

# Parallel transport dynamics for mixed quantum states with applications to time-dependent density functional theory

Dong An<sup>a</sup>, Di Fang<sup>b,\*</sup>, Lin Lin<sup>b,c</sup>

<sup>a</sup> Joint Center for Quantum Information and Computer Science (QuICS), University of Maryland, College Park, MD 20742, USA

<sup>b</sup> Department of Mathematics and Challenge Institute for Quantum Computation, University of California, Berkeley, CA 94720, USA

<sup>c</sup> Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA



## ARTICLE INFO

### Article history:

Received 4 June 2021

Received in revised form 11 October 2021

Accepted 9 November 2021

Available online 19 November 2021

### Keywords:

Quantum dynamics

Parallel transport gauge

Metallic systems

TDDFT

## ABSTRACT

Direct simulation of the von Neumann dynamics for a general (pure or mixed) quantum state can often be expensive. One prominent example is the real-time time-dependent density functional theory (rt-TDDFT), a widely used framework for the first principle description of many-electron dynamics in chemical and materials systems. Practical rt-TDDFT calculations often avoid the direct simulation of the von Neumann equation, and solve instead a set of Schrödinger equations, of which the dynamics is equivalent to that of the von Neumann equation. However, the time step size employed by the Schrödinger dynamics is often much smaller. In order to improve the time step size and the overall efficiency of the simulation, we generalize a recent work of the parallel transport (PT) dynamics for simulating pure states [An, Lin, Multiscale Model. Simul. 18, 612, 2020] to general quantum states. The PT dynamics provides the optimal gauge choice, and can employ a time step size comparable to that of the von Neumann dynamics. Going beyond the linear and near adiabatic regime in previous studies, we find that the error of the PT dynamics can be bounded by certain commutators between Hamiltonians, density matrices, and their derived quantities. Such a commutator structure is not present in the Schrödinger dynamics. We demonstrate that the parallel transport-implicit midpoint (PT-IM) method is a suitable method for simulating the PT dynamics, especially when the spectral radius of the Hamiltonian is large. The commutator structure of the error bound, and numerical results for model rt-TDDFT calculations in both linear and nonlinear regimes, confirm the advantage of the PT dynamics.

© 2021 Elsevier Inc. All rights reserved.

## 1. Introduction

Consider the problem of solving a finite dimensional, (possibly) nonlinear von Neumann equation

$$i\partial_t \rho(t) = [H(t, \rho(t)), \rho(t)], \quad \rho(0) = \rho_0, \quad (1.1)$$

where  $\rho_0 \in \mathbb{C}^{N_g \times N_g}$  is a Hermitian matrix satisfying  $\rho_0^2 \preceq \rho_0$ . Here  $[A, B] = AB - BA$  is the commutator of  $A$  and  $B$ , and  $A \preceq B$  means that  $A - B$  is a negative semidefinite matrix. The initial quantum state  $\rho_0$  is called a pure state if  $\rho_0^2 = \rho_0$ , and a mixed state if  $\rho_0^2 \prec \rho_0$ . Eq. (1.1) can be used to describe the dynamics of a closed quantum system in a very general

\* Corresponding author.

E-mail addresses: [dongan@umd.edu](mailto:dongan@umd.edu) (D. An), [difang@berkeley.edu](mailto:difang@berkeley.edu) (D. Fang), [linlin@math.berkeley.edu](mailto:linlin@math.berkeley.edu) (L. Lin).

setting, and we allow the time-dependent Hamiltonian  $H(t, \rho(t)) \in \mathbb{C}^{N_g \times N_g}$  to have a nonlinear dependence on entries of  $\rho(t)$ . One prominent application is the real-time time-dependent density functional theory (rt-TDDFT) [39,45,35,43], which is one of the most widely used techniques for studying ultrafast properties of electrons, and has resulted in a variety of applications in quantum physics, chemistry, and materials science.

