, Ronnie Kosloff, Ido Schaefer, New, highly accurate propagator for the linear and nonlinear Schrödinger equation, J. Sci. Comput. 53 (1) (2012) 211–221.
[51] Amrendra Vijay, Horia Metiu, A polynomial expansion of the quantum propagator, the Green's function, and the spectral density operator, J. Chem. Phys. 116 (1) (2002) 60–68.
[52] J. Werschnik, E.K.U. Gross, Quantum optimal control theory, J. Phys. B, At. Mol. Opt. Phys. 40 (18) (2007) R175.
[53] P. Zhao, H. De Raedt, S. Miyashita, F. Jin, K. Michielsen, Dynamics of open quantum spin systems: an assessment of the quantum master equation approach, Phys. Rev. E 94 (2016) 022126.
[54] Wei Zhu, Youhong Huang, D.J. Kouri, Colston Chandler, David K. Hoffman, Orthogonal polynomial expansion of the spectral density operator and the calculation of bound state energies and eigenfunctions, Chem. Phys. Lett. 217 (1–2) (1994) 73–79.

Chapter 4

Optimization of high harmonic generation by optimal control theory—climbing a mountain in extreme conditions

Unpublished.

In the present chapter, the optimization scheme is applied to typical HHG problems. The theory established in Chapter 2 is further developed and adjusted to the HHG problem. The simulation of the dynamics is performed by the tools developed in Chapter 3 for non-Hermitian dynamics. The absorbing boundary conditions are realized by the method employed in Chapter 3. The optimization is performed by a second-order gradient method (BFGS), replacing the relaxation scheme of Chapter 2. This was required due to the relative complexity of the HHG physics, which is reflected in the complexity of the optimization hypersurface. The application of the optimization method to the HHG problem is thoroughly discussed, with attention to several unique issues.

Optimization of high harmonic generation by optimal control theory—climbing a mountain in extreme conditions

Ido Schaefer and Ronnie Kosloff[1,*]

[1] *The Institute of Chemistry, The Hebrew University of Jerusalem, Jerusalem 9190401, Israel*

(Dated: November 10, 2019)

A theoretical optimization method of high-harmonic-generation (HHG) is developed in the framework of optimal-control-theory (OCT). The target of optimization is the emission radiation of a particular frequency. The OCT formulation includes restrictions on the frequency band of the driving pulse, the permanent ionization probability and the total energy of the driving pulse. The optimization task requires a highly accurate simulation of the dynamics. Absorbing boundary conditions are employed, where a complex-absorbing-potential is constructed by an optimization scheme for maximization of the absorption. A new highly accurate propagation scheme is employed, which can address explicit time dependence of the driving as well as a non-Hermitian Hamiltonian. The optimization process is performed by a second-order gradient scheme. The method is applied to a simple one-dimensional model system. The results demonstrate a significant enhancement of selected harmonics, with minimized total energy of the driving pulse and controlled permanent ionization probability. A successful enhancement of an even harmonic emission is also demonstrated.

* ido.schaefer@mail.huji.ac.il; kosloff1948@gmail.com

I. INTRODUCTION

When high intensity light is focused on a dilute gas novel phenomena emerge. One of the most intriguing is High-Harmonic-Generation (HHG), where a nonlinear response of the atomic system leads to emission in very high multiples of carrier frequency of the driving laser [1–3]. HHG is a crucial step in generating attosecond pulses [4–7]. In addition, HHG has become a light source in extreme UV or soft X-ray for novel spectroscopic applications [8–11].

HHG constitutes a major breakthrough in experimental physics. However, the main drawback is the low efficiency of the process (typically $\sim 10^{-5}$). Another major drawback is the energetic requirements, where very high power lasers are required to initiate the process (typically in the range of $\sim 10^{14} - 10^{15} W/cm^2$). In addition, the HHG process involves ionization. The permanent liberation of the electronic density into the macroscopic medium results in the production of plasma, which is responsible to severe experimental problems, such as dispersion and absorption of the emitted radiation.

The low yield of the HHG process has motivated novel approaches to enhance the process. One major direction is the application of *quantum coherent control* for the optimization of the HHG process. In the present study, we establish and explore a quantum Optimal Control Theory (OCT) scheme to enhance the emission of a specific harmonic output. The minimization of the energetic requirements and restriction of permanent ionization are also addressed.

In general, the optimization of a physical process can be achieved by two approaches:

1. *Rational optimization*, based on the intuitive application of physical insights;

2. *Search procedure optimization*, based on a computational optimization procedure.

Unravelling the mechanism of the process is a step toward a rational optimization. The three-step-mechanism was the first successful HHG model [12]. This is a single electron semiclassical model: At the peak of the radiation pulse, the bound electron tunnels out to free space. In turn, the field switches sign and the electron is launched back to the nucleus where it recombines and emits a high energy photon [13, 14].

This simple semiclassical theory has been successful in predicting control approaches able to enhance the yield. For example, control of the absolute phase of a few cycle pulse enhances the HHG yield. The model has inspired the use of polarization and polarization shaping to control the HHG process [15–17].

An important aspect ignored by the semiclassical three-step model is the importance of quantum effects in HHG, such as interference of the electronic wavefunction, or interference of pathways [18].

Other models for HHG have been proposed which were applicable to earlier experiments which used higher driving frequency [1, 19–21].

Contrary to rational optimization, the search procedure optimization approach does not require previous physical insights. Machine learning enables optimization of a process without a physical model [22]. HHG has also been optimized using learning algorithms based on simulations [23]. Experimental approaches to the optimization of HHG were based on open loop learning algorithms [24–27]. This is a successful pragmatic approach, however with the disadvantage that physical insight can only be gained after extensive analysis.

Quantum optimal control theory (OCT) was developed as a computational scheme able to obtain a control field for a given task [28, 29]. The control task is formulated as a *maximization problem* of a functional object, based on variational calculus. The first control tasks were the control of a chemical yield [29]. Other tasks emerged [30], such as cooling molecular internal degrees of freedom [31] or generating quantum gates [32]. A major effort has been devoted to create efficient algorithms to solve OCT problems [33–38]. These methods are based on iterative algorithms. The largest computational effort in the computational optimization process is devoted to solving the time-dependent Schrödinger equation with a time-dependent control field. The methods differ by their update scheme of the control field from iteration to iteration. The search process of OCT is significantly more efficient than learning algorithm approaches, since it is based explicitly or implicitly on *gradient information*, which is lacking in learning algorithms.

(*Terminological remark*: The present paper addresses the optimization of the HHG phenomenon; however, many of the discussed details apply also to the more general problem of optimization of harmonic generation (HG), including low-order harmonic generation phenomena. Other details are unique to the subproblem of optimization of HHG. The term HG refers to the general problem, while the term HHG refers specifically to the particular problem addressed in the paper.)

Control of HG poses a severe challenge on OCT. One difficulty lies in the fact that OCT is naturally formulated in the time domain, while the control requirements are defined in the frequency domain. An additional difficulty originates in the fact that frequency requirements extend over a duration of time. This results in a uniqueness of the OCT formulation: While the common OCT formulation addresses targets at a given time-point, the target of HG is extended in time. In OCT such targets require an inhomogeneous source term in the Schrödinger equation [39–41]. The treatment of inhomogeneity poses a numerical challenge.

The earliest study of an emission spectrum target dates back to 2008 [42]. However, this study is

confined to the context of control of coherent anti-Stokes Raman scattering (CARS) spectrum.

Our previous work [43, 44] first addressed the general problem of theoretical optimization of HG. It mainly focused on the optimization of low-order HG phenomena, such as HG based on a resonance-mediated-absorption mechanism in molecular model systems. The optimization method was demonstrated to efficiently enhance the emission in selected frequencies. In [44] the method was demonstrated also to a HHG problem. However, the chosen optimization problem was not a typical HHG process. The selected target frequency was within the Bohr frequencies of the system, and thus the optimized physical process was based on the excitation of one of the *characteristic frequencies of the system*. In contrast, in the typical HHG process the spectrum consists of multiples of the fundamental frequency of the *laser source*, where the system plays the role of an up-conversion medium for the incoming frequency. Thus, the physical process is fundamentally different.

Another problem in [43, 44] was the chosen restriction imposed on the source field spectrum. Only an upper bound on the control frequency was imposed. As a result, the control field included low frequencies which currently cannot be produced in high intensities. A more realistic restriction is a frequency band centered around the carrier frequency of the laser source (see Sec. IV).

The present study addresses the particular problem of HHG control, based on the principles developed in our previous work for the general HG problem. However, the HHG problem presents several unique challenges which require a special treatment.

The accurate numerical simulation of the HHG dynamics presents a computational challenge. The HHG process is characterized by extreme physical conditions: The central atomic or molecular potential has a Coulomb character, which is steep by nature; the process involves ionization, where the dynamics of liberated electron in the continuum extends to large spatial distances from the parent ion; the dynamics of the liberated electron under the influence of the driving field is characterized by a strongly accelerated motion. These extremes require large computational resources for a reliable and accurate description.

Another aspect of the numerical difficulty lies in the fact that HHG is a small amplitude effect. Hence, even tiny numerical artefacts lead to large distortions of the high-harmonic spectrum, where the magnitude of the spurious effect may be comparable to the physical effects. This necessitates the use of highly accurate tools for the description of the HHG process. In particular, a highly accurate solution of the time-dependent Schrödinger equation is required, which necessitates an appropriate solver. Another major challenge lies in the realization of absorbing boundary conditions. Imperfection in their absorption capabilities will lead to large spurious effects [45, 46].

The *optimization* procedure involves additional challenges. The complexity of the physical situation

is reflected in the optimization hypersurface, which becomes considerably more complex than that of the simpler low-order HG problems. This requires the use of more sophisticated optimization tools. Another problem is that the severity of the numerical artefacts can be enhanced by the optimization process, which tends to maximize them at the expense of real physical effects.

An issue which requires special treatment in the optimization of HHG is the requirement of prevention of permanent ionization. While ionization is an integral part of the typical HHG mechanism, the magnitude of the electronic probability liberated into the macroscopic medium should be minimized. This requirement has to be reflected in the OCT formulation.

Several studies dealing with theoretical optimization of HHG have been recently published [47–51]. Ref. [47] employs a maximization term similar to the one employed in our previous work, as well as in [42]. The study aims at the extension of the cutoff frequency. However, as noted by the authors, the available band of the source includes low frequencies (as in [44]), which increase the ponderomotive energy and thus extend the cutoff. The low frequencies dominate the spectra of the optimized pulses. Thus, the control achievement presented in this study is actually the maximization of the emission in a region which is below the cutoff frequency of the dominating frequency in the pulse. A demonstration of the extension of the cutoff without the extension of the available band to lower frequencies has not been achieved to date.

Ref. [48] utilizes the principles presented in [43, 44] to the development of an OCT formulation in the framework of time-dependent Density-Functional-Theory (TDDFT). This theory enables an optimization of HHG in multi-electron systems. Unlike in [43, 44, 47], the system is controlled by the variation of the profile of a *slowly varying envelope* of a fixed carrier frequency.

Refs. [49–51] employ genetic algorithm optimization schemes, an approach which has already been employed in both theoretical and experimental works, as was mentioned above.

The contribution of the present study lies in the thorough treatment of the numerical aspects of the problem, as well as in the theoretical treatment of the restriction of permanent ionization, a topic which has not been addressed in other studies.

The aim of this paper is to establish a comprehensive approach based on OCT to address the target of optimizing a single emission line in HHG. The ultimate goal is to identify novel mechanisms based on interference phenomena which can achieve this goal. Before this task can be achieved, the technical issues in adopting OCT to HHG have to be addressed and solved. This is the main topic of this paper. The discussion of the mechanisms of optimized fields remains to be addressed in a future publication.

The paper is organized as follows: In Sec. II, we develop the theoretical OCT formulation; in Sec. III,

the numerical tools are described; Sec. IV demonstrates the method in the maximization of selected harmonics in a simple model system; the paper is concluded in Sec. V.

II. THEORY

Our basic OCT formulation of the general HG problem is reviewed in Sec. II A. A more detailed description can be found in [43, 44]. This formulation establishes the foundations of the method. In Secs. II C, II D, the basic formulation is augmented by the required elements for the HHG problem.

A. The basic OCT formulation of harmonic generation

We consider the optimization of a laser pulse defined in the time-interval $t \in [0, T]$. We assume a linear polarization in the x direction. The temporal profile of the pulse is defined by the time-dependent electric field $\epsilon(t)$, where the effect of the magnetic field is ignored.

The basic requirements of the HG optimization consist of two elements:

1. The spectrum of the driving field has to be restricted to the spectral band available by the laser source.

2. The emission of the system has to be maximized at the target frequency band.

The two requirements can be treated separately.

Our control requirements are *spectral*, which are naturally expressed in the *frequency domain*. However, the *quantum dynamics* is naturally expressed in the *time domain*. These two descriptions are related by a *spectral transform*. For this study we employed the *cosine transform*. It was preferred over other spectral transforms (the Fourier transform and the sine transform) due to the properties of its discrete version, the *discrete cosine transform* (DCT) (see Appendix C 1), as will be clarified in Sec. II D and Appendix A.

The operation of the cosine transform on an arbitrary function $g(t)$ will be denoted by the symbol $\mathcal{C}$, and the transformed function by $\bar{g}(\omega)$:

$$\bar{g}(\omega) \equiv \mathcal{C}[g(t)] \equiv \sqrt{\frac{2}{\pi}} \int_0^\infty g(t) \cos(\omega t)\, dt \tag{1}$$

The inverse cosine transform will be denoted by $\mathcal{C}^{-1}$:

$$\mathcal{C}^{-1}[\bar{g}(\omega)] \equiv \sqrt{\frac{2}{\pi}} \int_0^{\infty} \bar{g}(\omega) \cos(\omega t)\, d\omega = g(t) \tag{2}$$

The cosine-transform is its own inverse. It has the important property of being an *orthogonal transformation*.

In practice, the upper integration limit of the transform does not extend to infinity. The problem is discretized in both time and frequency; this restricts the upper limits of integration of both the direct and the inverse transforms, as follows. The upper integration limit of the inverse transform is always limited by the sampling frequency of the temporal signal, which determines the maximal available frequency, denoted here as Ω. Similarly, the length of the represented time-interval is determined by the density of the discretized ω grid. The theory is defined within the time-interval $t \in [0, T]$; hence, the density of the ω grid is naturally chosen to represent a time-interval of length T. Thus, the direct transform is replaced by a *finite-time transform*, with an upper limit T.

We begin from the treatment of the maximization of the emission in the target band, which is the second control requirement mentioned above. The emission spectrum is generated by the dipole dynamics. This, in turn, is determined by the observables related to the dipole dynamics, e. g., the dipole operator, the momentum operator, or the dipole acceleration operator (see Appendix A). The emission spectrum consists of the same frequencies contained in the spectra of the time-dependent expectation values of these observables. Thus, the emission spectrum can be controlled by controlling the spectrum of one of these observables. In what follows, we shall address the general problem of the control of the spectrum of an arbitrary Hermitian observable. The HG problem becomes a particular case of the general problem.

Let us denote the controlled observable by $\hat{\mathbf{O}}$. The $\left\langle \hat{\mathbf{O}} \right\rangle(t)$ spectrum becomes

$$\overline{\left\langle \hat{\mathbf{O}} \right\rangle}(\omega) \equiv \mathcal{C}\left[\left\langle \hat{\mathbf{O}} \right\rangle(t)\right] = \sqrt{\frac{2}{\pi}} \int_0^{T} \left\langle \hat{\mathbf{O}} \right\rangle(t) \cos(\omega t)\, dt \tag{3}$$

The problem of maximization of the magnitude of $\overline{\left\langle \hat{\mathbf{O}} \right\rangle}(\omega)$ in the target frequency band can be translated into the maximization of the following functional term:

$$J_{max} \equiv \frac{1}{2} \int_0^{\Omega} f_O(\omega) \overline{\left\langle \hat{\mathbf{O}} \right\rangle}^2(\omega)\, d\omega, \tag{4}$$

$$f_O(\omega) \geq 0, \qquad \max\left[f_O(\omega)\right] = 1 \tag{5}$$

where $f_O(\omega)$ is a *filter function* which represents the target frequency band, i. e. it has pronounced values only in the target spectral region (note that the normalization chosen here is different from [43, 44]). A target maximization term of this type has been employed previously in [42].

We proceed to the treatment of the restriction of the driving field spectrum, the first requirement mentioned above. The function $\bar{\epsilon}(\omega)$ defines the spectral representation of the time-dependent driving field $\epsilon(t)$ as follows:

$$\epsilon(t) = \mathcal{C}^{-1}[\bar{\epsilon}(\omega)] = \sqrt{\frac{2}{\pi}} \int_0^\Omega \bar{\epsilon}(\omega) \cos(\omega t)\, d\omega, \qquad t \in [0, T] \tag{6}$$

The spectral restriction of the driving field is achieved by a frequency-dependent *penalty term* inserted into the maximization functional:

$$J_{energy} \equiv -\alpha \int_0^\Omega \frac{1}{f_\epsilon(\omega)} \bar{\epsilon}^2(\omega)\, d\omega, \tag{7}$$

$$f_\epsilon(\omega) > 0, \qquad \max\left[f_\epsilon(\omega)\right] = 1 \tag{8}$$

$$\alpha > 0 \tag{9}$$

where $f_\epsilon(\omega)$ is another *filter function*, which represents the available laser source band, i. e. it has pronounced values in the frequency band available to the control field (note that the normalization chosen here is different from [43, 44]). α is an adjustable penalty factor. J_{energy} penalizes the undesirable frequency regions, and thus restricts the band of the optimized $\bar{\epsilon}(\omega)$.

J_{energy} has also the role of imposing a restriction on the *total energy* of the pulse. By (8) we have:

$$\int_0^\Omega \frac{1}{f_\epsilon(\omega)} \bar{\epsilon}^2(\omega)\, d\omega \geq \int_0^\Omega \bar{\epsilon}^2(\omega)\, d\omega \tag{10}$$

The RHS represents the *norm* of $\bar{\epsilon}(\omega)$. When $\bar{\epsilon}(\omega)$ is transformed to the time-domain, its norm is preserved by the orthogonality property of the inverse cosine-transform. Thus, the magnitude of the RHS becomes equivalent to the *fluence* of the driving field,

$$\Phi\left[\epsilon(t)\right] \equiv \int_0^T \epsilon^2(t)\, dt \tag{11}$$

Consequently, the fluence becomes the lower limit of the magnitude of the LHS of (10). This means that the magnitude of the fluence is also penalized by J_{energy}. Since the fluence is proportional to the total energy of the pulse, J_{energy} also restricts the total energy.

The energetic restriction is particularly important in the context of HHG, which is typically a highly inefficient process energetically. The simultaneous maximization of the emission amplitude by J_{max} and minimization of the driving field amplitude by J_{energy} enhances the efficiency of the process.

It is convenient to define (Cf. Eq. (7)):

$$\tilde{f}_\epsilon(\omega) \equiv \frac{f_\epsilon(\omega)}{\alpha} \tag{12}$$

Penalization of the undesirable frequency components of the field has been suggested previously in [36]. However, Ref. [36] employs a time-domain formulation, while the present formulation is in the frequency domain, which has considerable advantages (see [43, Chapter 3], where the relation to other methods of restricting the driving field spectrum [36, 40, 52] was discussed).

The *dynamical requirements* are imposed by a *constraint* to the optimization problem. The constraint is introduced into the optimization formulation by the *Lagrange-multiplier method*.

The dynamics is governed by the time-dependent Schrödinger equation, subject to a given initial condition:

$$\frac{d\left|\psi(t)\right\rangle}{dt} = -i\hat{\mathbf{H}}(t)\left|\psi(t)\right\rangle, \qquad \left|\psi(0)\right\rangle = \left|\psi_0\right\rangle \tag{13}$$

(Atomic units are used throughout, thus we set: $\hbar = 1$.) The time-dependent Hamiltonian is given by

$$\hat{\mathbf{H}}(t) = \hat{\mathbf{H}}_0 - \hat{\mu}_x\epsilon(t) \tag{14}$$

where $\hat{\mathbf{H}}_0$ is the unperturbed Hamiltonian, and $-\hat{\mu}_x\epsilon(t)$ represents the interaction of the x component of the dipole with the x polarized driving field. The dipole approximation is employed. The Schrödinger equation becomes a time-dependent constraint to the optimization problem. The maximization functional is modified by the introduction of a corresponding Lagrange-multiplier term (see Appendix B for more details):

$$J_{Schr} = -2\,\mathrm{Re}\int_0^T \left\langle \chi(t)\left|\frac{d}{dt} + i\hat{\mathbf{H}}(t)\right|\psi(t)\right\rangle dt \tag{15}$$

The full maximization functional of the basic HG problem becomes

$$J[\bar{\epsilon}(\omega)] = J_{max} + J_{energy} + J_{Schr} \tag{16}$$

After imposing the extremum conditions, we obtain the following set of Euler-Lagrange equations

(see Appendix B for the full derivation):

$$\frac{d\left|\psi(t)\right\rangle}{dt} = -i\hat{\mathbf{H}}(t)\left|\psi(t)\right\rangle, \qquad \left|\psi(0)\right\rangle = \left|\psi_0\right\rangle \tag{17}$$

$$\frac{d\left|\chi(t)\right\rangle}{dt} = -i\hat{\mathbf{H}}(t)\left|\chi(t)\right\rangle - \mathcal{C}^{-1}\left[f_O(\omega)\overline{\left\langle\hat{\mathbf{O}}\right\rangle}(\omega)\right]\hat{\mathbf{O}}\left|\psi(t)\right\rangle, \qquad \left|\chi(T)\right\rangle = 0 \tag{18}$$

$$\bar{\epsilon}(\omega) = f_\epsilon(\omega)\mathcal{C}[\eta(t)], \qquad \eta(t) \equiv -\frac{\mathrm{Im}\left\langle\chi(t)\left|\hat{\mu}\right|\psi(t)\right\rangle}{\alpha} \tag{19}$$

$$\epsilon(t) = \mathcal{C}^{-1}[\bar{\epsilon}(\omega)] = \mathcal{C}^{-1}\left\{f_\epsilon(\omega)\mathcal{C}[\eta(t)]\right\} \tag{20}$$

The expression for $\eta(t)$ is identical to that obtained for the driving field $\epsilon(t)$ without the spectral restriction (see [40, 43, 44]). Thus, Eq. (20) can be interpreted as the unrestricted field, filtered by the filter function $f_\epsilon(\omega)$. A complete filtration of undesirable spectral regions is achieved in the limit $f_\epsilon(\omega) \longrightarrow 0$. Although J_{energy} is rigorously undefined for $f_\epsilon(\omega) = 0$, in practice $f_\epsilon(\omega)$ can be set to 0 achieving a complete filtration of undesirable regions.