In practice,  $\rho_0$  is often of low-rank, or can be very well approximated by a low-rank matrix. For simplicity, let the rank of  $\rho_0$  be denoted by  $N$  and assume  $N \ll N_g$ . The von Neumann dynamics neglects the low-rank structure and propagates  $\rho(t)$  directly as a dense matrix. For rt-TDDFT calculations with a fine discretization scheme (e.g. the planewave basis set, or the finite difference discretization),  $N_g$  can be  $10^6$  or larger, and the direct propagation of Eq. (1.1) becomes extremely expensive. In this case, the von Neumann dynamics is often replaced by a set of nonlinear Schrödinger equations (see Eq. (2.1)), and the simulation variables become the electron wavefunctions described by a much smaller matrix  $\Psi(t) \in \mathbb{C}^{N_g \times N}$ . However, such a rank reduction can come at the cost of the time step size, denoted by  $h$ . In many applications,  $h$  for the von Neumann dynamics (1.1) can be chosen to be at the sub-femtosecond scale ( $1 \text{ fs} = 10^{-15} \text{ s}$ ), while  $h$  for the Schrödinger dynamics needs to be sub-attosecond scale ( $1 \text{ as} = 10^{-18} \text{ s}$ ) [6,40,12].

This attosecond-femtosecond time scale separation has inspired the development of new numerical methods in the past few years [44,32,18,47,1]. In fact, when  $\rho_0$  is a pure state, such a time scale separation is *unphysical* and originates purely from a gauge choice of electron wavefunctions in the Schrödinger representation. In the simplest case, when the rank of the pure state  $\rho_0$  is 1, the gauge is only a time-dependent complex phase factor. When the rank of the pure state  $\rho_0$  is  $N$ , the gauge is an  $N \times N$  unitary matrix. The gauge choice does not affect  $\rho$ . In other words, a gauge transformed dynamics may solve a different set of variables from those in the Schrödinger dynamics, but results in the same physical observables.

Given a pure initial state, among all possible gauge choices, the parallel transport (PT) gauge [18,1] yields the slowest gauge-transformed dynamics at any given time. Compared to the Schrödinger dynamics, the time step  $h$  in the PT dynamics can be chosen to be much larger and is comparable to that of the von Neumann dynamics Eq. (1.1). When combined with implicit integrators (such as the Crank-Nicolson method or the implicit midpoint rule), the PT dynamics has been applied to rt-TDDFT simulations for real materials with thousands of atoms at the level of semi-local exchange-correlation functionals (such as the Perdew–Burke–Ernzerhof [36] functional) [18] and hybrid exchange-correlation functionals (such as the Heyd–Scuseria–Ernzerhof [16] functional) [19,20].

In previous works, the PT dynamics was derived for a pure initial state, and its efficiency has been justified in the linear, near adiabatic regime [1] in terms of a singularly perturbed linear system. The pure initial state is suitable for describing molecules and insulating materials at zero temperature. On the other hand, in practice, the initial state is often a low-energy excited state [8,10,4], or a thermal state [46] especially for metallic systems. This inspires us to consider the most general setting when  $\rho_0$  is given by a mixed state (for instance, the occupation number of  $\rho_0$  is given by the Fermi-Dirac distribution).

### Contribution:

By assuming a dynamical low-rank factorization  $\rho(t) \approx \Phi(t)\sigma(t)\Phi^\dagger(t)$ , where  $\Phi(t) \in \mathbb{C}^{N_g \times N}$  and  $\sigma(t) \in \mathbb{C}^{N \times N}$ , we derive the PT dynamics in terms of its low-rank factors  $\Phi(t), \sigma(t)$ . The PT dynamics with a pure initial state is recovered by setting  $\sigma(t) = I_N$ . When the spectral radius of the Hamiltonian is large, the time step  $h$  is simultaneously constrained by accuracy and stability requirements, and implicit integrators are more suited for efficient propagation of the PT dynamics. Using the implicit midpoint (IM) rule (also known as the second order Gauss-Legendre method, GL2) as an example, we derive the discretized numerical scheme, and prove that the resulting PT-IM scheme has certain orthogonality and trace-preserving properties.