The advantage of the current formulation for the spectral restriction of the driving field lies in the flexibility of Eq. (19), where the freedom in the choice of $f_\epsilon(\omega)$ can have a prominent effect on the profile of the optimized pulse. Smooth filtration can be achieved by choosing a smooth filter function. In addition, $f_\epsilon(\omega)$ can be chosen so as to provide an envelope shape to the spectral profile of the field (limitations of this practice are discussed in Sec. IV). Finally, in certain experiments the system is irradiated by several sources, which differ both in the spectral band and the available intensity; an appropriate choice of $f_\epsilon(\omega)$ can address these scenarios.

General comments on the presented OCT formulation are in order.

The form of J_{max} enables a *selective enhancement* of spectral emission regions. However, it should the noted that there is an ambiguity in the term "selectivity", which can be interpreted in two different ways:

1. Maximization of the yield of the selected target, regardless of the appearance of "by-products" (in our case, the emission in other spectral regions);

2. Simultaneous maximization of the yield of the selected target, and minimization of the yield of any "by-product" (in our case, suppression of other harmonics).

The formulation presented here is aimed at selectivity of the first type. However, several other studies target the selectivity of the second type, and the control requirements include suppression of undesirable harmonics (see, for example, [26]). In principle, the suppression of harmonics can be achieved by

modifying the definition of $f_O(\omega)$ (Eq. (5)) to allow negative values. Then the J_{max} term can be used to penalize undesirable spectral emission regions. Note that this changes the interpretation of $f_O(\omega)$ as a filter function. However, this option has not been investigated yet.

Selectivity was the aim of the earliest quantum coherent control studies, which addressed the selective enhancement of the yield in chemical reactions. However, it should be noted that there is a fundamental difference between the optimization of chemical reactions and optimization of HHG. In chemical reactions, the sum of the yields of all products is always 100%. Consequently, the by products are always produced at the expense of the yield of the desired chemical product. Thus, there is no ambiguity in the term of selectivity in the chemical context. In contrary, in the present context, the high-harmonic products never sum to 100% of the energetic yield. The efficiency of the HHG process is very low—typically, the total energy of the emission is orders of magnitude lower than the invested energy of the incoming pulse. Therefore, the appearance of high-harmonic "by products" (typically, neighbouring harmonics; see Sec. IV) is unnecessarily at the expense of the desired harmonics. On the contrary, the suppression of other harmonics inserts an additional requirement into the control problem, which may be at the expense of the maximization target.

Both J_{energy} and J_{Schr} set constraints on the optimization problem. However, the two terms are fundamentally different in nature. The Schrödinger equation constraint is a "hard" constraint, which is well defined, and should strictly not be violated. In contrast, J_{energy} represents softer requirements of desirable trends—the energy should be kept as small as possible, and the spectrum of the pulse should gradually decay in undesirable regions. The requirements represented by a penalty term are sometimes referred to as "soft constraints". In what follows, the additional requirements of HHG will be realized by both hard and soft constraint terms.

B. The target operator

The target operator $\hat{\mathbf{O}}$ in the present study is chosen as the *stationary acceleration operator* of the dipole, which will be defined in what follows. The *dipole acceleration operator* is defined as

$$\hat{\ddot{\mu}}_x = \hat{\ddot{\mathbf{X}}} = -\frac{dV\left(\hat{\mathbf{X}}\right)}{d\hat{\mathbf{X}}} \tag{21}$$

where $V(x)$ is the potential. Note that atomic units are used, and the electron has a unit mass and charge. The expression in the RHS of Eq. (21) represents the operator expression of the classical force.

The potential can be divided into a stationary part and a time-dependent part:

$$V\left(\hat{\mathbf{X}}\right) = V_0\left(\hat{\mathbf{X}}\right) - \hat{\mathbf{X}}\epsilon(t) \tag{22}$$

Accordingly, the acceleration operator can also be divided into a stationary part and a time-dependent part:

$$\ddot{\hat{\mathbf{X}}} = -\frac{dV_0\left(\hat{\mathbf{X}}\right)}{d\hat{\mathbf{X}}} + \epsilon(t) \tag{23}$$

The *stationary acceleration operator* is defined by the stationary part of the acceleration operator:

$$\hat{\mathbf{C}} \equiv -\frac{dV_0\left(\hat{\mathbf{X}}\right)}{d\hat{\mathbf{X}}} \tag{24}$$

The time-dependent part, $\epsilon(t)$, contributes only large linear response components to the emission spectrum (as will be explained in Appendix A), and thus is not of interest in the present context [48]. Moreover, it is advantageous to eliminate these components from the spectrum due to numerical considerations (see Appendix A). Accordingly, we set:

$$\hat{\mathbf{O}} \equiv \hat{\mathbf{C}} \tag{25}$$

A thorough discussion of the considerations in the choice of the target operator is given in Appendix A. With the choice (25) of the target operator, the maximization term J_{max} becomes (Cf. Eq. (4)):

$$J_{max} \equiv \frac{1}{2}\int_0^{\Omega} f_C(\omega)\overline{\left\langle \hat{\mathbf{C}} \right\rangle}^2(\omega)\, d\omega, \qquad f_C(\omega) \geq 0 \tag{26}$$

where $f_C(\omega)$ is the target filter function for the $\left\langle \hat{\mathbf{C}} \right\rangle(t)$ spectrum.

C. Restriction of permanent ionization

The HHG process involves partial ionization of the system. Part of the amplitude of the liberated electron reverts to the parent ion, and participates in the generation of high harmonics via the overlap with the ionic core. However, part of the electronic amplitude does not return, and the system becomes *permanently ionized*. This results in the production of *plasma* in the medium, which is a source of experimental problems (see, for example, [53]). Therefore, it is highly desirable to reduce the permanent

ionization in the process. This task can be achieved by incorporating an additional control requirement in the OCT formalism.

In our previous work [44] the restriction of permanent ionization was realized by imposing a penalty on spatial regions which are beyond a chosen spatial threshold. The problem with this formulation is that it is suitable for a complete elimination of permanent ionization. However, we found that this requirement is incompatible for typical HHG problems; some permanent ionization has to be allowed to enable the appearance of significant HHG effects. In principle, the magnitude of the penalty can be reduced to allow some permanent ionization. The difficulty is that it becomes very intricate to quantify the allowed permanent ionization probability in this method.

In the present work we used a different formulation. The penalty is imposed on the permanent ionization itself. We assume that *absorbing boundaries* are employed to eliminate the outgoing amplitude at the edges of the grid. The penalty on permanent ionization can be formulated by the following penalty term:

$$J_{ion} \equiv \sigma\left(\langle\psi(T)|\psi(T)\rangle\right), \tag{27}$$

$$\sigma(y) \leq 0, \tag{28}$$

$$\sigma(1) = 0 \tag{29}$$

where $\sigma(y)$ is a *monotonically increasing function*, which has the role of a penalty function. When absorbing boundaries are employed, the effective Hamiltonian becomes non-Hermitian. As a result, the norm of $|\psi(t)\rangle$ is not conserved. The magnitude of $\langle\psi(t)|\psi(t)\rangle$ is gradually decreasing during the process, as the electron's probability is being absorbed by the boundaries. $\langle\psi(T)|\psi(T)\rangle$ represents the survival probability in the process, which can be identified as the probability which is not permanently ionized (see Appendix E). Condition (28) implies that the lost electronic density is penalized. The requirement that $\sigma(y)$ is a monotonically increasing function implies that the magnitude of the penalty increases with the permanent ionization probability. Condition (29) ensures that J remains unaltered in the limit of zero permanent ionization probability.

The main advantage of this formulation lies in the flexibility of the form of $\sigma(y)$, which can be adjusted to represent the complexity of the control requirement. As has already been mentioned, some permanent ionization should be allowed; this can be reflected by the definition of the maximal allowed permanent ionization probability, or, equivalently, the *minimal allowed survival probability*. The magnitude of the penalty in the "allowed" $\langle\psi(T)|\psi(T)\rangle$ region should be close to 0. As $\langle\psi(T)|\psi(T)\rangle$ approaches the

minimal allowed survival probability from the upper limit, the magnitude of the penalty should increase rapidly. This can be realized by a sigmoid profile of $\sigma(y)$ (for an example see Fig. 3 in Sec. IV).

A further discussion on the interpretation of J_{ion} is given in Appendix E.

D. Imposing boundary conditions on the driving field

A realistic pulse shape must be a continuous function. Rapid jumps in $\epsilon(t)$ cannot be produced in laboratory conditions. Very rapid variations in $\epsilon(t)$ will be referred to as "discontinuities".

In the basic formulation of Sec. II A, the continuity of $\epsilon(t)$ *within* the time-interval of the problem, $t \in (0, T)$, is ensured by the limitation of the pulse to a low frequency regime. However, at the boundaries of the time-domain, $t = 0$, $t = T$, we encounter a problem: The physical value of $\epsilon(t)$ outside the time-domain $t \in [0, T]$ is zero by definition. However, in practice, $\epsilon(t)$ is constructed by a discrete spectral transform (see Appendix C 1), in which the ω sampling is discretized; discrete spectral transforms represent *infinite periodic patterns*, which are non-zero outside $t \in [0, T]$. Thus, the *physical* $\epsilon(t)$ is defined by

$$\epsilon(t) \equiv \begin{cases} \mathcal{C}^{-1}[\bar{\epsilon}(\omega)] & 0 \le t \le T \\ 0 & t < 0,\ t > T \end{cases} \tag{30}$$

The restriction of the frequency band by J_{energy} prevents the appearance of discontinuities in the infinite periodic function; it does not prevent the possibility of discontinuities in Eq. (30) at the boundaries of the time-interval of the problem, $t = 0$, $t = T$.

The problem can be solved by formulation of additional control requirements; the driving pulse has to satisfy the following *boundary conditions*:

$$\epsilon(0) = 0 \tag{31}$$

$$\epsilon(T) = 0 \tag{32}$$

The boundary conditions can be treated as additional *constraints* to the optimization problem. These can be realized by the Lagrange-multiplier method. Eqs. (31), (32), are treated as *constraint equations*.

Accordingly, the following Lagrange-multiplier terms are added to J:

$$J_{\epsilon(0)} \equiv -\sqrt{2\pi}\lambda_0\epsilon(0) = -2\lambda_0 \int_0^\Omega \bar{\epsilon}(\omega)\cos(0)\,d\omega = -2\lambda_0 \int_0^\Omega \bar{\epsilon}(\omega)\,d\omega \tag{33}$$

$$J_{\epsilon(T)} \equiv -\sqrt{2\pi}\lambda_T\epsilon(T) = -2\lambda_T \int_0^\Omega \bar{\epsilon}(\omega)\cos(\omega T)\,d\omega \tag{34}$$

where $\sqrt{2\pi}\lambda_0$, $\sqrt{2\pi}\lambda_T$, are the Lagrange-multipliers of the constraint equations (31), (32), respectively.

Prevention of discontinuities in $\epsilon(t)$ has a fundamental theoretical importance. If discontinuities are present at the boundaries, the spectrum of the physical $\epsilon(t)$ contains very high frequencies. This has considerable physical significance in the context of HHG, since very energetic photons are contained in the driving pulse. The HHG process is an up-conversion process, in which low energy photons in the driving pulse are converted by the system into high energy photons. The presence of highly energetic photons in the driving pulse completely changes the physical picture, and leads to spurious HHG effects.

A dynamical analysis of these spurious effects enables a more thorough understanding of their physical origin. A sudden change in the Hamiltonian leads to the violation of the *adiabatic approximation*. We found that non-adiabatic processes are a prerequisite to the generation of high harmonics. Typically, this is achieved by extremely high intensities. A discontinuity in the field generates spurious non-adiabatic transitions in much lower intensities. This topic requires a dedicated paper.

While this effect is typically small in magnitude, it is of considerable importance in the context of HHG, which is governed by small amplitude effects. In our previous work [44] the topic of the boundary conditions of the field was ignored (as in many OCT works). Thus, the physical significance of the results presented in the HHG problem of Ref. [44] is quite limited.

Actually, conditions (31), (32), are insufficient to prevent the appearance of highly energetic components in the physical $\epsilon(t)$ spectrum; discontinuities in the time derivatives,

$$\frac{d^n\epsilon(t)}{dt^n}, \qquad n \geq 1$$

are also a source of high frequency components in the spectrum, which can be questionable both experimentally and theoretically (this was ignored, e. g., in [48]). We found that a discontinuity in the first derivative is responsible for significant spurious HHG effects. However, the effect of a discontinuity in the second derivative on the high-harmonic spectrum was found to be negligible. Thus, the following

boundary conditions to $\epsilon(t)$ are added:

$$\left.\frac{d\epsilon(t)}{dt}\right|_{t=0} = 0 \tag{35}$$

$$\left.\frac{d\epsilon(t)}{dt}\right|_{t=T} = 0 \tag{36}$$

These boundary conditions are satisfied automatically by the DCT, in which $\epsilon(t)$ is spanned by a cosine series. This is an important advantage of the DCT over the discrete Fourier transform, in which the derivative boundary conditions define additional constraints in the optimization problem.

E. The full OCT formulation of HHG control

The total maximization functional is the sum of all the individual terms:

$$J \equiv J_{max} + J_{ion} + J_{energy} + J_{\epsilon(0)} + J_{\epsilon(T)} + J_{Schr} \tag{37}$$

where the different components are given by Eqs. (26), (27), (7), (33), (34), (15).

The extremum conditions yield the following set of Euler-Lagrange equations:

$$\frac{d\left|\psi(t)\right\rangle}{dt} = -i\hat{\mathbf{H}}(t)\left|\psi(t)\right\rangle,$$
$$\left|\psi(0)\right\rangle = \left|\psi_0\right\rangle \tag{38}$$
$$\frac{d\left|\chi(t)\right\rangle}{dt} = -i\hat{\mathbf{H}}^{\dagger}(t)\left|\chi(t)\right\rangle - \mathcal{C}^{-1}\left[f_C(\omega)\overline{\left\langle\hat{\mathbf{C}}\right\rangle}(\omega)\right]\hat{\mathbf{C}}\left|\psi(t)\right\rangle,$$
$$\left|\chi(T)\right\rangle = \sigma'\left(\left\langle\psi(T)|\psi(T)\right\rangle\right)\left|\psi(T)\right\rangle \tag{39}$$
$$\hat{\mathbf{H}}(t) = \hat{\mathbf{H}}_0 - \hat{\mathbf{X}}\epsilon(t) \tag{40}$$
$$\epsilon(t) = \mathcal{C}^{-1}[\bar{\epsilon}(\omega)] \tag{41}$$
$$\bar{\epsilon}(\omega) = \bar{\epsilon}_{unc}(\omega) - \tilde{f}_{\epsilon}(\omega)[\lambda_0 + \lambda_T\cos(\omega T)] \tag{42}$$
$$\bar{\epsilon}_{unc}(\omega) \equiv f_{\epsilon}(\omega)\mathcal{C}[\eta(t)], \qquad \eta(t) \equiv -\frac{\mathrm{Im}\left\langle\chi(t)\left|\hat{\mathbf{X}}\right|\psi(t)\right\rangle}{\alpha} \tag{43}$$
$$\epsilon_{unc}(t) = \mathcal{C}^{-1}[\bar{\epsilon}_{unc}(\omega)] \tag{44}$$
$$\lambda_0 = \frac{\epsilon_{unc}(0)d - \epsilon_{unc}(T)b}{ad - b^2} \tag{45}$$
$$\lambda_T = \frac{\epsilon_{unc}(T)a - \epsilon_{unc}(0)b}{ad - b^2} \tag{46}$$
$$a \equiv \mathcal{C}^{-1}\left[\tilde{f}_{\epsilon}(\omega)\right]\Big|_{t=0} \tag{47}$$
$$b \equiv \mathcal{C}^{-1}\left[\tilde{f}_{\epsilon}(\omega)\right]\Big|_{t=T} \tag{48}$$
$$d \equiv \mathcal{C}^{-1}\left[\tilde{f}_{\epsilon}(\omega)\cos(\omega T)\right]\Big|_{t=T} \tag{49}$$

The derivation of the equations is given in Appendix B. These equations form the base for the optimization procedure.

III. NUMERICAL METHODS

Optimal control equations are solved by iterative forward backward propagation of the Schrödinger equation with an update scheme to change the control field from iteration to iteration. The control of HG is a particularly difficult problem and therefore the standard procedures that have been developed for OCT do not apply. The solution of the Schrödiner equation is complicated due to explicit time dependence of the control Hamiltonian. An additional difficulty is that due to absorbing boundaries the Hamiltonian is non-Hermitian. High accuracy is required since the HHG phenomenon is generated from

a minor fraction of the wavefunction.

A. Optimization procedure

The optimization procedure is based on a second-order gradient method (quasi-Newton)—the BFGS method (see [54, Chapter 3] and references therein).

The BFGS method replaces the relaxation process employed in our previous work (see in detail [43, Chapter 3.2.3]). The relaxation process was found to be successful in the optimization of low-order HG processes. However, we found it to be rather slow in the optimization of the HHG process. This can be attributed to the complexity of the physical situation, which results in an optimization hypersurface which is more complex than that of the simpler HG problems. This necessitates the use of second-order information in the optimization (the Hessian), which can be approximated by a quasi-Newton method.

We found that the BFGS method yields a drastic improvement compared to the relaxation process. However, this required adjustments of several details of the implementation to the present problem. The implementation details are described in Appendix C 2.

B. Dynamics

The propagation with an explicit time dependent Hamiltonian is solved by a new highly accurate and efficient algorithm [55, 56]. The algorithm is based on a *semi-global propagation approach*, which is governed by multiple considerations which are both local in time, and global in time. The propagation is performed in relatively large time-steps, where each time-interval is approximated globally as a unified unit by an interpolation based approach. The application of the algorithm to the physical situation of HHG has already been described in detail in [56, Sec. 4].

The Fourier grid method [57, 58] is employed for the Hamiltonian operation.

Absorbing boundary conditions are employed in order to prevent a wraparound of the wave-function at the boundaries of the grid. The absorbing boundaries are implemented by a complex absorbing potential (see [59] for a comprehensive review). This implies that the Hamiltonian becomes non-Hermitian. The propagation method was extended to the treatment of non-Hermitian dynamics in Ref. [56].

The choice of the complex absorbing potential is of utmost importance in the present setting. It has already been recognized in the early HHG simulations (see [45]) that reflection of the wave-function from the absorbing boundaries leads to large distortions of the high-harmonic spectrum. Different absorbing potentials vary in their absorption capabilities. Reliable HHG simulations require absorbing potentials

which reduce the reflection and transmission of the wave-function to extremely low rates. Attempts to locate absorbing potentials with this property by inspection have failed [45].

This issue is crucial in the context of optimization. The optimization process cannot distinguish between a physical effect and a numerical artefact. As a result, it might tend to amplify the magnitude of a numerical artefact instead of maximizing a real physical effect.

In the present work, we employed an *optimization procedure* for the construction of the absorbing potential. The optimization is performed to locate an absorbing potential with minimal reflection and transmission. The method relies on the principles presented in [60], but with several necessary modifications. The real part of the absorbing potential is constructed from a finite cosine series. The imaginary part is constructed from another finite cosine series, where the imaginary potential is given by squaring the cosine series and adding a minus sign. This prevents the presence of positive imaginary components in the potential, which is a source of numerical instability in the propagation process. The optimization parameters are the cosine coefficients. The reflection and transmission amplitudes are obtained by static scattering calculations (see [61]). Further details are available in [56, Sec. 4.2]. However, a full description of the method has not been published to date.

IV. RESULTS

The optimization of selected harmonics in the range of the thirteenth to the seventeenth harmonic of the fundamental of a Ti-sapphire laser is demonstrated in a simple single electron model system. The selected target frequencies are above the ionization threshold. (*Remark*: Atomic units are used throughout.)

A. The system

Our system is a simplified one-dimensional model of an atomic system. A "one-dimensional electron" is placed in a central potential of a Coulombic nature, which represents the parent atom potential. The central potential has the form of a *truncated Coulomb potential* (see Fig. 1):

$$V(x) = 1 - \frac{1}{\sqrt{x^2 + 1}} \tag{50}$$

This form eliminates the singularity of the Coulomb potential at $x = 0$.

The one-dimensional model has been extensively studied in the context of intense laser atomic physics

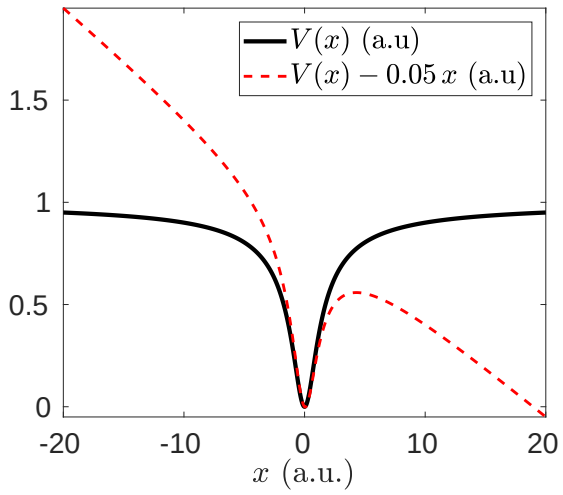

FIG. 1. The truncated Coulomb potential (Eq. (50), black), and the distorted potential under the influence of a strong field (dashed red).

in general, and HHG in particular. It is by no means a realistic atomic model; nevertheless, it preserves the fundamental features of the high-harmonic spectrum [45]. The model constitutes the simplest system for which the HHG phenomenon can be observed. The simplicity of the model is a considerable advantage for the study of HHG. The realistic HHG process is extremely complicated and rich with physical effects of secondary importance, such as multielectron effects [14, 48], effects of higher dimensionality (e. g. spreading of the electron in transversal directions to the polarization of the driving field; nonlinear interaction between spatial degrees of freedom), and macroscopic propagation effects [53]. The current simplified model refines the most fundamental elements of the HHG process, which can contribute to their study and understanding.

The time-dependent *physical* Hamiltonian becomes

$$\hat{\mathbf{H}}(t) = \frac{\hat{\mathbf{P}}^2}{2} + 1 - \frac{1}{\sqrt{\hat{\mathbf{X}}^2 + 1}} - \hat{\mathbf{X}}\epsilon(t) \tag{51}$$

This Hamiltonian is supplemented by a complex potential to account for the absorbing boundaries. In addition, the potential is modified such that the classical force induced by the physical potential is smoothly "turned off" near the absorbing boundaries, as described in detail in [56, Sec. 4].

The fundamental Bohr frequency of the model system, $\omega_{1,0} = 0.395$ a.u., is similar to that of the hydrogen atom (0.375 a.u.) or the argon atom (0.424 a.u.).

The spatial domain is $x \in [-240, 240)$. We use an equidistant grid, with 768 points. The distance between adjacent grid points becomes 0.625 a.u.. The absorbing potential extends over 40 a.u. at each

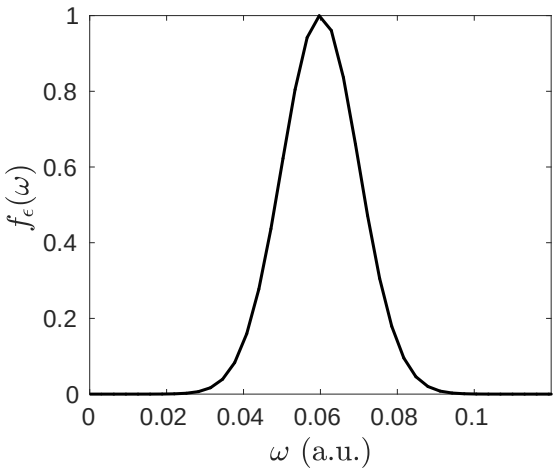

FIG. 2. The driving field filter function

boundary of the x grid (see [56, Sec. 4.2]). Thus, the actual *physical domain* is $x \in [-200, 200]$.

B. General details of the control problems

The initial state $|\psi_0\rangle$ in all problems is the ground state of the stationary Hamiltonian. The time interval allocated to the process is $t \in [0, 1000\,\text{a.u.}]$, which corresponds to a pulse duration of 24.2_{fs}. The driving field filter function has a Gaussian profile (see Fig. 2):

$$f_\epsilon(\omega) = \exp\left[-\frac{(\omega - 0.06)^2}{2 \cdot 0.01^2}\right] \tag{52}$$

The Gaussian is centred at $\omega = 0.06\,\text{a.u.}$, which corresponds to a wavelength $\lambda = 760_{nm}$—similar to the central wavelength of the Titanium-Sapphire laser. Let us denote:

$$\omega_0 = 0.06\,\text{a.u.} \tag{53}$$

ω_0 will be referred to as the *fundamental frequency* of the laser source.

The filter function of the target frequency has also a Gaussian profile:

$$f_C(\omega) = \exp\left[-\frac{(\omega - n\omega_0)^2}{2 \cdot 0.01^2}\right] \tag{54}$$

where n denotes the harmonic order of the target frequency. n varies with the specific optimization problem.

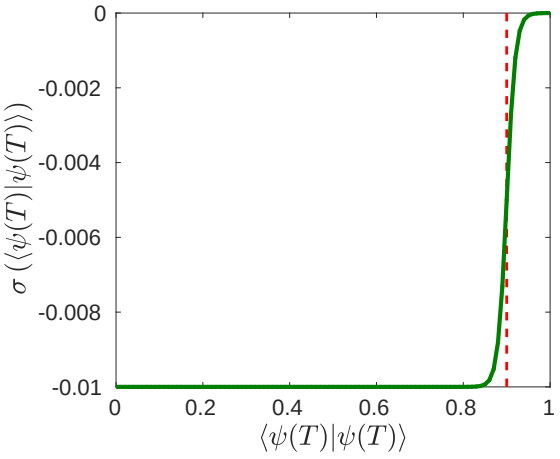

FIG. 3. The ionization penalty function; the minimal allowed survival probability (0.9) is marked by a vertical dashed red line.

The total energy penalty factor is $\alpha = 2 \times 10^{-6}$.

The ionization penalty function is (see Fig. 3)

$$\sigma(y) = 5 \times 10^{-3} \left\{ \tanh\left[50\,(y - 0.9)\right] - \tanh(5) \right\} \tag{55}$$

$\sigma(y)$ is chosen so as to restrict the maximal allowed permanent ionization to 10% of the probability.

The initial guess for the driving field is constructed from an unconstrained function $\bar{\epsilon}^{(0)}_{unc}(\omega)$, substituted into Eqs. (42), (44)-(46) (as explained in Appendix C 2 d). The following unconstrained field has been used for all problems:

$$\bar{\epsilon}^{(0)}_{unc}(\omega) = 5 \, \exp\left[-\frac{(\omega - 0.06)^2}{2 \cdot 0.01^2} \right] \, \sin\left[\frac{(\omega - 0.06)\pi}{0.015} \right] \tag{56}$$

The termination condition is given by Eq. (C26).

C. Maximization of selected harmonics

The maximization of the 13'th harmonic is the first target to be studied. The target filter function is given by the substitution of $n = 13$ into Eq. (54). The target frequency, $\omega = 0.78$ a.u., is the first odd-harmonic frequency which is above the ionization threshold frequency, 0.67 a.u..

The inverse-Hessian approximation was reset after 7 iterations (see Appendix C 2 f), when the termination condition (C26) was first matched. The convergence curve is shown in Fig. 4, where the optimization process is divided in two stages—before and after the reset of the Hessian.

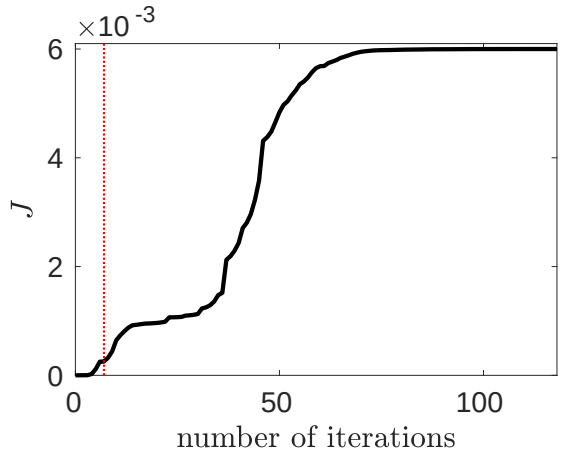

FIG. 4. The convergence curve for the optimization of the 13'th harmonic; the process is divided into two stages by a doted vertical red line—before and after the reset of the Hessian (see Appendix C 2 f).

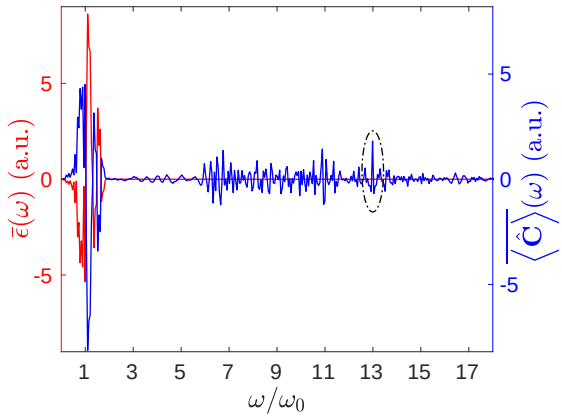

FIG. 5. The spectra of the driving field (red) and the stationary acceleration expectation (blue) Vs. the harmonic order, for the 13'th harmonic problem; the response at the target frequency is marked.

The spectra of the optimized driving field and the stationary acceleration expectation are plotted in Fig. 5 Vs. the harmonic order. The driving field is successfully restricted to the region of the fundamental frequency of the source. The response at the 13'th harmonic is marked. The enhanced emission at the target frequency is apparent.

Several other important peaks are present in the emission spectrum: There is a large linear response around ω_0; a large peak is present at the fundamental Bohr frequency of the system, which equals $6.6\omega_0$; a significant response is observed at the 11'th harmonic, which is the neighbouring odd-harmonic of the target harmonic. The significant response at neighbouring harmonics is typical since no suppression of other harmonics was included in the control requirements (see Sec. II A).

Let us examine more carefully the spectral restriction of the driving field frequency by the filter

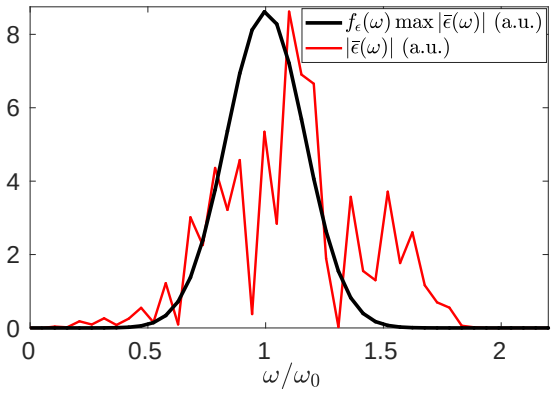

FIG. 6. A comparison of the forms of the filter function and the driving field spectrum of the 13'th harmonic problem; the absolute value of the driving field spectrum, $|\bar{\epsilon}(\omega)|$ (red), and the filter function $f_\epsilon(\omega)$ normalized to the peak of the driving field spectrum (black), are plotted Vs. the harmonic order. It is shown that there is a significant deviation from the filter function envelope, in particular at the blue side of the spectrum. An additional general blue shift can be also observed.

function $f_\epsilon(\omega)$. In Fig. 6, we compare the magnitude of $\bar{\epsilon}(\omega)$ with the form of $f_\epsilon(\omega)$. It can be observed that the general form of $\bar{\epsilon}(\omega)$ is influenced by the Gaussian envelope shape imposed by $f_\epsilon(\omega)$. However, it can be also observed that there is a significant deviation from the filter function envelope. The decay rate of $\bar{\epsilon}(\omega)$ is slower than that of $f_\epsilon(\omega)$, in particular at the blue side of the spectrum. A blue shift can be observed also at the central region of the emission spectrum. It can be found that the vast majority of the contribution to the magnitude of J_{energy} originates from the wings of the driving field spectral profile, in particular at the blue side. This implies that the extension of the available spectrum from the laser source has a fundamental role in the enhancement of the HHG process, in particular the blue extension. This statement can be verified by the variation of the standard-deviation of $f_\epsilon(\omega)$. A more thorough physical analysis is left for a future publication.

Let us compare the harmonic yield produced by the optimized pulse at the target frequency with that of a reference pulse. Our reference pulse is constructed from a periodic wave with frequency ω_0, which is constrained to the required boundary conditions of the present problem. We start from a periodic waveform, confined to a finite time-duration:

$$\epsilon_{harmonic}(t;\beta) \equiv \beta \sin[\omega_0(t-500)], \qquad t \in [0, 1000] \tag{57}$$

This form is used to construct a field which satisfies the cosine series boundary conditions, (35), (36):

$$\epsilon_{ref,unc}(t;\beta) = \mathcal{C}^{-1}\left\{f_\epsilon(\omega)\mathcal{C}\left[\epsilon_{harmonic}(t;\beta)\right]\right\} \tag{58}$$

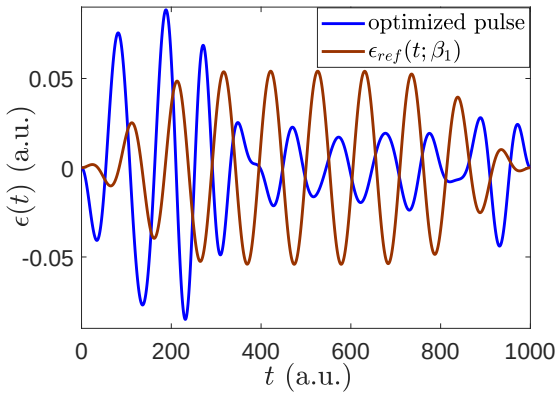

FIG. 7. The temporal profiles of the optimized pulse (blue) and the reference pulse normalized to the same total energy (brown).

Finally, $\epsilon_{ref,unc}(t;\beta)$ is substituted in Eqs. (42), (44)-(46), in order to obtain a new pulse $\epsilon_{ref}(t;\beta)$ which satisfies also the zero boundary conditions of $\epsilon(t)$, (31), (32) (as explained in Appendix B).

The optimized pulse is compared to $\epsilon_{ref}(t;\beta)$ with two different values of β. The two values correspond to normalization of $\epsilon_{ref}(t;\beta)$ to two different physical properties of the optimized pulse:

1. The fluence $\Phi\left[\epsilon(t)\right]$ (Eq. (11)), or equivalently, the total energy;

2. The peak intensity, or equivalently, $\max|\epsilon(t)|$.

The two choices of β will be marked as β_1, β_2, respectively.

The temporal profiles of the optimized pulse and $\epsilon_{ref}(t;\beta_1)$ are plotted in Fig. 7.

The response of the different pulses at the 13'th harmonic is compared in Fig. 8. It can be observed that $\epsilon_{ref}(t;\beta_1)$ does not induce a significant response. The response is significantly enhanced under the influence of $\epsilon_{ref}(t;\beta_2)$. However, the response of $\epsilon_{ref}(t;\beta_2)$ is still not comparable to that of the optimized pulse.

In Table I we compare for the three pulses the survival probability at the end of the process, the fluence, and the J_{max} value (as defined in Eq. (26)). The ionization probability of the optimized pulse is successfully restricted to the allowed rate, defined by $\sigma(\langle\psi(T)|\psi(T)\rangle)$ (less than 10%). The ionization probability of $\epsilon_{ref}(t;\beta_1)$ is very small, which implies that the response of the system to the exerted field is not significant. The ionization probability of $\epsilon_{ref}(t;\beta_2)$ is close to 22%, which is significantly higher than the optimized pulse, and beyond our control requirements. The fluence of $\epsilon_{ref}(t;\beta_2)$ is also considerably larger. It is shown that the optimized pulse induces a much larger response in the target frequency, with considerably lower ionization probability and total energy.

Other target harmonics were optimized: The 15'th, 17'th and 14'th harmonics.

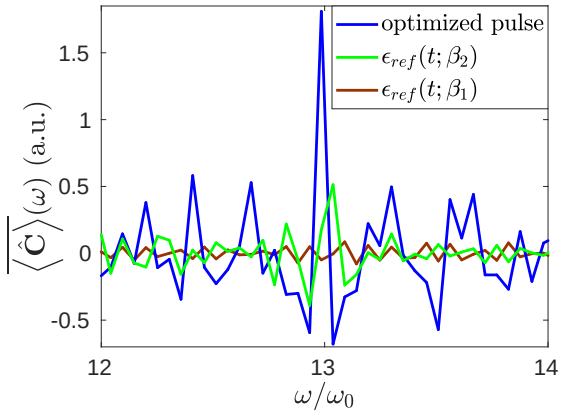

FIG. 8. The stationary acceleration expectation spectra of the optimized pulse (blue), the reference pulse normalized to the same peak intensity (green) and the reference pulse normalized to the same total energy (brown); the region of the target harmonic of the optimized pulse (the 13'th harmonic) is shown.

Pulse	$\langle\psi(T)\|\psi(T)\rangle$	$\Phi\left[\epsilon(t)\right]$ (a.u.)	J_{max} (a.u.)
optimized	0.926	0.992	6.97×10^{-3}
$\epsilon_{ref}(t;\beta_1)$	0.999	0.992	3.87×10^{-5}
$\epsilon_{ref}(t;\beta_2)$	0.783	2.65	8.73×10^{-4}

TABLE I. The survival probability at $t=T$, the fluence, and the J_{max} value of the optimized pulse and the two reference pulses.

The targeting of an even harmonic (the 14'th) is used to test the limits of the control opportunities. It is well known that the typical HHG spectrum consists of odd harmonics. This property of the HHG spectrum can be referred to as the *selection rules* of the process. It is shown in [62, 63] that the HHG selection rules originate in the symmetry properties of the problem. The derivation of the selection rules is based on the *Floquet formalism*, which is defined for a CW field. Nevertheless, a quasi-periodic pulse of finite duration can be approximated by the Floquet formalism. The symmetry of the problem in the idealized Floquet formalism is combined from spatial symmetry properties (the central potential of the atom is symmetric under inversion; the dipole operator is anti-symmetric under inversion) and temporal symmetry properties (the periodic harmonic waveform is anti-symmetric under a temporal shift of a half period; see [62, 63] for a fuller explanation).

However, the assumptions underlying the derivation of the selection rules in [62, 63] may not hold in an optimized pulse. We shall mention several important issues:

1. The temporal symmetry properties of the harmonic waveform assumed in [62, 63] can be broken

in an optimized field.

2. It is assumed in [62, 63] that the system is found in a single Floquet state. This assumption needn't be valid if the system is appropriately controlled to be in a superposition of such states.

3. The assumption that the system can be approximately described by the Floquet formalism may totally collapse in an optimized pulse, which may considerably deviate from a quasi-periodic template.

All these issues are related to the *breaking of symmetry* of the problem. The question is to which extent this broken symmetry can lead to the enhancement of the "forbidden" even harmonics.

During the optimization processes of the 14'th and 15'th harmonics, it was necessary to reset the Hessian approximation, after the termination condition (C26) was first matched.

In Fig. 9, we present the response of the system in the region of interest for the four problems (including the 13'th harmonic problem). It is shown that the response is selectively enhanced in all the required frequencies.

Remarkably, the enhancement of the response at the even 14'th harmonic target is as successful compared to the other odd harmonic targets. In Fig. 10, the temporal profile of the optimized pulse is shown. The field seems to be considerably more complex than the field of the optimized pulse of the 13'th harmonic problem, presented in Fig. 7. The reason could be the necessity of breaking symmetry in the even harmonic problem. It can be readily observed that the optimized field does not satisfy the symmetry properties of a harmonic waveform. In addition, the Floquet formalism clearly becomes inappropriate for the description of this very complex profile.

As in the spectrum of the 13'th harmonic problem, significant response in neighbouring harmonics appears also in the other spectra presented in Fig. 9. Interestingly, for the odd harmonic targets, only odd neighbouring harmonics are observed. This implies that the symmetry properties of HHG inferred from a CW field, remain relevant for the description of the process induced by the optimized pulses. In contrary, in the 14'th harmonic problem we can observe the appearance of both an odd harmonic (the 13'th) and an even one (the 16'th). This could be expected by the break of symmetry which allows the appearance of the target frequency.

In Table II we present the survival probability at the end of the process for all our target frequencies. The ionization probability in all problems is successfully restricted to less than 10%, and lies in the range of $6-8$ %.

The validity of the approximations introduced by the employment of absorbing boundary conditions

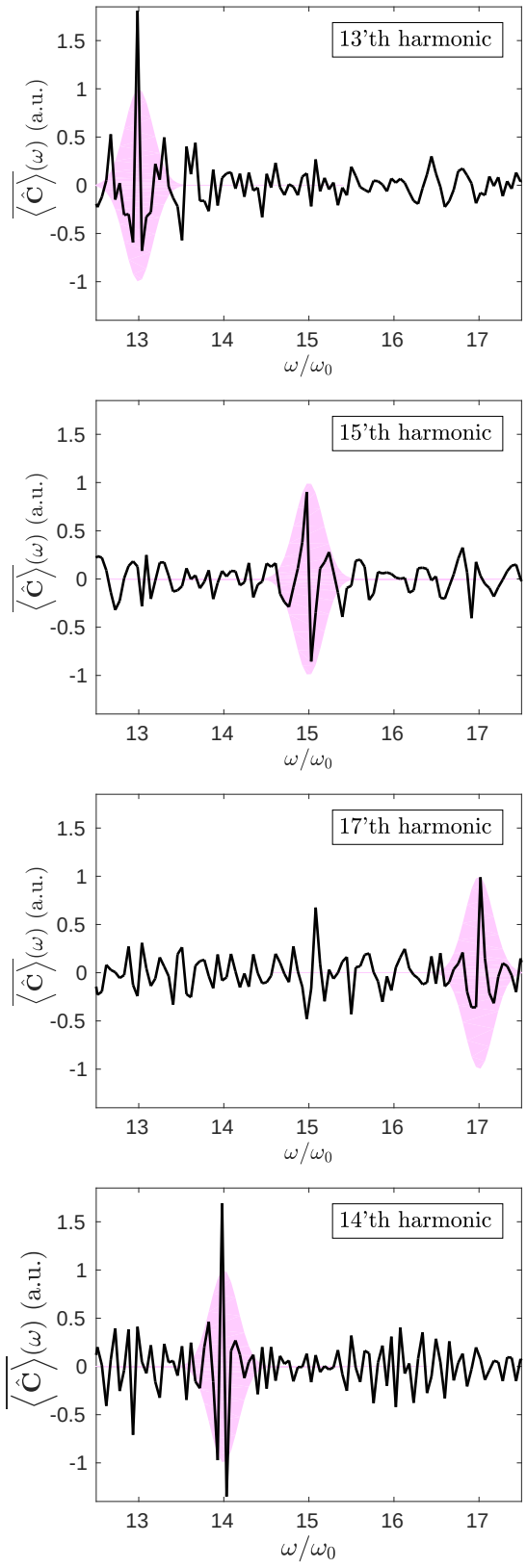

FIG. 9. The stationary acceleration expectation spectra of the various target harmonic problems in the region of interest; the profile of the target filter function is marked in lite magenta for each problem.

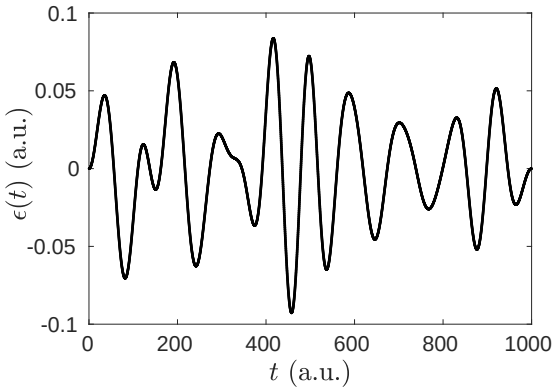

FIG. 10. The temporal profile of the optimized pulse in the 14'th harmonic problem.

n	$\langle\psi(T)\|\psi(T)\rangle$
13	0.926
14	0.923
15	0.932
17	0.937

TABLE II. The survival probability at $t = T$ of the optimized pulses for the different target frequencies.

has been tested for all problems, as described in Appendix D.

V. CONCLUSION

In the present study, an optimization method of HHG has been developed in the framework of quantum OCT. The target was a specific emission frequency. Several restrictions have been imposed in the OCT formulation as "hard" and "soft" constraints. In particular, the restriction of permanent ionization has been addressed by its formulation as a soft constraint. This requirement has not been addressed in previous theoretical studies.

Special emphasis was given to the numerical implementation. The simulation of the dynamics was performed by highly accurate methods, which is crucial to achieve a reliable description of the HHG process. The solver of the explicitly time dependent Schrödinger equation was performed by a new highly accurate approach [55, 56]. A new optimization method was employed for the construction of the complex-absorbing-potential, used for the realization of the absorbing boundary conditions. The complexity of the HHG optimization problem required the employment of a second-order gradient method

for the optimization process. Several details of the implementation to the present problem required special care.

The results demonstrated significant selective enhancement of the harmonic yield, with simultaneous minimization of the total energy of the driving pulse and control of the permanent ionization probability. The violation of the high harmonic selection rules has also been demonstrated.