We then derive a new error bound for the discretized PT dynamics. Instead of relying on the linear quantum adiabatic theorem to obtain an *a priori* error bound of the solution, our new error bound expresses the local truncation error directly in terms of the Hamiltonian, density matrix, and their derived quantities. Our analysis shows that an upper bound of the local truncation error of PT dynamics only involves certain commutators between the Hamiltonian (or its time derivatives) and the density matrix (or the associated spectral projector), while that of the Schrödinger gauge involves additional terms lacking such commutator structures. Using the commutator type error bound, in the near adiabatic regime when the *a priori* estimate is available from the quantum adiabatic theorem, our new result shows the PT dynamics gains one extra order of accuracy in terms of the singularly perturbed parameter  $\epsilon$  than the Schrödinger dynamics, which reproduces the previous result [1]. Recently, the quantum adiabatic theorem has been extended to certain weakly nonlinear systems [9,11]. Our commutator type error analysis can be directly combined with such analysis leading to results comparable to that of [1] in the weakly nonlinear regime. Away from the near adiabatic regime, the commutator scaling of the PT dynamics can still lead to a significantly smaller error than that of the Schrödinger dynamics. We illustrate the numerical performance of the PT dynamics for a number of one-dimensional model metallic systems, which also verifies the effectiveness of the new error bound.

### Related works:

Numerical integrators for rt-TDDFT simulation following the Schrödinger dynamics is a well-studied subject (see an early paper [6], and also [12,38] for recent comparative studies of a variety of standard numerical integrators), but the importance and the benefit of gauge-transformed dynamics have only been realized recently (see [47] for another type of gauge-transformed dynamics using Wannier functions).

At the continuous level, the PT dynamics is a special case of the dynamical low-rank approximation (DLRA) developed by Lubich *et al.* (see [26,30] for examples; DLRA is intimately related to the Dirac-Frenkel/McLachlan variational principle in the physics literature). The basic strategy of DLRA is to update a low-rank decomposition (such as eigenvalue or singular value decomposition) of a large matrix (in this case  $\rho(t)$ ) on the fly. For a mixed initial state, a direct application of DLRA involves  $\sigma^{-1}(t)$  in the equation of the low-rank factors, which in general can be a source of numerical instability [26,31]. Our derivation of the PT dynamics with a mixed initial state uses the structure of the von Neumann equation and can be viewed as a simplified derivation of DLRA. It also naturally shows that the pathological term  $\sigma^{-1}(t)$  does not appear, so the PT dynamics is numerically stable even if one overestimates the numerical rank of  $\rho(t)$ .

Regarding the time discretization, existing works of DLRA mostly use explicit integrators, although the possibility of using implicit integrators has also been mentioned in certain settings [31]. Our previous studies suggest that for rt-TDDFT calculations, the combined use of the PT dynamics and implicit integrators is the key for efficient propagation in real chemical and materials systems [18,1]. The PT dynamics with a mixed initial state can also be viewed as a special case of the low-rank approximation for solving Lindblad equations by Le Bris *et al.* [27,28] (since the von Neumann equation can be viewed as the Lindblad equation without the decoherence operator), which is also derived independently of DLRA. It is worth pointing out that [27] introduces an arbitrary Hermitian matrix that can be freely determined. We demonstrate that in the context of the von Neumann dynamics, setting this arbitrary matrix to  $H(t)$  (the instantaneous Hamiltonian matrix), and 0 (the zero matrix) leads to the Schrödinger dynamics and the PT dynamics, respectively.

### Organization:

The rest of the paper is organized as follows. In Section 2, we introduce some preliminaries of rt-TDDFT and the PT dynamics with a pure initial state. We derive the PT dynamics with a mixed initial state in Section 3, and an implicit numerical propagator for the PT dynamics in Section 4. Section 5 analyzes the numerical errors of the PT and the Schrödinger dynamics. Finally, we validate the error analysis with numerical results in Section 6. For completeness, an alternative derivation of the PT dynamics that explicitly uses the structure of the tangent manifold (which is also a simplified derivation of [27]) is given in Appendix A. Appendix B describes Anderson's mixing method for solving the set of nonlinear equations in the PT-IM method, and provides a brief analysis of the computational cost.

## 2. Preliminaries

In this section, we briefly review the key idea of deriving PT dynamics with a pure initial state [1]. In the setting with a pure initial state, real-time time-dependent density functional theory (rt-TDDFT) solves the following set of Schrödinger equations

$$i\partial_t \Psi(t) = H(t, \rho(t))\Psi(t), \quad \Psi(0) = \Psi_0. \quad (2.1)$$

Here  $\Psi(t) = [\psi_1(t), \dots, \psi_N(t)]$  is the collection of electron wavefunctions (also called electron orbitals), and the number of columns  $N$  is equal to the number of electrons denoted by  $N_e$  (spin degeneracy omitted). The initial set of wavefunctions satisfy the orthonormality condition  $\Psi(0)^\dagger \Psi(0) = I_N$ . Here  $A^\dagger$  denotes the Hermitian conjugate of a matrix or vector  $A$ . The density matrix is  $\rho(t) = \Psi(t)\Psi^\dagger(t) \equiv \sum_{i=1}^N \psi_i(t)\psi_i^\dagger(t)$ .