The present paper is devoted to the control aspects of the optimization method. The physical interpretation of optimized fields requires a separate thorough discussion, and thus is beyond the framework of the present paper. Nevertheless, we shall briefly mention a particular insight into the mechanisms underlying the optimized fields in the odd harmonic problems of Sec. IV. We found that the optimized HHG processes consist of two stages:

1. Liberation of the electron from the adiabatic regime; it was found that adiabaticity imposes a barrier on the initiation of the HHG process (as has already been mentioned in Sec. II D).

2. Generation of harmonics by quasi-periodic patterns of the driving fields.

The two stage pattern is demonstrated in Fig. 11; the temporal profiles of the optimized fields in the odd harmonic problem are plotted, where the division into two stages in each optimized field is marked by a vertical dashed red line (it should be noted that the transition from the first stage to the second one does not take place in a well-defined time-point, and thus there is some arbitrariness in this marking). The first stage is characterized by non-periodic patterns, with higher intensities and a tendency to higher frequencies, while the second one is characterized by quasi-periodic patterns, with lower intensities and frequencies around the fundamental frequency of the source. Both the high intensities and the high frequencies in the first stage contribute to the deviation from the adiabatic regime. A fuller discussion is left for a future publication.

We hope that the present optimization method will contribute to a fuller understanding of physics of the HHG process in general, and of optimal pulses for controlling an emission line in particular.

ACKNOWLEDGMENTS

We thank Daniel Strasser, Roi Baer and Nimrod Moiseyev for useful discussions. This research was supported by the Israel Science Foundation, Grant No. 2244/14.

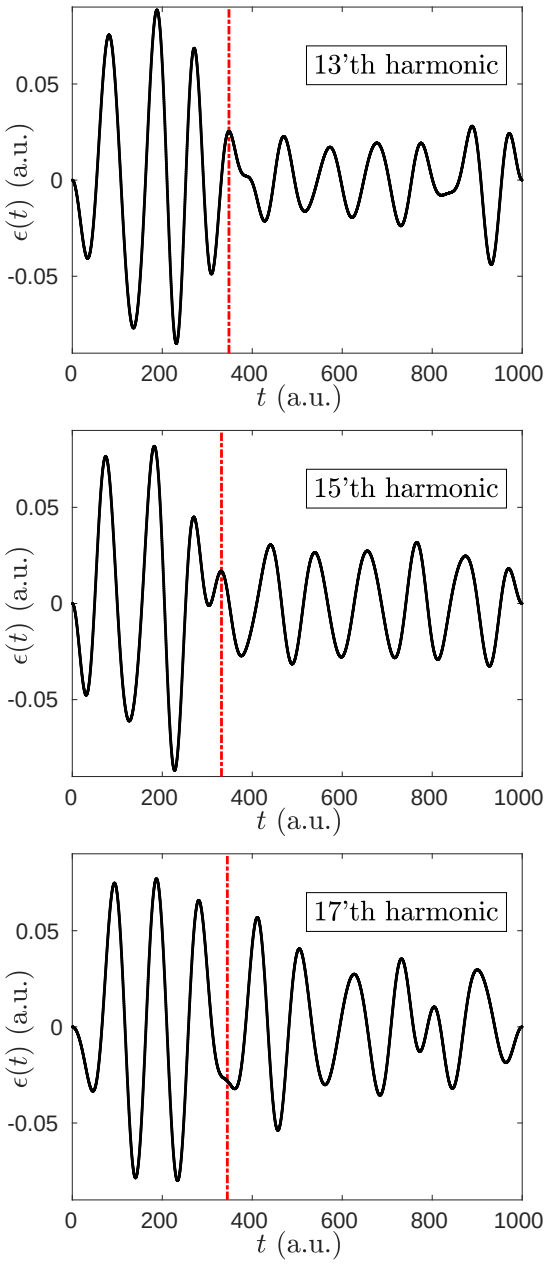

FIG. 11. The temporal profiles of the optimized pulses in the odd-harmonic problems; the process can be divided into two stages, where the first stage is characterized by non-periodic patterns, higher intensities and a tendency to higher frequencies, while the second one is characterized by quasi-periodic patterns, with lower intensities and frequencies around the fundamental frequency of the source. The division into two stages is marked by a vertical dashed red line.

Appendix A: The choice of the target operator

The target operator $\hat{\mathbf{O}}$ has been chosen as the stationary acceleration operator $\hat{\mathbf{C}}$ (see Sec. II B). The present appendix discusses the considerations which lead to this choice.

As an introduction, we shall present several obvious candidates for $\hat{\mathbf{O}}$ which are related to the dipole motion. For simplicity, we consider the dipole of a single electron. Only the x component is considered, in accordance with the polarization of the driving field. The following observables are directly related to the dipole motion:

1. The dipole operator:

$$\hat{\mu}_x = \hat{\mathbf{X}} \tag{A1}$$

2. The dipole velocity operator:

$$\hat{\dot{\mu}}_x = \hat{\dot{\mathbf{X}}} = \hat{\mathbf{P}}_x \tag{A2}$$

3. The dipole acceleration operator (see Sec. II B):

$$\hat{\ddot{\mu}}_x = \hat{\ddot{\mathbf{X}}} = -\frac{dV\left(\hat{\mathbf{X}}\right)}{d\hat{\mathbf{X}}} \tag{A3}$$

Note that atomic units are used, and the electron has a unit mass.

The spectra of the three observables can be related mathematically. The simplest relations are between the *Fourier* spectral functions. Let us use the following notation for the Fourier transform of an arbitrary function $g(t)$:

$$g^f(\omega) \equiv \mathcal{F}\left[g(t)\right] \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g(t) \exp(-i\omega t)\, dt \tag{A4}$$

$g(t)$ is restored from $g^f(\omega)$ by the inverse Fourier transform:

$$g(t) = \mathcal{F}^{-1}\left[g^f(\omega)\right] \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g^f(\omega) \exp(i\omega t)\, d\omega \tag{A5}$$

The three Fourier spectral functions are related by

$$\left\langle \hat{\ddot{\mu}}_x \right\rangle^f(\omega) = i\omega \left\langle \hat{\dot{\mu}}_x \right\rangle^f(\omega) = -\omega^2 \langle \hat{\mu}_x \rangle^f(\omega) \tag{A6}$$

The relations (A6) can be derived from the inverse Fourier expression for the dipole expectation,

$$\langle \hat{\mu}_x \rangle(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \langle \hat{\mu}_x \rangle^f(\omega) \exp(i\omega t)\, d\omega \tag{A7}$$

by considering the single and the double differentiation of the equation w.r.t. t. It can be seen from

(A6) that the three spectra *contain the same spectral regions*. However, the higher frequency regions are emphasized in higher derivatives of $\hat{\mu}_x$. The three spectral functions also differ by a phase shift.

The *cosine transform* spectra of the dipole and the acceleration are related by

$$\overline{\left\langle \ddot{\hat{\mu}}_x \right\rangle}(\omega) = -\omega^2 \overline{\langle \hat{\mu}_x \rangle}(\omega) \tag{A8}$$

The relation to $\overline{\left\langle \dot{\hat{\mu}}_x \right\rangle}(\omega)$ is more complex. However, the general picture remains similar, where the higher frequencies are emphasized in higher derivatives.

The choice of the target operator $\hat{\mathbf{O}}$ requires discussions on two different grounds:

1. In the *physical ground*, we first have to identify the observable associated as the source of emission.

2. In the *numerical ground*, we should identify the observable which is the preferable one from a numerical viewpoint. The spectrum of this observable can be related mathematically to the emission spectrum, which may be associated with another observable.

We begin from the physical discussion. In classical electrodynamics an accelerated charge emits radiation proportional to the *acceleration* of the charge. Therefore, it has been assumed that the high-harmonic spectrum is proportional to the dipole acceleration spectrum. However, in 2008 it was claimed in [64] that the emission high-harmonic spectrum is proportional to the *velocity* spectrum. Currently, there was a debate on this topic in the literature (see [65–67]). In recent publications there is still no consensus on this point.

The present study focuses on the optimization of selected harmonics (see Sec. IV). Therefore, the target band extends over a small interval in the spectrum. Hence, the ω factor difference between the acceleration and velocity spectra does not change significantly over the target region, and thus does not make an important difference between the spectra. Therefore, the physical question is not of fundamental importance for our results.

However, the numerical question has considerable importance. First, we have to introduce the numerical problem of concern. Spectra produced by discrete spectral transforms almost always contain background of numerical origin. This numerical background might hinder small magnitude physical effects. There are three important sources of background:

1. *Discretization*: The discrete spectral transforms are based on discrete basis functions, with a discrete frequency sampling. If the signal contains a frequency which is not present in the discrete set, the transformed signal contains peaks at the nearest frequencies, but with a long "tail", which

induces background noise throughout the spectrum. The magnitude of the background depends on the magnitude of the peak, and the distance of the represented frequency from the nearest frequency in the discrete set. In the discrete cosine and sine transforms, the frequency sampling is twice as dense as the Fourier sampling; however, the phase flexibility of the discrete Fourier transform is lost, with a fixed phase for each frequency component. The effect of mismatch of phase of the signal with the basis functions has the same effect as that of frequency mismatch.

2. *Truncation error*: The discrete time-sampling results in a truncation of the spectral series at a cutoff frequency Ω. The truncation error induces the effect of *aliasing*.

3. *Boundary effects*: Each spectral transform is adjusted to certain boundary conditions of the signal. $\left\langle \hat{\mathbf{O}} \right\rangle(t)$ is not expected to satisfy any of these boundary conditions at the final time T. This results in discontinuities at the boundary (a discrete spectral series represents an *infinite periodic signal*, which is defined also for $t > T$). A discontinuity contains frequencies from the entire spectrum, and induces noise throughout the spectrum (see [43, Chapter 3]). The DCT has the advantage that the discontinuity is in the *derivative* of the signal, and not in the function value; this reduces drastically the boundary effects compared to the discrete Fourier and sine transforms. However, the effect is not completely eliminated.

In our previous study [43, 44] we targeted the dipole spectrum, $\overline{\langle \hat{\mu}_x \rangle}(\omega)$. The problem with this choice is that the spectrum is scaled by $1/\omega$ compared to the velocity spectrum, or $1/\omega^2$ compared to the acceleration spectrum. This led to difficulties in the optimization of higher frequencies, which were hindered by the background of the spectrum. The problem disappeared when we adopted the choice of the acceleration operator spectrum as the optimization target, as in [47, 48]. In the acceleration spectrum, the magnitude of the physical information in higher frequencies is amplified compared to the other spectra discussed here. Hence, it becomes the preferable choice for harmonic generation problems. Conversely, in down conversion optimization problems the dipole spectrum is expected to be preferable.

In principle, if necessary, it is possible to obtain further amplification of higher frequencies; the target operator can be chosen as a time-derivative of the dipole of a higher order than the second. This was not required in the present study.

The acceleration spectrum still contains a large linear response component, which induces background throughout the spectrum. A further improvement can be achieved with some insight into the acceleration spectrum. As was mentioned in Sec. II B, the acceleration operator can be divided into a stationary

part and a time-dependent part:

$$\ddot{\hat{\mathbf{X}}} = \hat{\mathbf{C}} + \epsilon(t) \tag{A9}$$

By linearity of the spectral transform, the acceleration spectrum can also be divided into two parts:

$$\overline{\left\langle \ddot{\hat{\mathbf{X}}} \right\rangle}(\omega) = \overline{\left\langle \hat{\mathbf{C}} \right\rangle}(\omega) + \bar{\epsilon}(\omega) \tag{A10}$$

The second term of the RHS is the driving field spectrum, which contains only *linear response* in the low frequency regime. It is irrelevant to the region of interest in the spectrum—the high harmonic regime. Thus, the part of interest in the acceleration spectrum is the *stationary acceleration* only [48]. Accordingly, we can set:

$$\hat{\mathbf{O}} \equiv \hat{\mathbf{C}} \tag{A11}$$

The omission of the time-dependent part of the acceleration drastically reduces the magnitude of the linear response peaks in the spectrum. Consequently, the background is considerably reduced. This reveals another advantage of the acceleration spectrum over the dipole or velocity spectra—the majority of the linear response can be isolated and eliminated from the spectrum.

Appendix B: The derivation of the Euler-Lagrange equations

For the sake of clarity, let us first summarize the different components of the maximization functional. The full maximization functional is

$$J \equiv J_{max} + J_{ion} + J_{energy} + J_{\epsilon(0)} + J_{\epsilon(T)} + J_{Schr} \tag{B1}$$

The different terms in the functional can be classified as follows:

1. A maximization term,

$$J_{max} \equiv \frac{1}{2}\int_0^{\Omega} f_C(\omega)\overline{\left\langle \hat{\mathbf{C}} \right\rangle}^2(\omega)\, d\omega, \qquad f_C(\omega) \geq 0, \qquad \max\left[f_C(\omega)\right] = 1 \tag{B2}$$

   where

$$\overline{\left\langle \hat{\mathbf{C}} \right\rangle}(\omega) \equiv \sqrt{\frac{2}{\pi}}\int_0^{T} \left\langle \hat{\mathbf{C}} \right\rangle(t)\cos(\omega t)\, dt \tag{B3}$$

2. Two penalty terms, which represent "soft constraints":

(a) The constraint on the ionization probability is represented by

$$J_{ion} \equiv \sigma\left(\langle\psi(T)|\psi(T)\rangle\right), \qquad \sigma(y) \leq 0, \qquad \sigma(1) = 0 \tag{B4}$$

(b) The following term represents the constraint on the energy of the incident field, as well as the restriction of its spectrum:

$$J_{energy} \equiv -\int_0^\Omega \frac{1}{\tilde{f}_\epsilon(\omega)} \bar{\epsilon}^2(\omega)\, d\omega, \qquad \tilde{f}_\epsilon(\omega) > 0 \tag{B5}$$

where $\bar{\epsilon}(\omega)$ defines the spectral representation of $\epsilon(t)$ as follows:

$$\epsilon(t) = \sqrt{\frac{2}{\pi}} \int_0^\Omega \bar{\epsilon}(\omega) \cos(\omega t)\, d\omega \tag{B6}$$

3. Three Lagrange-multiplier terms, which represent three "hard" constraints:

(a) The constraint of the zero boundary condition on $\epsilon(t)$ at $t = 0$,

$$\epsilon(0) = 0, \tag{B7}$$

is represented by the following Lagrange-multiplier term:

$$J_{\epsilon(0)} \equiv -\sqrt{2\pi}\lambda_0 \epsilon(0) = -2\lambda_0 \int_0^\Omega \bar{\epsilon}(\omega) \cos(0)\, d\omega = -2\lambda_0 \int_0^\Omega \bar{\epsilon}(\omega)\, d\omega \tag{B8}$$

(b) The constraint of the zero boundary condition on $\epsilon(t)$ at $T = 0$,

$$\epsilon(T) = 0, \tag{B9}$$

is represented by the following Lagrange-multiplier term:

$$J_{\epsilon(T)} \equiv -\sqrt{2\pi}\lambda_T \epsilon(T) = -2\lambda_T \int_0^\Omega \bar{\epsilon}(\omega) \cos(\omega T)\, d\omega \tag{B10}$$

(c) The J_{Schr} term represents the Scrödinger equation constraint on $|\psi(t)\rangle$,

$$\frac{d\,|\psi(t)\rangle}{dt} = -i\hat{\mathbf{H}}(t)\,|\psi(t)\rangle \tag{B11}$$

The Schrödinger equation relates $|\psi(t)\rangle$ to $\bar{\epsilon}(\omega)$ via the time-dependent Hamiltonian,

$$\hat{\mathbf{H}}(t) = \hat{\mathbf{H}}_0 - \hat{\mathbf{X}}\epsilon(t) = \hat{\mathbf{H}}_0 - \hat{\mathbf{X}}\left(\sqrt{\frac{2}{\pi}}\int_0^{\Omega}\bar{\epsilon}(\omega)\cos(\omega t)\,d\omega\right) \tag{B12}$$

The Schrödinger equation constraint has to be imposed at each time-point of the time-interval $t \in [0, T]$.

In order to formulate the constraint correctly, we should note that the vector $|\psi(t)\rangle$ is a complex entity. Hence, each of its components consists of two degrees of freedom in each time-point—the real and imaginary parts. The Schrödinger equation, being a complex equation, constrains both degrees of freedom. Each of these constraints requires a distinct Lagrange-multiplier term. It is convenient to treat $|\psi(t)\rangle$ and $\langle\psi(t)|$ as independent variables, in order to represent the two degrees of freedom of the complex state vector. The constraint equation on $|\psi(t)\rangle$ is (B11). The constraint equation on $\langle\psi(t)|$ is its adjoint equation,

$$\frac{d\langle\psi(t)|}{dt} = i\langle\psi(t)|\hat{\mathbf{H}}^{\dagger}(t) \tag{B13}$$

These constraint equations are subject to the following initial conditions, respectively:

$$|\psi(0)\rangle = |\psi_0\rangle \tag{B14}$$

$$\langle\psi(0)| = \langle\psi_0| \tag{B15}$$

The constraint equations, (B11), (B13), together with the initial conditions, (B14), (B15), ensure that

$$\langle\psi(t)| = |\psi(t)\rangle^{\dagger} \tag{B16}$$

in all t. Assuming this, only Eqs. (B11), (B14) are required for the performance of the computations in practice.

The resulting Lagrange-multiplier term is a continuous summation of the constraint terms over t:

$$J_{Schr} \equiv$$
$$-\left[\int_0^T \left\langle \chi(t)\left|\frac{d\psi(t)}{dt} + i\hat{\mathbf{H}}(t)\psi(t)\right.\right\rangle dt + \int_0^T \left\langle \frac{d\psi(t)}{dt} + i\hat{\mathbf{H}}(t)\psi(t)\right|\left.\chi(t)\right\rangle dt\right] \tag{B17}$$

In this stage, $|\chi(t)\rangle$ and $\langle\chi(t)|$ are treated as *independent Lagrange-multiplier functions*. Nevertheless, the symmetry of the problem under the adjoint operation suggests that

$$\langle\chi(t)| = |\chi(t)\rangle^{\dagger} \tag{B18}$$

This justifies the use of the same letter χ for both Lagrange-multipliers. The full justification to (B18) will be given in what follows. For the sake of brevity, we shall rely on Eq. (B18) before it was fully justified, and write J_{Schr} as

$$J_{Schr} = -2\,\mathrm{Re}\int_0^T \left\langle \chi(t) \left| \frac{d}{dt} + i\hat{\mathbf{H}}(t) \right| \psi(t) \right\rangle dt \tag{B19}$$

The extremum conditions are:

$$\frac{\delta J}{\delta \bar{\epsilon}(\omega)} = 0 \tag{B20}$$

$$\frac{\delta J}{\delta \left|\psi(t)\right\rangle} = 0 \tag{B21}$$

$$\frac{\delta J}{\delta \left\langle\psi(t)\right|} = 0 \tag{B22}$$

$$\frac{\delta J}{\delta \left|\psi(T)\right\rangle} = 0 \tag{B23}$$

$$\frac{\delta J}{\delta \left\langle\psi(T)\right|} = 0 \tag{B24}$$

J_{Schr} is more easily handled after integrating by part the following expression:

$$\int_0^T \left\langle \chi(t) \left| \frac{d\psi(t)}{dt} \right\rangle dt \right.$$

We obtain:

$$J_{Schr} = -2\,\mathrm{Re}\left[\left\langle\chi(T)\right|\left.\psi(T)\right\rangle - \left\langle\chi(0)\right|\left.\psi(0)\right\rangle - \int_0^T \left\langle \left(\frac{d}{dt} + i\hat{\mathbf{H}}^{\dagger}(t)\right)\chi(t) \right| \psi(t) \right\rangle dt\right] \tag{B25}$$

We start from condition (B20). First, we shall derive the equation for a simpler problem, in which the boundary condition constraints on the field (Eqs. (B7), (B9)) are excluded. Then, we shall derive the equation for the full problem, which is conveniently expressed in the terms of the simpler problem expression for the field.

In the unconstrained problem, the $J_{\epsilon(0)}$ and $J_{\epsilon(T)}$ terms (Eqs. (B8), (B10), respectively) are excluded

from J. The LHS of Eq. (B20) becomes:

$$\frac{\delta J}{\delta \bar{\epsilon}(\omega)} = \frac{\delta J_{energy}}{\delta \bar{\epsilon}(\omega)} + \frac{\delta J_{Schr}}{\delta \bar{\epsilon}(\omega)} \tag{B26}$$

where

$$\frac{\delta J_{energy}}{\delta \bar{\epsilon}(\omega)} = -\frac{2\bar{\epsilon}(\omega)}{\tilde{f}_\epsilon(\omega)} \tag{B27}$$

$$\begin{aligned}
\frac{\delta J_{Schr}}{\delta \bar{\epsilon}(\omega)} &= 2\ \mathrm{Re}\left[-i\int_0^T \left\langle \chi(t) \left| \frac{\delta \hat{\mathbf{H}}(t)}{\delta \bar{\epsilon}(\omega)} \right| \psi(t) \right\rangle\, dt\right] \\
&= -2\ \mathrm{Im}\left[\sqrt{\frac{2}{\pi}}\int_0^T \left\langle \chi(t) \left| \hat{\mathbf{X}} \right| \psi(t) \right\rangle \cos(\omega t)\, dt\right] \\
&= -2\ \mathrm{Im}\left\{\mathcal{C}\left[\left\langle \chi(t) \left| \hat{\mathbf{X}} \right| \psi(t) \right\rangle\right]\right\} \\
&= 2\,\mathcal{C}\left[-\mathrm{Im}\left\langle \chi(t) \left| \hat{\mathbf{X}} \right| \psi(t) \right\rangle\right]
\end{aligned} \tag{B28}$$

The resulting gradient expression is

$$\frac{\delta J}{\delta \bar{\epsilon}(\omega)} = 2\left(-\frac{\bar{\epsilon}(\omega)}{\tilde{f}_\epsilon(\omega)} + \mathcal{C}\left[-\mathrm{Im}\left\langle \chi(t) \left| \hat{\mathbf{X}} \right| \psi(t) \right\rangle\right]\right) \tag{B29}$$

The extremum condition (B20) yields the following expression for $\bar{\epsilon}(\omega)$:

$$\bar{\epsilon}(\omega) = \tilde{f}_\epsilon(\omega)\mathcal{C}\left[-\mathrm{Im}\left\langle \chi(t) \left| \hat{\mathbf{X}} \right| \psi(t) \right\rangle\right] \tag{B30}$$

The field in the time-domain is

$$\epsilon(t) = \mathcal{C}^{-1}\left\{\tilde{f}_\epsilon(\omega)\mathcal{C}\left[-\mathrm{Im}\left\langle \chi(t) \left| \hat{\mathbf{X}} \right| \psi(t) \right\rangle\right]\right\} \tag{B31}$$

Now we turn to the solution of the full constrained problem. The LHS of Eq. (B20) is modified to

$$\frac{\delta J}{\delta \bar{\epsilon}(\omega)} = \frac{\delta J_{energy}}{\delta \bar{\epsilon}(\omega)} + \frac{\delta J_{\epsilon(0)}}{\delta \bar{\epsilon}(\omega)} + \frac{\delta J_{\epsilon(T)}}{\delta \bar{\epsilon}(\omega)} + \frac{\delta J_{Schr}}{\delta \bar{\epsilon}(\omega)} \tag{B32}$$

where

$$\frac{\delta J_{\epsilon(0)}}{\delta \bar{\epsilon}(\omega)} = -2\lambda_0 \tag{B33}$$

$$\frac{\delta J_{\epsilon(T)}}{\delta \bar{\epsilon}(\omega)} = -2\lambda_T \cos(\omega T) \tag{B34}$$

The gradient expression is modified to

$$\frac{\delta J}{\delta \bar{\epsilon}(\omega)} = 2\left(-\frac{\bar{\epsilon}(\omega)}{\tilde{f}_\epsilon(\omega)} + \mathcal{C}\left[-\operatorname{Im}\left\langle \chi(t)\left|\hat{\mathbf{X}}\right|\psi(t)\right\rangle\right] - \lambda_0 - \lambda_T\cos(\omega T)\right) \tag{B35}$$

Eq. (B20) yields the following modified expression for $\bar{\epsilon}(\omega)$:

$$\bar{\epsilon}(\omega) = \tilde{f}_\epsilon(\omega)\mathcal{C}\left[-\operatorname{Im}\left\langle \chi(t)\left|\hat{\mathbf{X}}\right|\psi(t)\right\rangle\right] - \tilde{f}_\epsilon(\omega)[\lambda_0 + \lambda_T\cos(\omega T)] \tag{B36}$$

The first term in the RHS of Eq. (B36) is recognized as the same expression as $\bar{\epsilon}(\omega)$ in the unconstrained problem, Eq. (B30). Let us denote:

$$\bar{\epsilon}_{unc}(\omega) \equiv \tilde{f}_\epsilon(\omega)\mathcal{C}\left[-\operatorname{Im}\left\langle \chi(t)\left|\hat{\mathbf{X}}\right|\psi(t)\right\rangle\right] \tag{B37}$$

We can rewrite the expression to $\bar{\epsilon}(\omega)$ in the constrained problem as

$$\bar{\epsilon}(\omega) = \bar{\epsilon}_{unc}(\omega) - \tilde{f}_\epsilon(\omega)[\lambda_0 + \lambda_T\cos(\omega T)] \tag{B38}$$

Now, in order to find explicit expressions to λ_0 and λ_T, we should enforce the boundary constraints, (B7), (B9), on the solution (B38). We start from the constraint (B7). It can be rewritten in the terms of $\bar{\epsilon}(\omega)$ as follows:

$$\epsilon(0) = \sqrt{\frac{2}{\pi}}\int_0^\Omega \bar{\epsilon}(\omega)\cos(0)\,d\omega = \sqrt{\frac{2}{\pi}}\int_0^\Omega \bar{\epsilon}(\omega)\,d\omega = 0 \tag{B39}$$

Let us substitute the solution (B38) in the constraint equation:

$$\epsilon_{unc}(0) - \left(\lambda_0\,\mathcal{C}^{-1}\left[\tilde{f}_\epsilon(\omega)\right]\Big|_{t=0} + \lambda_T\,\mathcal{C}^{-1}\left[\tilde{f}_\epsilon(\omega)\right]\Big|_{t=T}\right) = 0 \tag{B40}$$

The constraint (B9) can be written in the terms of $\bar{\epsilon}(\omega)$ as follows:

$$\epsilon(T) = \sqrt{\frac{2}{\pi}}\int_0^\Omega \bar{\epsilon}(\omega)\cos(\omega T)\,d\omega = 0 \tag{B41}$$

Substituting Eq. (B38) in Eq. (B41) we obtain:

$$\epsilon_{unc}(T) - \left(\lambda_0 \mathcal{C}^{-1} \left[\tilde{f}_\epsilon(\omega) \right] \Big|_{t=T} + \lambda_T \mathcal{C}^{-1} \left[\tilde{f}_\epsilon(\omega) \cos(\omega T) \right] \Big|_{t=T} \right) = 0 \tag{B42}$$

Eqs. (B40), (B42), constitute a two equation system of the two variables λ_0, λ_T. Let us denote, for convenience:

$$a \equiv \mathcal{C}^{-1} \left[\tilde{f}_\epsilon(\omega) \right] \Big|_{t=0} \tag{B43}$$

$$b \equiv \mathcal{C}^{-1} \left[\tilde{f}_\epsilon(\omega) \right] \Big|_{t=T} \tag{B44}$$

$$d \equiv \mathcal{C}^{-1} \left[\tilde{f}_\epsilon(\omega) \cos(\omega T) \right] \Big|_{t=T} \tag{B45}$$

The system of equations can be written in a matrix-vector form in the following way:

$$\begin{bmatrix} a & b \\ b & d \end{bmatrix} \begin{bmatrix} \lambda_0 \\ \lambda_T \end{bmatrix} = \begin{bmatrix} \epsilon_{unc}(0) \\ \epsilon_{unc}(T) \end{bmatrix} \tag{B46}$$

Let us define:

$$M \equiv \begin{bmatrix} a & b \\ b & d \end{bmatrix} \tag{B47}$$

The solution of this system is

$$\lambda_0 = \frac{\epsilon_{unc}(0)d - \epsilon_{unc}(T)b}{\det M} \tag{B48}$$

$$\lambda_T = \frac{\epsilon_{unc}(T)a - \epsilon_{unc}(0)b}{\det M} \tag{B49}$$

where $\det M = ad - b^2$ is the determinant of M.

Note that the derivation of Eqs. (B38), (B48), (B49), does not require any knowledge of the forms of the other terms in the functional (J_{max} and J_{ion}), which do not have an explicit dependence on $\bar{\epsilon}(\omega)$. Thus, the same technique of imposing frequency restrictions and boundary constraints can be used directly for various control problems, without altering the expression of the field.

Eq. (B38) can be utilized to produce a field which satisfies the boundary condition constraints, (B7), (B9), from an arbitrary unconstrained field. The given unconstrained field can be substituted in Eq. (B38) as $\bar{\epsilon}_{unc}(\omega)$, even though it was not produced by Eq. (B37). The derivation of the expressions

of the λ's does not rely on any previous knowledge on $\bar{\epsilon}_{unc}(\omega)$. Thus, it holds in general for any unconstrained field. If, furthermore, the given $\bar{\epsilon}_{unc}(\omega)$ is restricted to the required frequency region which is represented by $\tilde{f}_\epsilon(\omega)$, then $\bar{\epsilon}(\omega)$ becomes also restricted to the required frequency region. This technique can be used to produce a guess field which satisfies both the boundary conditions and the frequency requirements.

We proceed to the condition (B21). Let us derive the LHS of this equation:

$$\frac{\delta J}{\delta |\psi(t)\rangle} = \frac{\delta J_{max}}{\delta |\psi(t)\rangle} + \frac{\delta J_{Schr}}{\delta |\psi(t)\rangle} \tag{B50}$$

$$\begin{aligned}\frac{\delta J_{max}}{\delta |\psi(t)\rangle} &= \frac{1}{2}\int_0^\Omega f_C(\omega) \frac{d\overline{\left\langle \hat{\mathbf{C}} \right\rangle}^2(\omega)}{d\overline{\left\langle \hat{\mathbf{C}} \right\rangle}(\omega)} \frac{\delta\overline{\left\langle \hat{\mathbf{C}} \right\rangle}(\omega)}{\delta \left\langle \hat{\mathbf{C}} \right\rangle(t)} \frac{\delta \left\langle \hat{\mathbf{C}} \right\rangle(t)}{\delta |\psi(t)\rangle} d\omega \\ &= \sqrt{\frac{2}{\pi}} \int_0^\Omega f_C(\omega) \overline{\left\langle \hat{\mathbf{C}} \right\rangle}(\omega) \cos(\omega t)\, d\omega \, \langle\psi(t)| \hat{\mathbf{C}} \\ &= \mathcal{C}^{-1}\left[f_C(\omega) \overline{\left\langle \hat{\mathbf{C}} \right\rangle}(\omega) \right] \langle\psi(t)| \hat{\mathbf{C}}\end{aligned} \tag{B51}$$

$$\frac{\delta J_{Schr}}{\delta |\psi(t)\rangle} = \frac{d\langle\chi(t)|}{dt} + \left\langle i\hat{\mathbf{H}}^\dagger(t)\chi(t) \right| \tag{B52}$$

Eq. (B21) becomes

$$\frac{d\langle\chi(t)|}{dt} = -\left\langle i\hat{\mathbf{H}}^\dagger(t)\chi(t) \right| - \langle\psi(t)|\, \mathcal{C}^{-1}\left[f_C(\omega) \overline{\left\langle \hat{\mathbf{C}} \right\rangle}(\omega) \right] \hat{\mathbf{C}} \tag{B53}$$

Following analogous steps, the condition (B22) yields:

$$\frac{d|\chi(t)\rangle}{dt} = -i\hat{\mathbf{H}}^\dagger(t)\,|\chi(t)\rangle - \mathcal{C}^{-1}\left[f_C(\omega) \overline{\left\langle \hat{\mathbf{C}} \right\rangle}(\omega) \right] \hat{\mathbf{C}}\,|\psi(t)\rangle \tag{B54}$$

It can be seen that $\langle\chi(t)|$ is subject to the adjoint evolution equation of $|\chi(t)\rangle$.

Let us derive the LHS of the condition (B23):

$$\frac{\delta J}{\delta |\psi(T)\rangle} = \frac{\delta J_{ion}}{\delta |\psi(T)\rangle} + \frac{\delta J_{Schr}}{\delta |\psi(T)\rangle} \tag{B55}$$

$$\begin{aligned}\frac{\delta J_{ion}}{\delta |\psi(T)\rangle} &= \frac{d\left[\sigma\left(\langle\psi(T)|\psi(T)\rangle\right)\right]}{d\langle\psi(T)|\psi(T)\rangle} \frac{\delta\langle\psi(T)|\psi(T)\rangle}{\delta|\psi(T)\rangle} \\ &= \sigma'\left(\langle\psi(T)|\psi(T)\rangle\right)\langle\psi(T)|\end{aligned} \tag{B56}$$

$$\frac{\delta J_{Schr}}{\delta |\psi(T)\rangle} = -\langle\chi(T)| \tag{B57}$$

Eq. (B23) becomes

$$\langle\chi(T)| = \sigma'\left(\langle\psi(T)|\psi(T)\rangle\right)\langle\psi(T)| \tag{B58}$$

Following analogous steps, the condition (B24) yields:

$$|\chi(T)\rangle = \sigma'\left(\langle\psi(T)|\psi(T)\rangle\right)|\psi(T)\rangle \tag{B59}$$

From Eqs. (B58), (B59), we have:

$$\langle\chi(T)| = |\chi(T)\rangle^\dagger \tag{B60}$$

The evolution equations (B53), (B54), together with Eq. (B60), lead conclusively to Eq. (B18), as expected. Assuming this, only Eqs. (B54), (B59) are required to perform the computations in practice.

We collect the equations which define $\bar{\epsilon}(\omega)$, $|\psi(t)\rangle$ and $|\chi(t)\rangle$, i. e. Eqs. (B11), (B14), (B54), (B59), (B12), (B38), (B37), (B48), (B49), (B43), (B44), (B45):

$$\begin{aligned}
\frac{d|\psi(t)\rangle}{dt} &= -i\hat{\mathbf{H}}(t)|\psi(t)\rangle\,, \\
|\psi(0)\rangle &= |\psi_0\rangle && \text{(B61)}\\
\frac{d|\chi(t)\rangle}{dt} &= -i\hat{\mathbf{H}}^\dagger(t)|\chi(t)\rangle - \mathcal{C}^{-1}\left[f_C(\omega)\overline{\left\langle\hat{\mathbf{C}}\right\rangle}(\omega)\right]\hat{\mathbf{C}}|\psi(t)\rangle\,, \\
|\chi(T)\rangle &= \sigma'\left(\langle\psi(T)|\psi(T)\rangle\right)|\psi(T)\rangle && \text{(B62)}\\
\hat{\mathbf{H}}(t) &= \hat{\mathbf{H}}_0 - \hat{\mathbf{X}}\epsilon(t) && \text{(B63)}\\
\epsilon(t) &= \mathcal{C}^{-1}[\bar{\epsilon}(\omega)] && \text{(B64)}\\
\bar{\epsilon}(\omega) &= \bar{\epsilon}_{unc}(\omega) - \tilde{f}_\epsilon(\omega)[\lambda_0 + \lambda_T\cos(\omega T)] && \text{(B65)}\\
\bar{\epsilon}_{unc}(\omega) &\equiv \tilde{f}_\epsilon(\omega)\mathcal{C}\left[-\operatorname{Im}\left\langle\chi(t)\left|\hat{\mathbf{X}}\right|\psi(t)\right\rangle\right] && \text{(B66)}\\
\epsilon_{unc}(t) &= \mathcal{C}^{-1}[\bar{\epsilon}_{unc}(\omega)] && \text{(B67)}\\
\lambda_0 &= \frac{\epsilon_{unc}(0)d - \epsilon_{unc}(T)b}{ad - b^2} && \text{(B68)}\\
\lambda_T &= \frac{\epsilon_{unc}(T)a - \epsilon_{unc}(0)b}{ad - b^2} && \text{(B69)}\\
a &\equiv \mathcal{C}^{-1}\left[\tilde{f}_\epsilon(\omega)\right]\Big|_{t=0} && \text{(B70)}\\
b &\equiv \mathcal{C}^{-1}\left[\tilde{f}_\epsilon(\omega)\right]\Big|_{t=T} && \text{(B71)}\\
d &\equiv \mathcal{C}^{-1}\left[\tilde{f}_\epsilon(\omega)\cos(\omega T)\right]\Big|_{t=T} && \text{(B72)}
\end{aligned}$$

These constitute the Euler-Lagrange equations of the problem.

Appendix C: Numerical details

1. Discretization of the cosine transform

In practice, the cosine transform (1) is replaced by a discrete-cosine-transform (DCT). There are several types of DCT's. We used a boundary including DCT, which is sometimes referred to as the *DCT of the first type* (DCT-1). The DCT-1 of a time-dependent function, $g(t)$, sampled at $N_t + 1$ equidistant time points,

$$t_k, \qquad k = 0, 1, \ldots, N_t$$

is defined as:

$$\bar{g}(\omega_j) = \sqrt{\frac{2}{N_t}} \sum_{k=0}^{N_t} \frac{1}{h_k} g(t_k) \cos\left(\frac{jk\pi}{N_t}\right), \qquad j = 0, 1, \ldots, N_t \tag{C1}$$

$$h_k = \begin{cases} 2 & k = 0 \text{ or } k = N_t \\ 1 & 1 \leq k \leq N_t - 1 \end{cases} \tag{C2}$$

The inverse transform is defined by

$$g(t_k) = \sqrt{\frac{2}{N_t}} \sum_{j=0}^{N_t} \frac{1}{h_j} \bar{g}(\omega_j) \cos\left(\frac{jk\pi}{N_t}\right), \qquad k = 0, 1, \ldots, N_t \tag{C3}$$

It can be observed from Eqs. (C1), (C3) that DCT-1 is its own inverse.

The argument of the cosine function in Eqs. (C1), (C3) is equivalent to a discretized variant of the continuous argument ωt in the continuous transform (1), where the variables are discretized as follows:

$$t_k = k\frac{T}{N_t} \qquad k = 0, 1, \ldots, N_t \tag{C4}$$

$$\omega_j = j\frac{\pi}{T} \qquad j = 0, 1, \ldots, N_t \tag{C5}$$

In order to be consistent with the continuous formulation, the direct transform was multiplied by the factor $T/\sqrt{N_t \pi}$:

$$\bar{g}(\omega_j) = \sqrt{\frac{2}{\pi}} \frac{T}{N_t} \sum_{k=0}^{N_t} \frac{1}{h_k} g(t_k) \cos\left(\frac{jk\pi}{N_t}\right) \tag{C6}$$

where T/N_t replaces dt in the continuous integral form (Cf. Eq. (C4)). The inverse transform was divided by the same factor:

$$\begin{aligned} g(t_k) &= \frac{\sqrt{2\pi}}{T}\sum_{j=0}^{N_t}\frac{1}{h_j}\bar{g}(\omega_j)\cos\left(\frac{jk\pi}{N_t}\right) \\ &= \sqrt{\frac{2}{\pi}}\frac{\pi}{T}\sum_{j=0}^{N_t}\frac{1}{h_j}\bar{g}(\omega_j)\cos\left(\frac{jk\pi}{N_t}\right) \end{aligned} \tag{C7}$$

where π/T replaces $d\omega$ in the continuous integral form (Cf. Eq. (C5)). The consistency with the continuous formulation has an important advantage: When using Eq. (C1) as is, the definition of the spectral function $\bar{g}(\omega)$, represented by the transform, varies with the sampling frequency. The consistency with the continuous form makes $\bar{g}(\omega)$ independent of the sampling.

The choice of this type of DCT is important for the following reason: The boundary constraints, (B7), (B9), are imposed at the boundaries of the time-domain, i. e. at $t = 0$ and $t = T$. Consequently, $\epsilon_{unc}(t)$ and the inverse cosine transform in Eqs. (B40), (B42), are evaluated at the time-domain boundaries. Therefore, it is important that the discrete transform will include both boundaries. Moreover, the Euler-Lagrange equations define a *two-point boundary value problem*: $|\psi(t)\rangle$ is propagated from the initial condition at $t = 0$, and $|\chi(t)\rangle$ is backward propagated from the final condition at $t = T$. Thus, the boundary including DCT is compatible with the control equations.

In this context, we shall give a practical remark on the computation of d in Eqs. (B48), (B49). d is computed in the DCT-1 form as

$$\begin{aligned} d &= \frac{\sqrt{2\pi}}{T}\sum_{j=0}^{N_t}\frac{1}{h_j}\tilde{f}_\epsilon(\omega_j)\cos^2(\omega_j T) \qquad\qquad \omega_j = j\frac{\pi}{T} \\ &= \frac{\sqrt{2\pi}}{T}\sum_{j=0}^{N_t}\frac{1}{h_j}\tilde{f}_\epsilon\left(\frac{j\pi}{T}\right)\cos^2(j\pi) \\ &= \frac{\sqrt{2\pi}}{T}\sum_{j=0}^{N_t}\frac{1}{h_j}\tilde{f}_\epsilon\left(\frac{j\pi}{T}\right)(-1)^2 \\ &= \frac{\sqrt{2\pi}}{T}\sum_{j=0}^{N_t}\frac{1}{h_j}\tilde{f}_\epsilon\left(\frac{j\pi}{T}\right) \end{aligned} \tag{C8}$$

which is equivalent to the discretized expression to a. Thus, Eq. (B48) can be computed as

$$\lambda_0 = \frac{\epsilon_{unc}(0)a - \epsilon_{unc}(T)b}{\det M} \tag{C9}$$

where $\det M$ can be computed as

$$\det M = a^2 - b^2 \tag{C10}$$

The $\det M$ in Eq. (B49) is also computed by Eq. (C10).

2. Details of the optimization procedure

As was mentioned in Sec. III, the optimization procedure is based on the BFGS method (see, for example, [54, Chapter 3]). The BFGS method is a universal method of optimization which is well established and widespread. However, several details are crucial for a successful implementation of the method in the present context. Commercial programs might completely fail when they are used without certain modifications.

a. Discretization of the optimization space

The optimization is performed in a discrete variable space. Thus, we first have to reformulate the problem by discrete means in order to obtain the necessary expressions for the optimization process.

The optimized function, $\bar{\epsilon}(\omega)$, is discretized as follows:

$$\bar{\epsilon}(\omega) \longrightarrow \vec{\bar{\boldsymbol{\epsilon}}} \equiv \begin{bmatrix} \bar{\epsilon}(\omega_0) \\ \bar{\epsilon}(\omega_1) \\ \vdots \\ \bar{\epsilon}(\omega_{N_t}) \end{bmatrix} \qquad \omega_j = j\frac{\pi}{T} \tag{C11}$$

The ω sampling is the same as the DCT sampling (Cf. Eq. (C5)). The functional has to be expressed by the terms of the discrete optimization space:

$$J\left[\bar{\epsilon}(\omega)\right] \longrightarrow J^d\left[\vec{\bar{\boldsymbol{\epsilon}}}\right] \tag{C12}$$

where J^d is a discretized variant of J. The $\bar{\epsilon}(\omega)$ dependence in J is reformulated by discrete means; J_{energy} is replaced by

$$J^d_{energy} = -\sum_{j=0}^{N_t} \frac{1}{\tilde{f}_\epsilon(\omega_j)} \bar{\epsilon}^2(\omega_j) w_j, \qquad w_j = \frac{1}{h_j}\frac{\pi}{T} \tag{C13}$$

where the h_j's are defined by Eq. (C2). The w_j's are the *integration weights*. The $\bar{\epsilon}(\omega)$ dependence of

$\epsilon(t)$ in $J_{\epsilon(0)}$, $J_{\epsilon(T)}$ and J_{Schr} is expressed by the inverse DCT expression (Cf. Eq. (C7)):

$$\begin{aligned}\epsilon(t) &= \sqrt{\frac{2}{\pi}}\frac{\pi}{T}\sum_{j=0}^{N_t}\frac{1}{h_j}\bar{\epsilon}(\omega_j)\cos(\omega_j t) \\ &= \sqrt{\frac{2}{\pi}}\sum_{j=0}^{N_t}\bar{\epsilon}(\omega_j)\cos(\omega_j t)\, w_j\end{aligned} \tag{C14}$$

The resulting gradient expression is

$$\frac{dJ^d}{d\bar{\epsilon}(\omega_j)} = w_j \left.\frac{\delta J}{\delta\bar{\epsilon}(\omega)}\right|_{\omega=\omega_j} \tag{C15}$$

where $\frac{\delta J}{\delta\bar{\epsilon}(\omega)}$ is given by Eq. (B35).

b. *Definition of the numerical optimization problem*

Typically, the space of $\vec{\boldsymbol{\epsilon}}$ is much larger than what is required in practice. The reason lies in the form of Eq. (B36)—it can be easily seen that $\bar{\epsilon}(\omega)$ attains negligible values in the regions in which the filter function $f_\epsilon(\omega)$ is negligible. Hence, we can neglect the terms of $\vec{\boldsymbol{\epsilon}}$ in these regions and fix them to zero in advance. Only the terms in the non-negligible regions participate in the optimization process. This results in a dramatic decrease of the dimension of the optimization space. This practice is important also for another reason—it solves the problem of division by zero in the first term in the RHS of Eq. (B35).