Throughout the paper we are concerned with time propagation instead of spatial discretization. Unless otherwise specified, Eq. (2.1) represents a discrete, finite dimensional quantum system, i.e.  $H(t, \rho)$  is a Hermitian matrix with finite dimension  $N_g$ . If the quantum system is spatially continuous, we may first find a set of orthonormal basis functions and expand the continuous wavefunction under this basis. Then after a Galerkin projection, Eq. (2.1) becomes an  $N_g$ -dimensional quantum system, and  $\psi_j(t)$  represents the coefficient vector under the basis for the  $j$ -th wavefunction.

The time-dependent Hamiltonian operator  $H(t, \rho(t))$  is Hermitian for all  $t$  and  $\rho$ , and its precise form is not important for the purpose of this paper. Starting from a pure initial state  $\rho_0 := \rho(0)$  (that is when  $\rho_0^2 = \rho_0 = \Psi(0)\Psi(0)^\dagger$ ), the orthogonality condition  $\Psi(t)^\dagger \Psi(t) = I_N$  is satisfied for all  $t \geq 0$ , and hence  $\rho(t)$  is a pure state for all  $t$  satisfying  $\rho^2(t) = \rho(t)$ . Throughout the paper, we may use the notations  $\partial_t \rho = \rho_t = \dot{\rho}$  interchangeably for the time-derivatives. For composite functions such as  $H(t, \rho(t))$ , we use the notation  $\dot{H} := \frac{d}{dt} H(t, \rho(t)) = H_t + H_\rho \rho_t$ , where the tensor contractions are defined such that the chain rule holds. For example, the tensor contraction between the 4-tensor  $H_\rho$  and the matrix  $\rho_t$  is defined such that the chain rule  $\frac{d}{dt} H(t, \rho(t)) = H_t + H_\rho \rho_t$  follows the element-wise operation

$$\frac{d}{dt} H_{ij}(t, \rho(t)) = \partial_t H_{ij}(t, \rho(t)) + \sum_{k,l} \frac{\partial H_{ij}}{\partial \rho_{kl}}(t, \rho(t)) \frac{\partial \rho_{kl}(t)}{\partial t}.$$

The set of Schrödinger equations (2.1) is equivalent to the von Neumann dynamics (1.1). Note that if we right multiply  $\Psi(t)$  by a time-dependent unitary matrix  $U(t) \in \mathbb{C}^{N \times N}$  and let  $\Phi(t) = \Psi(t)U(t)$ , then

$$\rho(t) = \Psi(t)\Psi^\dagger(t) = \Phi(t) \left[ U^\dagger(t)U(t) \right] \Phi^\dagger(t) = \Phi(t)\Phi^\dagger(t). \quad (2.2)$$

The unitary rotation matrix  $U(t)$  is called the gauge matrix, and Eq. (2.2) indicates that the density matrix is *gauge-invariant*. In particular, the choice  $U(t) = I_N$  is referred to as the Schrödinger gauge.

Since all physical observables can be derived from the von Neumann equation (1.1) and the density matrix  $\rho(t)$ , the choice of the gauge matrix  $U(t)$  has no measurable effects. On the other hand, the gauge matrix introduces additional degrees of freedom, and can oscillate at a different time scale from that of the corresponding wavefunctions. It is then desirable to optimize the gauge matrix, so that the transformed wavefunctions  $\Phi(t)$  vary *as slowly as possible*, without changing the density matrix. This results in the following variational problem

$$\min_{U(t)} \|\dot{\Phi}\|_F^2, \text{ s.t. } \Phi(t) = \Psi(t)U(t), U^\dagger(t)U(t) = I_N. \quad (2.3)$$

Here  $\|\dot{\Phi}\|_F^2 := \text{Tr}[\dot{\Phi}^\dagger \dot{\Phi}]$  measures the Frobenius norm of the time derivative of the transformed orbitals.