Let us denote the vector of the ω_j values which participate in the optimization by $\vec{\boldsymbol{\omega}}_r$. In our calculations, we have chosen $\vec{\boldsymbol{\omega}}_r$ as to satisfy the following condition:

$$f_\epsilon(\vec{\boldsymbol{\omega}}_r) \geq 2.22\times10^{-16} \tag{C16}$$

The value in the RHS represents the machine precision of double type variables.

For convenience, let us define the solution vector in the reduced space:

$$\vec{\boldsymbol{\epsilon}}_r \equiv \bar{\epsilon}(\vec{\boldsymbol{\omega}}_r) \tag{C17}$$

We naturally formulated our optimization problem as a *maximization* problem. This is also the common convention in quantum OCT texts for the vast majority of the optimization problems. However, the BFGS method is traditionally formulated as a *minimization* process, which is the common convention

in the optimization literature in general. In order to apply the BFGS method to our problem, we have to use a uniform formulation. There exist two options:

1. The BFGS method can be reformulated as a maximization process.

2. We can perform a minimization of $-J^d$ instead of the maximization of J^d.

We adopt the second option, in order to prevent confusion, and to enable a direct use of existing BFGS codes.

In summary, our optimization problem is defined as the minimization of the objective $-J^d$ with respect to the solution vector $\vec{\bar{\boldsymbol{\epsilon}}}_r$.

We denote the gradient vector of the *minimization problem* as $\vec{\mathbf{g}}$. Its general term is given by

$$\begin{aligned} -\frac{dJ^d}{d\bar{\epsilon}(\omega_j)} &= -w_j \left.\frac{\delta J}{\delta\bar{\epsilon}(\omega)}\right|_{\omega=\omega_j} \\ &= 2w_j \left(\frac{\bar{\epsilon}(\omega_j)}{\tilde{f}_\epsilon(\omega_j)} - \mathcal{C}\left[-\operatorname{Im}\left\langle \chi(t) \left| \hat{\mathbf{X}} \right| \psi(t) \right\rangle \right] \Big|_{\omega=\omega_j} + \lambda_0 + \lambda_T \cos(\omega_j T) \right) \end{aligned} \tag{C18}$$

$\vec{\mathbf{g}}$ can be expressed by a vector notation in the following way:

$$\vec{\mathbf{g}} = 2\vec{\mathbf{w}} \circ \left(\vec{\bar{\boldsymbol{\epsilon}}}_r \oslash \tilde{f}_\epsilon(\vec{\boldsymbol{\omega}}_r) - \mathcal{C}\left[-\operatorname{Im}\left\langle \chi(t) \left| \hat{\mathbf{X}} \right| \psi(t) \right\rangle \right] \Big|_{\vec{\boldsymbol{\omega}}=\vec{\boldsymbol{\omega}}_r} + \lambda_0 + \lambda_T \cos(\vec{\boldsymbol{\omega}}_r T) \right) \tag{C19}$$

where $\vec{\mathbf{w}}$ is the weight vector. $\circ$ denotes the Hadamard product (elementwise product) and $\oslash$ the Hadamard division.

The approximated Hessian matrix will be denoted by S.

c. *Iterative formulation*

The optimization process is mathematically described by an iterative *update rule* for the solution vector $\vec{\bar{\boldsymbol{\epsilon}}}_r$. This requires the formulation the optimization equations by iterative means.

First, we have to index the iteration number of the mathematical objects involved in the iterative process. The iteration index will be denoted by a superscript in parentheses. For example, the solution vector in the k'th iteration is denoted by $\vec{\bar{\boldsymbol{\epsilon}}}_r^{(k)}$. $\vec{\bar{\boldsymbol{\epsilon}}}_r^{(0)}$ denotes the initial guess solution.

$\epsilon^{(k)}(t)$ is given by the substitution of $\vec{\bar{\boldsymbol{\epsilon}}}_r^{(k)}$ into Eq. (C14). $\left|\psi^{(k)}(t)\right\rangle$ and $\left|\chi^{(k)}(t)\right\rangle$ are given by Eqs. (B61)–(B63) with the substitution of $\epsilon^{(k)}(t)$ into the Hamiltonian (Eq. (B63)). $\lambda_0^{(k)}$ and $\lambda_T^{(k)}$ are given by Eqs. (B66)–(B72), where Eq. (B66) is computed by the substitution of $\left|\psi^{(k)}(t)\right\rangle$ and $\left|\chi^{(k)}(t)\right\rangle$.

The objective of the minimization process in the k'th iteration is given by $-J^{d\,(k)}$, where $J^{d\,(k)}$ is computed by $\vec{\bar{\boldsymbol{\epsilon}}}_r^{(k)}$ and $\left|\psi^{(k)}(t)\right\rangle$. Note that the Lagrange-multiplier terms, $J_{\epsilon(0)}$, $J_{\epsilon(T)}$ and J_{Schr}, are identically zero; thus, the Lagrange-multipliers do not participate in the computation of $J^{d\,(k)}$.

The gradient in the k'th iteration is given by the following expression (Cf. Eq. (C19)):

$$\vec{\mathbf{g}}^{(k)} = 2\vec{\mathbf{w}} \circ \left(\vec{\bar{\boldsymbol{\epsilon}}}_r^{(k)} \oslash \tilde{f}_\epsilon(\vec{\boldsymbol{\omega}}_r) - \mathcal{C}\left[-\operatorname{Im}\left\langle \chi^{(k)}(t) \left| \hat{\mathbf{X}} \right| \psi^{(k)}(t) \right\rangle \right] \Big|_{\vec{\boldsymbol{\omega}}=\vec{\boldsymbol{\omega}}_r} + \lambda_0^{(k)} + \lambda_T^{(k)} \cos(\vec{\boldsymbol{\omega}}_r T) \right) \tag{C20}$$

Let us define:

$$\vec{\boldsymbol{\delta}}^{(k)} \equiv \vec{\bar{\boldsymbol{\epsilon}}}_r^{(k+1)} - \vec{\bar{\boldsymbol{\epsilon}}}_r^{(k)} \tag{C21}$$

$$\vec{\boldsymbol{\gamma}}^{(k)} \equiv \vec{\mathbf{g}}^{(k+1)} - \vec{\mathbf{g}}^{(k)} \tag{C22}$$

The BFGS update rule for the approximated inverse-Hessian is

$$S^{(k+1)^{-1}} = S^{(k)^{-1}} + \left(1 + \frac{\vec{\boldsymbol{\gamma}}^{(k)\,T} S^{(k)^{-1}} \vec{\boldsymbol{\gamma}}^{(k)}}{\vec{\boldsymbol{\delta}}^{(k)\,T} \vec{\boldsymbol{\gamma}}^{(k)}} \right) \frac{\vec{\boldsymbol{\delta}}^{(k)} \vec{\boldsymbol{\delta}}^{(k)\,T}}{\vec{\boldsymbol{\delta}}^{(k)\,T} \vec{\boldsymbol{\gamma}}^{(k)}} - \frac{\vec{\boldsymbol{\delta}}^{(k)} \vec{\boldsymbol{\gamma}}^{(k)\,T} S^{(k)^{-1}} + S^{(k)^{-1}} \vec{\boldsymbol{\gamma}}^{(k)} \vec{\boldsymbol{\delta}}^{(k)\,T}}{\vec{\boldsymbol{\delta}}^{(k)\,T} \vec{\boldsymbol{\gamma}}^{(k)}} \tag{C23}$$

The initial inverse-Hessian guess, $S^{(0)^{-1}}$, is supplied by the user.

The direction of search in the k'th iteration will be denoted by $\vec{\mathbf{p}}^{(k)}$. It is given by the general quasi-Newton expression:

$$\vec{\mathbf{p}}^{(k)} = -S^{(k)^{-1}} \vec{\mathbf{g}}^{(k)} \tag{C24}$$

The update rule for the solution vector is

$$\vec{\bar{\boldsymbol{\epsilon}}}_r^{(k+1)} = \vec{\bar{\boldsymbol{\epsilon}}}_r^{(k)} + \kappa^{(k)} \vec{\mathbf{p}}^{(k)}, \qquad \kappa^{(k)} > 0 \tag{C25}$$

where $\kappa^{(k)}$ is determined by the *line-search* procedure. The line-search requires the knowledge of $\vec{\bar{\boldsymbol{\epsilon}}}_r^{(k)}$, $\vec{\mathbf{p}}^{(k)}$, $-J^{d\,(k)}$ and $\vec{\mathbf{g}}^{(k)}$. It also requires a procedure which returns $-J^d$ and $\vec{\mathbf{g}}$ for any solution vector $\vec{\bar{\boldsymbol{\epsilon}}}_r$ which is a candidate for $\vec{\bar{\boldsymbol{\epsilon}}}_r^{(k+1)}$.

Note that the only Euler-Lagrange equation which does not participate in the iterative calculation is the equation for the solution $\bar{\epsilon}(\omega)$ (B65). Nevertheless, the solution equation has been utilized in the considerations which led to the reduction of the optimization space (Sec. C 2 b). In what follows, it turns out that the solution equation form is actually contained implicitly in the optimization procedure.

The termination condition of the iterative process is chosen as

$$\frac{\left\| \vec{\vec{\epsilon}}_r^{(k)} - \vec{\vec{\epsilon}}_r^{(k-1)} \right\|}{\left\| \vec{\vec{\epsilon}}_r^{(k)} \right\|} \leq 10^{-4} \tag{C26}$$

d. Initial guess for the solution vector and the Hessian

The update equation (C25) presents a problem: The RHS consists of two terms. The first is the previous solution $\vec{\vec{\epsilon}}_r^{(k)}$, which is assumed to satisfy the boundary constraints, (B7) and (B9). However, the second term $-\kappa^{(k)} {S^{(k)}}^{-1} \vec{\mathbf{g}}^{(k)}$ does not satisfy the boundary constraints in general. As a result, the update rule does not conserve the boundary constraints.

Another property which is not conserved by the update rule is the restriction of the driving field spectrum. The direction of search $\vec{\mathbf{p}}^{(k)}$ is not confined, in general, to the allowed frequency region. As an illustration of this statement, let us take $S^{(k)} = I$, where I is the identity matrix. The identity matrix is usually chosen as the initial guess for the Hessian in the BFGS process. With this choice, the direction of search becomes $\vec{\mathbf{p}}^{(k)} = -\vec{\mathbf{g}}^{(k)}$, as in the first order gradient method. It can be readily observed from Eq. (C18) that the gradient direction does not satisfy the spectral restriction. The terms λ_0, $\lambda_T \cos(\omega_j T)$ obviously have significant values throughout the spectrum of ω_j values. Actually, the different terms in parenthesis in Eq. (C18) can be recognized as the same terms as in both sides of Eq. (B36), *divided by* $\tilde{f}_\epsilon(\omega)$. The division by the filter function simply *cancels the filtration* which is present in Eq. (B36). Thus, the gradient expression is not restricted to the required spectral region, represented by $f_\epsilon(\omega)$.

When $\vec{\mathbf{p}}^{(k)}$ is not spectrally restricted, the optimization process cannot proceed; very large penalties are put on the "forbidden" frequency regions, and thus the result cannot improve unless $\kappa^{(k)}$ is extremely small. Consequently, the optimization process becomes highly ineffective and impractical. Moreover, the update of $\vec{\vec{\epsilon}}_r$ is often so small that it does not change beyond the roundoff error regime.

Note that the conservation problems of the update rule originate from the form (C25), which determines $\vec{\vec{\epsilon}}_r$ independently of the Euler-Lagrange equation for the solution, (B65). These problems do not exist in optimization methods in which the update rule is based directly on the solution equation. An example of a method of this class is the relaxation process proposed in our previous work (Ref. [43, Chapter 3.2.3], [44]). An update rule of this type satisfies automatically the boundary and spectral restrictions which are present in Eq. (B65).

The conservation problems in the BFGS process can be solved by an appropriate choice of the *initial guess solution* $\vec{\vec{\epsilon}}_r^{(0)}$ and the *initial Hessian* $S^{(0)}$. First, $\vec{\vec{\epsilon}}_r^{(0)}$ should satisfy the boundary conditions and

the spectral restrictions. A guess solution which satisfies both requirements can be constructed by the following steps:

1. Choose a spectral function $\bar{\epsilon}^{(0)}_{unc}(\omega)$ which satisfies the spectral restrictions, but unnecessarily the boundary constraints.

2. Obtain a constrained spectral function $\bar{\epsilon}^{(0)}(\omega)$ by substituting $\bar{\epsilon}^{(0)}_{unc}(\omega)$ into Eqs. (B65), (B67)–(B72) (the justification to this practice is explained in Appendix B).

3. $\vec{\bar{\boldsymbol{\epsilon}}}^{(0)}_r$ is simply given by the terms of $\bar{\epsilon}^{(0)}(\omega)$ which participate in the optimization procedure.

Note that if $\bar{\epsilon}^{(0)}_{unc}(\omega) \propto \tilde{f}_\epsilon(\omega)$, then Eq. (B65) yields $\bar{\epsilon}^{(0)}(\omega) \equiv 0$. This initial field is not very useful—there seems to be a local minimum or a saddle point in the zero field solution. Consequently, the optimization procedure cannot proceed from this point in the optimization space.

$S^{(0)}$ should be chosen as the Hessian of $-J^d_{energy}$ (see Eq. (C13)). The general term of the Hessian is given by

$$-\frac{\partial^2 J^d_{energy}}{\partial\bar{\epsilon}(\omega_j)\partial\bar{\epsilon}(\omega_l)} = \delta_{jl}\frac{2w_j}{f_\epsilon(\omega_j)} \tag{C27}$$

Thus, $S^{(0)}$ is a *diagonal matrix*. It can be written in a matrix notation in the following way:

$$S^{(0)} = -\nabla_{\vec{\boldsymbol{\omega}}_r}(\nabla_{\vec{\boldsymbol{\omega}}_r} J^d_{energy})^T = 2\ \mathrm{diag}\left[\vec{\mathbf{w}} \oslash \tilde{f}_\epsilon(\vec{\boldsymbol{\omega}}_r)\right] \tag{C28}$$

The initial inverse-Hessian matrix is given by

$$S^{(0)^{-1}} = \frac{1}{2}\ \mathrm{diag}\left[\tilde{f}_\epsilon(\vec{\boldsymbol{\omega}}_r) \oslash \vec{\mathbf{w}}\right] \tag{C29}$$

This choice of $S^{(0)}$ yields the following initial direction of search:

$$\begin{aligned}\vec{\mathbf{p}}^{(0)} &= -S^{(0)^{-1}}\vec{\mathbf{g}}^{(0)} \\ &= -\vec{\bar{\boldsymbol{\epsilon}}}^{(0)}_r + \tilde{f}_\epsilon(\vec{\boldsymbol{\omega}}_r) \circ \left\{\mathcal{C}\left[-\mathrm{Im}\left\langle\chi^{(0)}(t)\left|\hat{\mathbf{X}}\right|\psi^{(0)}(t)\right\rangle\right]\Big|_{\vec{\boldsymbol{\omega}}=\vec{\boldsymbol{\omega}}_r} - \left[\lambda^{(0)}_0 + \lambda^{(0)}_T\cos(\vec{\boldsymbol{\omega}}_r T)\right]\right\}\end{aligned} \tag{C30}$$

Let us write the function version of the second term in the RHS:

$$\tilde{f}_\epsilon(\omega)\left\{\mathcal{C}\left[-\mathrm{Im}\left\langle\chi^{(0)}(t)\left|\hat{\mathbf{X}}\right|\psi^{(0)}(t)\right\rangle\right] - \left[\lambda^{(0)}_0 + \lambda^{(0)}_T\cos(\omega T)\right]\right\}$$

This expression can be recognized as the expression for $\bar{\epsilon}(\omega)$ from the solution Euler-Lagrange equation

(see Eqs. (B65), (B66)), with the substitution of $\left|\psi^{(0)}(t)\right\rangle$, $\left|\chi^{(0)}(t)\right\rangle$, $\lambda_0^{(0)}$ and $\lambda_T^{(0)}$. Let us denote:

$$\bar{\epsilon}_{EL}^{(k)}(\omega) \equiv \tilde{f}_\epsilon(\omega)\left\{\mathcal{C}\left[-\operatorname{Im}\left\langle\chi^{(k)}(t)\left|\hat{\mathbf{X}}\right|\psi^{(k)}(t)\right\rangle\right]-\left[\lambda_0^{(k)}+\lambda_T^{(k)}\cos(\omega T)\right]\right\} \tag{C31}$$

$$\vec{\bar{\boldsymbol{\epsilon}}}_{r,EL}^{(k)} \equiv \bar{\epsilon}_{EL}^{(k)}(\vec{\boldsymbol{\omega}}_r) \tag{C32}$$

Using this notation, Eq. (C30) can be rewritten as

$$\vec{\mathbf{p}}^{(0)} = -\vec{\bar{\boldsymbol{\epsilon}}}_r^{(0)} + \vec{\bar{\boldsymbol{\epsilon}}}_{r,EL}^{(0)} \tag{C33}$$

$\vec{\bar{\boldsymbol{\epsilon}}}_{r,EL}^{(k)}$ satisfies the boundary constraints and spectral restrictions for all k. $\vec{\bar{\boldsymbol{\epsilon}}}_r^{(0)}$ is also assumed to satisfy both requirements by an appropriate choice, as above. According to Eq. (C33), $\vec{\mathbf{p}}^{(0)}$ is a linear combination of $\vec{\bar{\boldsymbol{\epsilon}}}_r^{(0)}$ and $\vec{\bar{\boldsymbol{\epsilon}}}_{r,EL}^{(0)}$. Thus, $\vec{\mathbf{p}}^{(0)}$ also satisfies both requirements. Consequently, the update rule for $\vec{\bar{\boldsymbol{\epsilon}}}_r^{(1)}$ preserves the boundary and spectral restrictions.

Note that the relaxation method proposed in our previous work yields the same update direction as that of Eq. (C33), generalized to all k. It has been shown in [43, Chapter 3.2.3] that the relaxation method is equivalent to a quasi-Newton method with the same approximated Hessian (C28) for all iterations.

The situation is more complicated in the BFGS method, where the Hessian is updated in each iteration. Nevertheless, the BFGS update rule preserves the required properties for all k, with the proper choice of $\vec{\bar{\boldsymbol{\epsilon}}}_r^{(0)}$ and $S^{(0)}$, as above. It can be shown that the resulting direction of search $\vec{\mathbf{p}}^{(k)}$ is spanned by the following set of vectors:

$$\vec{\bar{\boldsymbol{\epsilon}}}_r^{(0)},\ \vec{\bar{\boldsymbol{\epsilon}}}_{r,EL}^{(0)},\ \vec{\bar{\boldsymbol{\epsilon}}}_{r,EL}^{(1)},\ldots,\vec{\bar{\boldsymbol{\epsilon}}}_{r,EL}^{(k)}$$

The full justification is left for a future publication. All the vectors in this set satisfy the required properties. Consequently, the required properties are preserved by the resulting update rule.

Thus, it has been shown that by a proper choice of $S^{(0)}$, the BFGS method yields an update rule which is based implicitly on the solution equation (B65).

It is noteworthy that there is an important particular case in which the common choice for the initial Hessian, $S^{(0)} = I$, conserves the required properties. This situation is characterized by the following conditions:

1. $f_\epsilon(\omega) \in \{0, 1\}$, i. e. $f_\epsilon(\omega)$ can attain the values 0 and 1 only. For instance, $f_\epsilon(\omega)$ is a rectangular function.

2. The boundary frequencies, $\omega_0 = 0$ and $\omega_{N_t} = N_t\pi/T$, are not contained in the allowed frequency region represented by $f_\epsilon(\omega)$.

The issue of conservation of the spectral restriction becomes irrelevant whenever the first condition is satisfied, since the "forbidden" spectral regions are completely eliminated from the optimization process in advance. If, in addition, the second condition is satisfied, ${S^{(0)}}^{-1}$ from Eq. (C29) differs from the identity matrix I just by a constant factor. This contributes a constant factor to $\vec{\mathbf{p}}^{(0)}$, which does not affect its direction. Thus, the common choice $S^{(0)} = I$ yields an update rule which conserves the boundary conditions as well.

e. The line-search

The line-search procedure is based on the scheme outlined in Ref. [54, Chapter 2.6]. We shall comment on several details in which we deviated from this scheme.

In [54, Chapter 2.6] it is assumed that the computation of the gradient is much more expensive numerically than the computation of the optimized function value. This assumption is reflected in the details of the line-search scheme—the gradient is computed only when it is essential for the search process. This excludes a certain situation in which the gradient information is not essential, although its knowledge improves the search process—the choice of the next point in the search is based on interpolation considerations; in the scheme outlined in [54, Chapter 2.6] it is preferred to employ a quadratic interpolation, rather than a cubic one with the cost of an extra gradient evaluation. In our situation, the economical considerations regarding the gradient computation are irrelevant—the vast majority of the numerical effort consists of the propagation process, and the extra computational effort in the gradient calculation is negligible. Thus, the gradient should be computed whenever it can be utilized for the improvement of the process. Hence, we always employed a cubic interpolation for the choice of the next point in the search.

The conditions for an acceptable point in [54, Chapter 2.5] include the requirement that the point is located below a decreasing "ρ-line" (the first Goldstein condition). This condition is unnecessary in our case, since it is employed to exclude a situation which is irrelevant in our problem (see [54, Page 30]). In an existing line-search program we can set $\rho = 0$.

It is also unnecessary to include a lower bound for the functional value ($\bar{f}$ in [54, Chapter 2.6]) from similar considerations. In an existing program we can set $\bar{f} = -\infty$.

There is a unique issue in quantum optimal control theory problems which requires a specialized

treatment in the context of the line-search scheme. The computation of the functional involves propagation under the solution vector field. Typically, the precise description of the dynamics becomes more demanding numerically as the magnitude of the field increases. Of course, the high numerical requirements cannot be avoided if the optimal field is actually large in magnitude. However, it may occur that the optimization search path passes through fields which are much larger in magnitude than the optimal field. Thus, it is desirable to control the optimization path such that large field regions are avoided.

In the context of the line-search process, it often happens that the bracket located in the bracketing phase is much larger than required. In other words, the search unnecessarily extends "too far" in the search direction $\vec{\mathbf{p}}$. The corresponding fields might be considerably larger than the optimal field. As a result, the propagation process might become much time-consuming, inaccurate, or unstable.

We issued this problem in a primitive way—we restricted the peak magnitude of the field which is allowed in the line-search process. Let us denote the maximal allowed magnitude of $\epsilon(t)$ by ϵ_{max}; the updated field in each iteration should satisfy the following condition:

$$\left|\epsilon^{(k+1)}(t)\right| \leq \epsilon_{max}, \qquad t \in [0,T] \tag{C34}$$

which is equivalent to the following set of conditions:

$$\epsilon^{(k+1)}(t) \leq \epsilon_{max} \tag{C35}$$

$$\epsilon^{(k+1)}(t) \geq -\epsilon_{max} \tag{C36}$$

$$t \in [0,T]$$

These conditions yield a condition on the maximal allowed $\kappa^{(k)}$ in the search, as will be readily shown. First, conditions (C35), (C36) have to be discretized as follows:

$$\vec{\boldsymbol{\epsilon}}^{(k+1)} \leq \epsilon_{max} \tag{C37}$$

$$\vec{\boldsymbol{\epsilon}}^{(k+1)} \geq -\epsilon_{max} \tag{C38}$$

where we defined:

$$\vec{\boldsymbol{\epsilon}} \equiv \begin{bmatrix} \epsilon(t_0) \\ \epsilon(t_1) \\ \vdots \\ \epsilon(t_{N_t}) \end{bmatrix} \tag{C39}$$

$\vec{\boldsymbol{\epsilon}}^{(k+1)}$ can be expressed in the terms of the discretized solution $\vec{\tilde{\boldsymbol{\epsilon}}}$:

$$\vec{\boldsymbol{\epsilon}}^{(k+1)} = C^{-1}\vec{\tilde{\boldsymbol{\epsilon}}}^{(k+1)} \tag{C40}$$

where C denotes a matrix which represents the DCT linear transformation defined by Eq. (C6). Accordingly, C^{-1} represents the inverse DCT transformation defined by Eq. (C7). In order to express $\vec{\tilde{\boldsymbol{\epsilon}}}^{(k+1)}$ by the terms of $\kappa^{(k)}$, the update rule (C25) has to reformulated in the full N_t+1 dimensional space in which $\vec{\tilde{\boldsymbol{\epsilon}}}^{(k+1)}$ is defined. Let us define for $\vec{\mathbf{p}}^{(k)}$ a corresponding vector $\vec{\mathbf{p}}_f^{(k)}$ in the full space. The update rule in the full space becomes

$$\vec{\tilde{\boldsymbol{\epsilon}}}^{(k+1)} = \vec{\tilde{\boldsymbol{\epsilon}}}^{(k)} + \kappa^{(k)}\vec{\mathbf{p}}_f^{(k)} \tag{C41}$$

Eqs. (C40), (C41) yield:

$$\vec{\boldsymbol{\epsilon}}^{(k+1)} = C^{-1}\vec{\tilde{\boldsymbol{\epsilon}}}^{(k)} + \kappa^{(k)}C^{-1}\vec{\mathbf{p}}_f^{(k)} = \vec{\boldsymbol{\epsilon}}^{(k)} + \kappa^{(k)}\vec{\mathbf{q}}^{(k)} \tag{C42}$$

where

$$\vec{\mathbf{q}}^{(k)} \equiv C^{-1}\vec{\mathbf{p}}_f^{(k)} \tag{C43}$$

It can be easily found that the substitution of Eq. (C42) into conditions (C37), (C38) yields the following set of N_t+1 conditions on $\kappa^{(k)}$:

$$\kappa^{(k)} \leq \frac{\epsilon_{max} - \operatorname{sgn}\left(q_{j+1}^{(k)}\right)\epsilon^{(k)}(t_j)}{\left|q_{j+1}^{(k)}\right|}, \qquad 0 \leq j \leq N_t \tag{C44}$$

where $q_{j+1}^{(k)}$ is the $(j+1)$'th entry of $\vec{\mathbf{q}}^{(k)}$, which corresponds to the time-point t_j.

This practice should be distinguished from optimization with inequality constraints. Conditions (C35), (C36) do not represent an additional *constraint* on the *optimization problem*; they just impose a restriction on the *search process*.

When condition (C44) is imposed, it may occur that the bracketing phase fails. Usually, this indicates that the chosen ϵ_{max} is too small for the given problem, and it should be increased. Alternatively, the problem can be altered by the increment of the penalties on energy or ionization. These modifications will result in an optimal field of smaller magnitude.

The parameter values of the line-search procedure are summarized in Table III. We use the notations of [54, Chapter 2.6], with the exception of the parameter $\kappa^{(0)}$, which is defined in the present text.

σ	0.9
ρ	0
τ_1	9
τ_2	0.1
τ_3	0.5
$\bar{f}$	$-\infty$
$\kappa^{(0)}$	1
ϵ_{max}	0.15

TABLE III. Line-search parameters

f. Practical remarks

The choice of the initial guess field is important for the success of the computational method. It is recommended to choose $\vec{\bar{\boldsymbol{\epsilon}}}_r^{(0)}$ such that $J^{d\,(0)} > 0$, for the following reason. As has already been mentioned, there seems to be a local minimum of $-J^d$, or a saddle point, in the zero field solution, $\vec{\bar{\boldsymbol{\epsilon}}}_r \equiv \vec{\mathbf{0}}$. This solution yields $J^d = 0$. If $-J^{d\,(0)} > 0$, or equivalently, $J^{d\,(0)} < 0$, the optimization process has a strong tendency to converge to the zero field solution.

The task of locating a guess solution which yields $J^{d\,(0)} > 0$ often becomes non-trivial. The guess field should produce some response in the required region, such that the magnitude of $J_{max}^{(0)}$ is sufficiently large. One option is to use some physical insight in the choice of $\vec{\bar{\boldsymbol{\epsilon}}}_r^{(0)}$ such that the magnitude of $J_{max}^{(0)}$ is significant. Another option is to employ a *preparation optimization*—we perform several iterations of another optimization problem, in order to find a field for which $J_{max}^{(0)}$ becomes significant. In the preparation problem, the penalties on energy and ionization are reduced, such that the initial functional value *of the preparation problem* becomes positive.

Occasionally, a quasi-Newton scheme applied to the present problem fails in the approximation of the Hessian. The optimization "gets stuck" after several iterations, when the resulting direction of search $\vec{\mathbf{p}}^{(k)}$ ceases to be useful. Theoretically, the BFGS update rule of the inverse-Hessian (C23) conserves the positive definiteness of S^{-1}. According to Eq. (C24), this yields a direction of search $\vec{\mathbf{p}}^{(k)}$ which is always a descent direction. However, in practice, the magnitude of the negative slope in the $\vec{\mathbf{p}}^{(k)}$ direction may be too small to exceed the roundoff error regime. As a result, no improvement can be achieved in a search in the $\vec{\mathbf{p}}^{(k)}$ direction.

The origin of the problem lies in the failure of the assumptions underlying the inverse-Hessian update rule (C23). The update rule relies on the following quadratic approximation:

$$\vec{\boldsymbol{\gamma}}^{(k)} \approx S^{(k+1)} \vec{\boldsymbol{\delta}}^{(k)} \tag{C45}$$

The quadratic approximation is relevant only when the new point $\vec{\boldsymbol{\epsilon}}_r^{(k+1)}$ is located in the nearest valley in the $\vec{\mathbf{p}}^{(k)}$ direction. The situation might be different when there exist additional further valleys in the $\vec{\mathbf{p}}^{(k)}$ direction. In this case, the accepted $\vec{\boldsymbol{\epsilon}}_r^{(k+1)}$ in the line-search scheme may be located in these further valleys, where approximation (C45) does not hold. Consequently, the update rule (C23) becomes irrelevant for the approximation of the inverse Hessian. After the accumulation of several events of this type, the resulting ${S^{(k)}}^{-1}$ ceases to be useful.

The existence of several valleys in the search direction is typical to a guess field of small magnitude combined with small penalties on energy and ionization. This results in small magnitudes of the penalty terms in the region of $\vec{\boldsymbol{\epsilon}}_r^{(k)}$ in the optimization space. In general, the penalty terms introduce a "wall" or a barrier in the optimization space. This can prevent the appearance of additional valleys. If the penalty terms in the vicinity of $\vec{\boldsymbol{\epsilon}}_r^{(k)}$ are small in magnitude, they do not have a significant effect on the optimization space in the region of $\vec{\boldsymbol{\epsilon}}_r^{(k)}$, and additional further valleys may appear.

The problem can be solved by the "reset" of the inverse-Hessian approximation when the process "gets stuck". The inverse-Hessian is set again to the form of (C29), and the iterative update of the inverse-Hessian by Eq. (C23) is restarted from this point. This is equivalent to starting a new optimization from the final solution of the first optimization.

A failure of the inverse-Hessian approximation can usually be detected before the total failure of the optimization, since it is characterized by an unusual behaviour of the process. We may observe too many iterations in the different phases of the line-search, and certain iterations of the $\vec{\boldsymbol{\epsilon}}_r$ update with very small improvement. If an unusual behaviour is detected, it is recommended to reset the inverse-Hessian in this stage, instead of waiting until the optimization is completely "stuck".

An alternative way to address this problem is to use the form of (C29) as the inverse-Hessian approximation also for $k > 0$, until $\vec{\boldsymbol{\epsilon}}_r^{(k)}$ and the penalty terms become sufficiently large to prevent the appearance of the problem.

Appendix D: The effect of the absorbing boundaries

As was mentioned in Sec. III B, the employment of absorbing boundary conditions might be a primary source of inaccuracy in HHG simulations. In principle, two sources of inaccuracy are introduced when an infinitely long spatial grid is replaced by absorbing boundary conditions:

1. *Physical effect*: The absorption of a portion of the wave-function amplitude at the boundaries can have an effect on the dynamics in the central region of the wave-function. The possible physical effects can be classified as *direct* or *indirect* effects:

   (a) *Direct effect*: The unjustified elimination of an electronic amplitude which is due to revert to the central region in a later stage in the dynamical process;

   (b) *Indirect effect*: The wave-function dynamics in all spatial regions is *coupled* by the time-dependent Schrödinger equation. Thus, the elimination of the amplitude at the boundaries has some effect on the dynamics in the central region of the wave-function.

2. *Numerical effect*: Imperfection in the absorption capabilities of the absorbing boundaries results in the effects of reflection from the absorbing boundary, or transmission and wraparound of the electronic amplitude.

Both the physical and the numerical effects can be reduced by placing the absorbing boundaries further from the central region of the grid. However, neither of them can be completely eliminated. As was discussed in Sec. III B, the magnitude of the numerical effect strongly depends on the choice of the complex-absorbing-potential. This motivated the employment of the optimization scheme described in Sec. III B for its construction.

The validity of the approximations introduced by the employment of absorbing boundaries was tested in all problems of Sec. IV. As a test of validity, the optimized fields obtained in all problems were used to propagate the same initial condition in a doubled grid, i. e. $x \in [-480, 480)$ where the spacing between adjacent points remains unaltered (this test has failed in [45], due to the presence of the numerical effect mentioned above). The resulting $\overline{\left\langle \hat{\mathbf{C}} \right\rangle}(\omega)$ spectra were calculated. We found that there is no significant difference between the spectra of the original grid and the doubled grid. In order to quantify the magnitude of the deviation of the response in the original grid from the doubled grid, we define the relative difference of J_{max}:

$$\Delta_{rel} J_{max} \equiv \frac{J_{max} - J_{max,doubled}}{J_{max,doubled}} \tag{D1}$$

n	$\Delta_{rel} J_{max}$
13	2.42×10^{-3}
14	9.02×10^{-4}
15	1.44×10^{-3}
17	2.41×10^{-5}

TABLE IV. The relative difference of J_{max} from the resulting J_{max} in the doubled grid (Eq. (D1)) for the various optimized harmonics.

where $J_{max,doubled}$ is the J_{max} calculated for the doubled grid. In Table IV, we present $\Delta_{rel} J_{max}$ for the various optimized harmonics. It can be observed that the magnitude of the relative difference does not exceed the order of $\sim 10^{-3}$. Thus, the effects introduced by the absorbing boundaries are not expected to have a significant effect on the optimization problem.

Appendix E: Prevention of plasma production represented as a control requirement

The minimization of plasma production in HHG is one of the aims of the current study. In this appendix we shall further discuss its realization by the control formulation outlined in Sec. II.

There is a fundamental problem in formulating a *control requirement* representing the *physical requirement* of prevention of plasma production. Plasma production is a macroscopic effect, which takes place in the macroscopic medium. However, the treatment of the dynamics in the present study is in the isolated system level. Thus, there is no direct access to the physical effect of plasma production by the current dynamical treatment. Consequently, it is impossible to give a direct formulation of the physical requirement as a control requirement.

As an alternative, it is possible to define the physical requirement of *localization* of the liberated electron around the parent ion. The localization of the electron ensures that it is not liberated into the macroscopic medium, and thus the plasma production is prevented. In more precise means, we can define a radius around the parent ion, where the electronic probability inside the radius is considered as localized. We shall refer to this radius as the *localization radius*. Such a physical requirement is accessible by the dynamical treatment of the isolated system, and thus can be formulated as a control requirement.

The immediate question which arises is how the localization radius should be chosen. It is important

to avoid an over-localization of the electron around the parent ion. The ionization and liberation of the electron into large distances in the continuum is an integral part of the HHG mechanism, where the electron gains large energies in its accelerated motion reverted into the parent ion over a large distance. If the chosen localization radius is too small, the feasibility of production of the target harmonic can be completely prevented, or at least reduced.

It should be emphasized that the insertion of an additional control requirement always reduces the optimal yield, since the space of the allowed control opportunities becomes restricted. However, we should distinguish between two types of reduction of the optimal harmonic yield by the localization requirement:

1. *Direct reduction*: The spatial restriction imposed on the electron damages directly the opportunities provided by the *physical mechanism* responsible for HHG, as above. In this case, the localization requirement imposes a direct restriction on the electronic probability which *participates in the HHG process*.

2. *Indirect reduction*: The HHG process has the *side effect* of liberated electronic probability which does not revert to the parent ion, and consequently, is responsible for plasma production. This part of the electronic probability does not participate in the physical mechanism responsible for the production of high harmonics. The localization requirement restricts this side effect. An indirect consequence of this restriction is the reduction of the control opportunities, where the optimal mechanism should satisfy the additional restriction of the side effect. Thus, the localization requirement restricts directly an electronic probability which *does not participate in the HHG process*, and the probability which participates in the process also becomes restricted as an indirect consequence of this requirement.

When we state that over-localization of the electron should be avoided, we intend to the prevention of the *direct reduction* of the optimal yield. However, the indirect reduction is an unavoidable consequence of the physical requirement of prevention of plasma production.

In what follows, we claim that the considerations in the choice of the localization radius are similar to the considerations of the choice position of the *absorbing boundaries*.

We shall start from the considerations of the choice of the position of the absorbing boundaries. In Appendix D we distinguished between two sources of inaccuracy induced by the absorbing boundaries, which were classified as a *physical effect* and a *numerical effect*. Let us assume, for the moment, that the absorption capabilities of the absorbing boundaries are nearly perfect, and the numerical effect is negligible. In this case, the position of the absorbing boundaries is chosen by purely physical considerations;

the choice should reduce the physical effect to a negligible magnitude. The term "negligible" depends on the required accuracy. This ensures that the absorption of the outgoing electronic amplitude has no significant effect on the physics in the central region of the grid.

The appropriate position of the absorbing boundaries depends on the specific profile of the laser pulse.

The determination of the appropriate position in an optimization problem becomes problematic. The laser pulse profile varies during the optimization process, and thus the position of the absorbing boundaries has to be far enough from the central region to consider all fields which take place in the optimization path. The problem is that these fields are unknown in advance. However, the position of the absorbing boundaries must be determined in advance, since it defines the system, and consequently, the optimization problem.

Nevertheless, an upper limit for the position of the absorbing boundaries can be roughly estimated, since we actually have some idea about important physical properties of the fields in the optimization path—the control requirements restrict the intensity of the field and the available frequency band. The position of the absorbing boundaries can be estimated by comparison to HHG problems with similar physical conditions.

Now we return to the discussion of the choice of the localization radius. In order to avoid the direct reduction of the optimal yield by the localization requirement, the localization requirement should exclude only electronic probability which is far enough from the parent ion to become irrelevant to the physics responsible for the high harmonic production. In other words, the localization radius is chosen such that the electronic probability which is beyond the localization radius has a negligible effect on the dynamics in the central region. This coincides with the appropriate position of the absorbing boundaries. This justifies the formulation of the localization control requirement by J_{ion} (defined in Eq. (27)), where the absorbed probability is identified as the permanently ionized probability.

Let us consider also the situation in which the imperfection in the absorption capabilities of the absorbing potential cannot be ignored. In this case, the numerical effect of the absorbing boundaries is non-negligible. The position of the absorbing boundaries is not led by physical considerations only, but also by numerical considerations. The common practice is to place the absorbing boundaries further from the central region of the grid, such that the numerical effect becomes negligible. In this case, the correspondence between the localization radius and the position of the absorbing boundaries is lost.

Nevertheless, the method outlined in Sec. II C for the restriction of permanent ionization can still be applied. The considerations which were outlined here for the choice of the localization radius impose only

a lower limit on the localization radius, but not an upper limit. The only problem in the exaggeration of the magnitude of the localization radius is the extra numerical cost of larger grids. Hence, when the use of larger grid is crucial from numerical considerations, the localization radius can still be defined by the position of the absorbing boundaries. However, it should be emphasized that the minimal allowed survival amplitude has to be increased as the distance of the absorbing boundaries from the parent ion is increased. The profile of $\sigma(y)$ has to be altered accordingly. Thus, the current formulation couples between the *numerical realization of the dynamics* and the *optimization problem*.

---

[1] A. McPherson, G. Gibson, H. Jara, U. Johann, T. S. Luk, I. McIntyre, K. Boyer, and C. K. Rhodes, JOSA B **4**, 595 (1987).

[2] M. Ferray, A. Lhuillier, X. F. Li, L. A. Lompre, G. Mainfray and C. Manus, J. Phys. B-At. Mol. Opt. Phys. **21**, L31 (1988).

[3] X. F. Li, A. L'Huillier, M. Ferray, L. A. Lompré, and G. Mainfray, Phys. Rev. A **39**, 5751 (1989).

[4] J. Mauritsson, P. Johnsson, E. Gustafsson, A. L'Huillier, K. J. Schafer, and M. B. Gaarde, Phys. Rev. Lett. **97**, 013001 (2006).

[5] T. Brabec and F. Krausz, Reviews of Modern Physics **72**, 545 (2000).

[6] P. M. Paul, E. S. Toma, P. Breger, G. Mullot, F. Augé, P. Balcou, H. G. Muller, and P. Agostini, Science **292**, 1689 (2001), http://science.sciencemag.org/content/292/5522/1689.full.pdf.

[7] P. á. Corkum and F. Krausz, Nature physics **3**, 381 (2007).

[8] M. T. Hassan, T. T. Luu, A. Moulet, O. Raskazovskaya, P. Zhokhov, M. Garg, N. Karpowicz, A. Zheltikov, V. Pervak, F. Krausz, *et al.*, Nature **530**, 66 (2016).

[9] H. J. Wörner, J. B. Bertrand, B. Fabre, J. Higuet, H. Ruf, A. Dubrouil, S. Patchkovskii, M. Spanner, Y. Mairesse, V. Blanchet, *et al.*, Science **334**, 208 (2011).

[10] I. Luzon, K. Jagtap, E. Livshits, O. Lioubashevski, R. Baer, and D. Strasser, Physical Chemistry Chemical Physics **19**, 13488 (2017).

[11] A. Bhattacherjee and S. R. Leone, Accounts of chemical research **51**, 3203 (2018).

[12] P. B. Corkum, Physical review letters **71**, 1994 (1993).

[13] M. Lewenstein, P. Balcou, M. Y. Ivanov, A. Lhuillier and P. B. Corkum, Phys. Rev. A **49**, 2117 (1994).

[14] O. Smirnova and M. Ivanov, arXiv preprint arXiv:1304.2413 (2013).

[15] O. Smirnova, Y. Mairesse, S. Patchkovskii, N. Dudovich, D. Villeneuve, P. Corkum, and M. Y. Ivanov,

Nature **460**, 972 (2009).

[16] A. Fleischer, O. Kfir, T. Diskin, P. Sidorenko, and O. Cohen, Nature Photonics **8**, 543 (2014).

[17] O. Neufeld, E. Bordo, A. Fleischer, and O. Cohen, New Journal of Physics **19**, 023051 (2017).

[18] S. A. Rice, Science **258**, 412 (1992).

[19] K. Burnett, V. C. Reed, and P. L. Knight, Journal of Physics B: Atomic, Molecular and Optical Physics **26**, 561 (1993).

[20] C. Cerjan and R. Kosloff, Physical Review A **47**, 1852 (1993).

[21] N. Ben-Tal, N. Moiseyev, R. Kosloff, and C. Cerjan, Journal of Physics B: Atomic, Molecular and Optical Physics **26**, 1445 (1993).

[22] D. J. MacKay and D. J. Mac Kay, *Information theory, inference and learning algorithms* (Cambridge university press, 2003).

[23] X. Chu and S.-I. Chu, Physical Review A **64**, 021403 (2001).

[24] P. Villoresi, S. Bonora, M. Pascolini, L. Poletto, G. Tondello, C. Vozzi, M. Nisoli, G. Sansone, S. Stagira, and S. De Silvestri, Optics letters **29**, 207 (2004).

[25] R. A. Bartels, M. M. Murnane, H. C. Kapteyn, I Christov, and H. Rabitz , Phys. Rev. A **70**, 043404 (2004).

[26] C. Winterfeldt, C. Spielmann, and G. Gerber, Rev. Mod. Phys. **80**, 117 (2008).

[27] A. S. Johnson, D. R. Austin, D. A. Wood, C. Brahms, A. Gregory, K. B. Holzner, S. Jarosch, E. W. Larsen, S. Parker, C. S. Strüber, *et al.*, Science advances **4**, eaar3761 (2018).

[28] A. P. Peirce, M. A. Dahleh, H. Rabitz, Phys. Rev. A **37**, 4950 (1988).

[29] Ronnie Kosloff, Stuart A. Rice, Pier Gaspard, Sam Tersigni and David Tannor, Chem. Phys. **139**, 201 (1989).

[30] S. J. Glaser, U. Boscain, T. Calarco, C. P. Koch, W. Köckenberger, R. Kosloff, I. Kuprov, B. Luy, S. Schirmer, T. Schulte-Herbrüggen, *et al.*, The European Physical Journal D **69**, 279 (2015).

[31] A. Aroch, S. Kallush, and R. Kosloff, Physical Review A **97**, 053405 (2018).

[32] José P. Palao and Ronnie Kosloff, Phys. Rev. A **68**, 062308 (2003).

[33] J. Somlói, V. A. Kazakov, and D. J. Tannor, Chemical physics **172**, 85 (1993).

[34] Y. Ohtsuki, G. Turinici, and H. Rabitz, The Journal of chemical physics **120**, 5509 (2004).

[35] R. Eitan, M. Mundt, and D. J. Tannor, Physical Review A **83**, 053426 (2011).

[36] I. Degani, A. Zanna, L. Sælen, and R. Nepstad, SIAM J. Sci. Comput. **31**, 3566 (2009).

[37] S. G. Schirmer and P. de Fouquieres, New Journal of Physics **13**, 073029 (2011).

[38] D. M. Reich, M. Ndong, and C. P. Koch, The Journal of chemical physics **136**, 104103 (2012).

[39] I. Serban, J. Werschnik, and E. K. U. Gross, Phys. Rev. A **71**, 053810 (2005).

[40] J. Werschnik and E. K. U. Gross, Journal of Physics B: Atomic, Molecular and Optical Physics **40**, R175 (2007).

[41] José P. Palao, Ronnie Kosloff, and Christiane P. Koch, Phys. Rev. A **77**, 063412 (2008).

[42] S. Pezeshki, M. Schreiber, and U. Kleinekathöfer, Phys. Chem. Chem. Phys. **10**, 2058 (2008).

[43] I. Schaefer, arXiv:1202.6520 (2012).

[44] I. Schaefer and R. Kosloff, Phys. Rev. A **86**, 063417 (2012).

[45] J. L. Krause, K. J. Schafer, and K. C. Kulander, Phys. Rev. A **45**, 4998 (1992).

[46] Y. Yu and B. D. Esry, Journal of Physics B: Atomic, Molecular and Optical Physics **51**, 095601 (2018).

[47] J. Solanpää, J. A. Budagosky, N. I. Shvetsov-Shilovski, A. Castro, A. Rubio, and E. Räsänen, Phys. Rev. A **90**, 053402 (2014).

[48] A. Castro, A. Rubio, and E. K. U. Gross, The European Physical Journal B **88**, 191 (2015).

[49] E. Balogh, B. Bódi, V. Tosa, E. Goulielmakis, K. Varjú, and P. Dombi, Phys. Rev. A **90**, 023855 (2014).

[50] C. Jin and C. D. Lin, Chinese Physics B **25**, 094213 (2016).

[51] J. B. Schönborn, P. Saalfrank, and T. Klamroth, The Journal of Chemical Physics **144**, 044301 (2016), https://doi.org/10.1063/1.4940316.

[52] T. E. Skinner, N. I. Gershenzon, J. Mag. Res. **204**, 248 (2010).

[53] M. B. Gaarde, J. L. Tate, and K. J. Schafer, Journal of Physics B: Atomic, Molecular and Optical Physics **41**, 132001 (2008).

[54] R. Fletcher, *Practical methods of optimization* (John Wiley & Sons, 1987).

[55] Hillel Tal-Ezer, Ronnie Kosloff, Ido Schaefer, J. Sci. Comput. **53**, 211 (2012).

[56] I. Schaefer, H. Tal-Ezer, and R. Kosloff, Journal of Computational Physics **343**, 368 (2017).

[57] D. Kosloff and R. Kosloff, Journal of Computational Physics **52**, 35 (1983).

[58] R. Kosloff, The Journal of Physical Chemistry **92**, 2087 (1988).

[59] J. Muga, J. Palao, B. Navarro, and I. Egusquiza, Physics Reports **395**, 357 (2004).

[60] J. Palao and J. Muga, Chemical Physics Letters **292**, 1 (1998).

[61] T. Kalotas and A. Lee, American Journal of Physics **59**, 48 (1991).

[62] N. Ben-Tal, N. Moiseyev, and A. Beswick, Journal of Physics B: Atomic, Molecular and Optical Physics **26**, 3017 (1993).

[63] O. E. Alon, V. Averbukh, and N. Moiseyev, Phys. Rev. Lett. **80**, 3743 (1998).

[64] D. J. Diestler, Phys. Rev. A **78**, 033814 (2008).

[65] J. C. Baggesen and L. B. Madsen, Journal of Physics B: Atomic, Molecular and Optical Physics **44**, 115601 (2011).

[66] J. A. Pérez-Hernández and L. Plaja, Journal of Physics B: Atomic, Molecular and Optical Physics **45**, 028001 (2011).

[67] J. C. Baggesen and L. B. Madsen, Journal of Physics B: Atomic, Molecular and Optical Physics **45**, 028002 (2011).

Chapter 5

Discussion and conclusion

In the present thesis, a theoretical optimization method of HHG has been developed in the framework of quantum OCT. It represents the first theoretical work which addressed the task of HHG control. Studies which follow similar lines [4, 22] appeared after the publication of our first paper (Chapter 2). However, there is still a large ground for further research on this topic.

We first addressed the general problem of harmonic generation control. The general method was later adjusted to the HHG problem, which is considerably more complex physically.

Quantum OCT is naturally expressed in the time-domain, while the harmonic generation problem mainly consists of spectral requirements, which are naturally expressed in the frequency domain. The maximization functional which was employed is essentially formulated in the frequency domain, unlike in the common quantum OCT formalism. The more general issue of imposing restrictions on the driving field spectrum was also addressed by our formulation.

The required boundary conditions of the driving field were imposed by additional constraint equations, and by the choice of the cosine spectral representation of the driving

field.

An important issue in HHG control is the restriction of permanent ionization. This was realized by two different methods:

1. By imposing a penalty on spatial regions which are away from the parent ion;

2. By imposing a penalty on the permanent ionization probability.

The first method was found to be appropriate only when complete elimination of permanent ionization is required. We found this requirement to be unachievable in typical HHG problems. The second method was found to be quite efficient in the restriction of permanent ionization to a predefined allowed probability. A reliable method of ionization restriction was lacking, and the current research fills this gap.

A special emphasis was given on the numerical realization of the problem. This topic was found to be of utmost importance in the context of HHG control. The present thesis represents our efforts in two different grounds:

1. Reliable *numerical simulation* of the HHG dynamics;

2. Efficient *optimization* method.

The *simulation* of HHG requires highly accurate tools. We employed a new, highly accurate and efficient propagation method, which is based on a semi-global approach for the propagation (Chapter 3). Theoretical and practical aspects of the propagation method were thoroughly discussed. The application to the physics of HHG was demonstrated to be successful. The maximal achievable accuracy in a double precision considerably exceeds that of the common Runge-Kutta 4 (RK4) method. In addition, the semi-global approach was proved to yield a vast improvement in the efficiency compared to RK4, with much faster error decay rates with the decrement of the time-step size.

The absorbing boundaries, required for the description of the HHG dynamics, were constructed by a new optimization scheme. The resulting complex-absorbing-potential was found to be satisfactory in the elimination of notable reflection and transmission effects.

The optimization was performed by two different schemes:

1. Relaxation scheme, based on direct functional evaluations;

2. Second-order gradient scheme (BFGS), based on gradient information.

The first method was found to be satisfactory for the optimization of low-order harmonic-generation problems. It was applied also to a HHG problem in Chapter 2. Later, this method was considered by us as ineffective for the HHG problem, due to the complexity of the optimization surface, which reflects the complexity of the physical situation. This led to the employment of the more sophisticated BFGS scheme in Chapter 4. It was required to adjust this scheme to the current problem, as was discussed in length in the appendix of Chapter 4.

The first HHG problem in which the method was demonstrated was atypical (Chapter 2), where one of the Bohr-frequencies was targeted. In Chapter 4, the method was applied to typical HHG problems, where the target frequency is an above-threshold multiple of the fundamental frequency of the source. We demonstrated selective enhancement of harmonics, with minimal total energy and restricted permanent ionization probability. The maximization of an even harmonic target was also achieved quite impressively.

The physical interpretation of optimized fields was not discussed in this framework. Nevertheless, insights into the mechanisms underlying optimized fields were obtained. We found that the fields obtained in the odd-harmonic problems consist of two stages:

1. Escape from the regime in which the adiabatic approximation holds; we found that the liberation of the electron from the adiabatic regime is the main barrier for the initiation of the HHG process.

2. Generation of harmonics by quasi-periodic patterns.

This is a special case of the two-stage structure found in [20] in the more general context of harmonic generation control. The description of further details of the mechanism was postponed for future publications.

Although important insights of the underlying mechanisms have been obtained, a full understanding of the mechanisms requires a further research. We hope that a further research along the lines outlined in this thesis will lead to a better understanding of the HHG process, and will enable an intuitive prediction of fields for experimental selective enhancement of harmonics.

Our model system is one-dimensional. Although the one-dimensional model is non-realistic, it preserves the basic features of the HHG process. The investigation of this simplified model has two important advantages:

1. The relative simplicity of the numerical simulations, which require much less computational resources; this leads directly to reasonable requirements of time resources, which is crucial for the feasibility of a thorough research of the problem.

2. The simplified model enables a "refinement" of the most fundamental features of the problem, which is important for a better understanding of the underlying physics.

However, there is also a considerable importance in the application to more realistic models. The importance of this task has several aspects:

1. The simplified model does not represent all the features of the realistic physics; while the one-dimensional model represents the effects of primary importance, effects of secondary importance also should be investigated. These include multi-electron processes, spreading of the wave-function in transversal dimensions to the polarization of the field, nonlinear conjugation of transversal components of the dipole motion, and macroscopic propagation effects.

2. The effect of polarization of the field can be investigated only in multidimensional models.

3. The realistic modelling is crucial for experimental implementation of optimized fields.

The implementation of the new method to realistic problems requires considerable further work. The hope is that optimized fields obtained by the theoretical scheme presented here will be used as a starting-point in experimental optimization processes. The experimental optimization will correct for the inaccuracies of the theoretical models and approximations. The combined scheme will yield an optimal field for experimental use.

Bibliography

[1] A. P. Peirce, M. A. Dahleh, H. Rabitz, *Optimal control of quantum-mechanical systems: Existence, numerical approximation, and application*, Phys. Rev. A **37** (1988), 4950.

[2] E. Balogh, B. Bódi, V. Tosa, E. Goulielmakis, K. Varjú, and P. Dombi, *Genetic optimization of attosecond-pulse generation in light-field synthesizers*, Phys. Rev. A **90** (2014), 023855.

[3] Thomas Brabec and Ferenc Krausz, *Intense few-cycle laser fields: Frontiers of nonlinear optics*, Reviews of Modern Physics **72** (2000), no. 2, 545.

[4] Alberto Castro, Angel Rubio, and Eberhard K. U. Gross, *Enhancing and controlling single-atom high-harmonic generation spectra: a time-dependent density-functional scheme*, The European Physical Journal B **88** (2015), no. 8, 191.

[5] Xi Chu and Shih-I Chu, *Optimization of high-order harmonic generation by genetic algorithm and wavelet time-frequency analysis of quantum dipole emission*, Physical Review A **64** (2001), no. 2, 021403.

[6] P áB Corkum and Ferenc Krausz, *Attosecond science*, Nature physics **3** (2007), no. 6, 381.

[7] Paul B Corkum, *Plasma perspective on strong field multiphoton ionization*, Physical review letters **71** (1993), no. 13, 1994.

[8] E. Constant, D. Garzella, P. Breger, E. Mevel, C. Dorrer, C. Le Blanc, F. Salin, F and P. Agostini, *Optimizing high harmonic generation in absorbing gases: Model and experiment*, Phys. Rev. Lett. **82** (1999), no. 8, 1668.

[9] Cheng Jin and C D Lin, *Optimization of multi-color laser waveform for high-order harmonic generation*, Chinese Physics B **25** (2016), no. 9, 094213.

[10] José P. Palao and Ronnie Kosloff, *Optimal control theory for unitary transformations*, Phys. Rev. A **68** (2003), 062308.

[11] X. F. Li, A. L'Huillier, M. Ferray, L. A. Lompré, and G. Mainfray, *Multiple-harmonic generation in rare gases at high laser intensity*, Phys. Rev. A **39** (1989), 5751–5761.

[12] M. Ferray, A. Lhuillier, X. F. Li, L. A. Lompre, G. Mainfray and C. Manus, *MULTIPLE-HARMONIC CONVERSION OF 1064-NM RADIATION IN RARE-GASES*, J. Phys. B-At. Mol. Opt. Phys. **21** (1988), no. 3, L31.

[13] M. Lewenstein, P. Balcou, M. Y. Ivanov, A. Lhuillier and P. B. Corkum, *"THEORY OF HIGH-HARMONIC GENERATION BY LOW-FREQUENCY LASER FIELDS"*, Phys. Rev. A **49** (1994), no. 3, 2117.

[14] J. Mauritsson, P. Johnsson, E. Gustafsson, A. L'Huillier, K. J. Schafer, and M. B. Gaarde, *Attosecond pulse trains generated using two color laser fields*, Phys. Rev. Lett. **97** (2006), 013001.

[15] A McPherson, G Gibson, H Jara, U Johann, Ting S Luk, IA McIntyre, Keith Boyer, and Charles K Rhodes, *Studies of multiphoton production of vacuum-ultraviolet radiation in the rare gases*, JOSA B **4** (1987), no. 4, 595–601.

[16] N. H. Burnett, H. A. Baldis, M. C. Richardson, and G. D. Enright, *HARMONIC-GENERATION IN CO2-LASER TARGET INTERACTION*, Appl. Phys. Letter **31** (1977), no. 3, 172.

[17] P. M. Paul, E. S. Toma, P. Breger, G. Mullot, F. Augé, Ph. Balcou, H. G. Muller, and P. Agostini, *Observation of a train of attosecond pulses from high harmonic generation*, Science **292** (2001), no. 5522, 1689–1692.

[18] R. A. Bartels, M. M. Murnane, H. C. Kapteyn, I Christov, and H. Rabitz , *Learning from learning algorithms: Application to attosecond dynamics of high-harmonic generation*, Phys. Rev. A **70** (2004), 043404.

[19] Ronnie Kosloff, Stuart A. Rice, Pier Gaspard, Sam Tersigni and David Tannor, *Wavepacket Dancing: Achieving Chemical Selectivity by Shaping Light Pulses*, Chem. Phys. **139** (1989), 201–220.

[20] Ido Schaefer, *Quantum Optimal Control Theory of Harmonic Generation (Master's thesis)*, arXiv:1202.6520 (2012).

[21] Jan Boyke Schönborn, Peter Saalfrank, and Tillmann Klamroth, *Controlling the high frequency response of H_2 by ultra-short tailored laser pulses: A time-dependent configuration interaction study*, The Journal of Chemical Physics **144** (2016), no. 4, 044301.

[22] J. Solanpää, J. A. Budagosky, N. I. Shvetsov-Shilovski, A. Castro, A. Rubio, and E. Räsänen, *Optimal control of high-harmonic generation by intense few-cycle pulses*, Phys. Rev. A **90** (2014), 053402.

[23] Paolo Villoresi, Stefano Bonora, Michele Pascolini, Luca Poletto, Giuseppe Tondello, Caterina Vozzi, Mauro Nisoli, Giuseppe Sansone, Salvatore Stagira, and Sandro De Silvestri, *Optimization of high-order harmonic generation by adaptive control of a sub-10-fs pulse wave front*, Optics letters **29** (2004), no. 2, 207–209.

[24] J Werschnik and E K U Gross, *Quantum optimal control theory*, Journal of Physics B: Atomic, Molecular and Optical Physics **40** (2007), no. 18, R175–R211.

[25] Carsten Winterfeldt, Christian Spielmann, and Gustav Gerber, *Colloquium: Optimal control of high-harmonic generation*, Rev. Mod. Phys. **80** (2008), 117–140.

הפרופגציה של פונקציית הגל מבוצעת ע״י שיטה חדשה בעלת יעילות ודיוק גבוהים. הגישה העומדת ביסודה של שיטה זו מבוססת על שילוב שיקולים גלובליים ולוקליים עבור הקירוב של ההתפתחות בזמן של המערכת. שיטת הפרופגציה מותאמת לדינמיקה הלא-הרמיטית המאפיינת את המערכת תחת השפעת הפוטנציאל הבולע המרוכב (פרק 3). אנו מדגימים את היישום של שיטה זו עבור הדינמיקה האופיינית לבעיית יצירת ההרמוניות הגבוהות.

בפרסום המוקדם יותר המוצג בתזה, נעשה שימוש בשיטת אופטימיזציה פשוטה יחסית (פרק 2). בפרסום המאוחר יותר המוצג כאן (פרק 4), אנו מעדיפים להשתמש בשיטת אופטימיזציה מתוחכמת יותר (BFGS). הדבר נדרש בשל המורכבות היחסית של בעיית האופטימיזציה של יצירת ההרמוניות הגבוהות ביחס לבעיות הפשוטות יותר של יצירת הרמוניות מסדר נמוך, שהן הנושא המרכזי בפרק 2. התעורר צורך בהתאמה של פרטים מסויימים של שיטת האופטימיזציה לבעייה הנוכחית.

במחקר שלפנינו, אנו מדגימים את השיטה שפותחה הן עבור בעיות של יצירת הרמוניות מסדר נמוך והן עבור בעיות של יצירת הרמוניות גבוהות. התוצאות מדגימות שליטה בתחום התדרים של פולס הלייזר המאלץ, הגברה של הפליטה בתדרים נבחרים, וכן הגבלה של יוניזציה בלתי הפיכה של המערכת. במבחן של השוואה של הפולסים האופטימליים לפולסים טיפוסיים, ניתן להווכח כי קיימת הגברה משמעותית של הפליטה בתדרים הנבחרים; זאת, כאשר נעשה שימוש חסכוני בהרבה באנרגיה, וכן במחיר של שיעורים נמוכים יותר של יוניזציה בלתי הפיכה.

התוצאות מדגימות גם את היכולת של יצירת פולסים אופטימליים השוברים את כללי הברירה האופייניים לתהליך יצירת ההרמוניות הגבוהות, כללי ברירה המגבילים את ספקטרום הפליטה להרמוניות אי זוגיות בלבד. הדבר מתאפשר הודות לשבירת הסימטריה האופיינית ליצירת ההרמוניות הגבוהות תחת השפעתם של פולסי לייזר טיפוסיים.

תקציר

יצירת הרמוניות גבוהות (High Harmonic Generation) הינה תהליך שבו התדר של קרינה אלקטרומגנטית המוקרנת על מדיום לא-לינארי מוכפל לסדרים גבוהים ע״י המדיום. יצירת הרמוניות גבוהות משמשת כמקור לקרינה אלקטרומגנטית קוהרנטית בתחומי האולטרא-סגול הרחוק (XUV) וקרני ה-X הקרובות (soft X-ray). יישומים חשובים לקרינה קוהרנטית בתחומים אלו הם עבור ייצור לייזר אטו-שניות, וכן ניסויים בפיסיקה אטומית המתבצעים בתחום האנרגיות הגבוהות. בימינו, יצירת הרמוניות גבוהות הינה המקור היחיד לקרינה קוהרנטית בתחומי התדרים הנ״ל שהינו ישים לשימוש במעבדות המחקר המצויות.

היעילות של תהליך יצירת ההרמוניות הגבוהות היא נמוכה למדי. בשל כך, התפתח מחקר של אופטימיזציה של התהליך, בהתבסס על תחום המחקר המתפתח של בקרה של מערכות קוונטיות באמצעות קרינה אלקטרומגנטית קוהרנטית (quantum coherent control). גישה שהופעלה במידה של הצלחה היא אופטימיזציה של תהליך הבקרה ע״י שימוש באלגוריתם גנטי. נעשה שימוש בגישה זו הן במחקרים תיאורטיים והן במחקרים ניסיוניים. עם זאת, גישה זו הינה מוגבלת למדי בשל היעילות הנמוכה המאפיינת את תהליך האופטימיזציה של האלגוריתם הגנטי.

תורת הבקרה האופטימלית במערכות קוונטיות (Quantum Optimal Control Theory) הינה השיטה התיאורטית המוצלחת ביותר הקיימת בימינו לאופטימיזציה של תהליכי בקרה במערכות קוונטיות. שיטה זו מבוססת על ניסוח של בעיית הבקרה כבעיית מקסימום, ע״י שימוש בחשבון ואריאציה. תהליך האופטימיזציה מבוסס על אינפורמציה הנגזרת מהמשוואות אודות הגרדיינט של משטח האופטימיזציה.

במחקר זה, מפותחת שיטה תיאורטית לאופטימיזציה של תהליך יצירת ההרמוניות הגבוהות ע״י שימוש בתורת הבקרה האופטימלית במערכות קוונטיות. תהליך האופטימיזציה מבוסס על חיפוש נומרי של פרופיל של פולס לייזר עבור הגברה של הרמוניות נבחרות. ההרכב הספקטרלי של פולס הלייזר האופטימלי מבוסס על תחום תדרים שהוגדר מראש עבור מקור הלייזר.

הדרישות שהוצבו עבור בעיית הבקרה כוללות מקסימיזציה של הפליטה של המערכת בתדר נבחר או אזור ספקטרלי, מינימיזציה של האנרגיה הכוללת של פולס הלייזר המאלץ, וכן מינימיזציה של יוניזציה בלתי הפיכה של המערכת.

בשלב הראשון, נטפל בבעיה הכללית יותר של אופטימיזציה של תהליכי יצירת הרמוניות (לאו דווקא הרמוניות גבוהות). לאחר הפיתוח של העקרונות הבסיסיים, נפנה לטיפול הספציפי יותר בבעיה של יצירת הרמוניות גבוהות, שם כמה נושאים דורשים טיפול מיוחד.

אחד האתגרים המרכזיים הוא סימולציה אמינה של הדינמיקה של יצירת ההרמוניות הגבוהות, הידועה כבעייה קשה. לשם כך נדרשים כלים נומריים בעלי דיוק גבוה.

במחקר הנוכחי, תנאי השפה הבולעים מיושמים ע״י פוטנציאל בולע מרוכב. אנו מפעילים שיטת אופטימיזציה לבניית פוטנציאל בולע שממזער אפקטים של החזרה והעברה של פונקציית הגל ע״י הפוטנציאל.

עבודה זו נעשתה בהדרכתו של

פרופ׳ רוני קוזלוב

יצירת הרמוניות גבוהות באמצעות תורת הבקרה האופטימלית במערכות קוונטיות

חיבור לשם קבלת תואר דוקטור לפילוסופיה

מאת עידו שפר

הוגש לסנט האוניברסיטה העברית בירושלים

5/19