The minimizer of (2.3), in terms of  $\Phi$ , satisfies the following equation

$$\rho \dot{\Phi} = 0. \quad (2.4)$$

We refer readers to [1] for the derivation. Eq. (2.4) has an intuitive explanation that the optimal dynamics should minimize the “internal” rotations within the range of  $\rho$ . Eq. (2.4) implicitly defines a gauge choice for each  $U(t)$ , and this gauge is called the *parallel transport gauge*. The name “parallel transport” comes from that  $\Phi(t)$  can be identified as the unique horizontal lift [34] of  $\rho(t)$  from the Grassmann manifold to the Stiefel manifold, starting from the initial condition  $\Psi_0$ . This will be further explained in Section 5.

From Eq. (2.4), the governing equation of  $\Phi(t)$  can be concisely written down as

$$i\partial_t \Phi(t) = H(t, \rho(t))\Phi(t) - \Phi(t)(\Phi^\dagger(t)H(t, \rho(t))\Phi(t)), \quad \Phi(0) = \Psi_0, \quad (2.5)$$

where  $\rho(t) = \Phi(t)\Phi^\dagger(t)$ . Notice that Eq. (2.5) introduces one extra term compared to the original dynamics Eq. (2.1) under Schrödinger gauge, and directly provides a self-contained definition of the transformed wavefunctions under the optimal gauge. In practice, we can directly solve Eq. (2.5) by numerical schemes to approximate the dynamics, instead of computing the PT gauge explicitly.

To observe the advantage of the parallel transport dynamics, consider the extreme case that each column of  $\Psi_0$  is already an eigenstate of  $H(0)$  and  $H(t, \rho(t)) \equiv H(0)$  is a time-independent matrix. Then Eq. (2.5) is reduced to

$$i\partial_t \Phi(t) = 0.$$

Hence  $\Phi(t) = \Phi(0)$  holds for all  $t \geq 0$ , while each column of the solution Schrödinger dynamics (2.1) rotates with a time-dependent phase factor. For less trivial dynamics, the temporal oscillation of gauge-transformed wavefunctions  $\Phi(t)$  can still be significantly slower than that of  $\Psi(t)$ .

### 3. Parallel transport dynamics with a mixed initial state

In this section, we first discuss the rt-TDDFT set-up with a mixed initial state and its corresponding Schrödinger dynamics, and then provide a derivation of the PT dynamics with a mixed initial state, which is equivalent to Schrödinger dynamics under a gauge transformation.

#### 3.1. Schrödinger dynamics with a mixed initial state

In rt-TDDFT calculations, the pure initial state can be used for simulating insulating systems starting from the ground state, or a well-defined excited state. In many other cases the initial state should be a mixed state. For instance, for metallic systems at finite temperature, the initial state often takes the form of the Fermi-Dirac distribution

$$\rho(0) = (1 + \exp(\beta(H(0) - \mu)))^{-1}, \quad (3.1)$$

where  $\beta = 1/(k_B T)$ ,  $k_B$  is the Boltzmann constant,  $T$  is the temperature. The chemical potential  $\mu$  is a Lagrange multiplier, which should be adjusted to satisfy the normalization condition

$$\text{Tr}[\rho(0)] = N_e, \quad (3.2)$$

where  $N_e$  is the number of electrons. If we diagonalize  $H(0)$  according to  $H(0)\psi_i(0) = \varepsilon_i(0)\psi_i(0)$ , then the occupation number

$$s_i(0) := \langle \psi_i(0) | \rho(0) | \psi_i(0) \rangle = (1 + \exp(\beta(\varepsilon_i(0) - \mu)))^{-1}.$$

Hence when  $\beta$  is large (e.g. at room temperature 300 K,  $\beta \approx 10^3$  in the atomic unit),  $s_i(0)$  is very close to 0 when  $\varepsilon_i(0) - \mu \gg \beta^{-1}$ . Therefore,  $\rho(0)$  can be very well approximated by a low rank matrix, with its approximate rank denoted by  $N$ . In particular, the singular values of the Fermi-Dirac distribution (see Fig. 1b) decay exponentially beyond  $\mu$  and hence one can truncate the singular values according to a certain energy threshold. In other words, we can set

$$\rho(0) = \sum_{i=1}^N \psi_i(0) s_i(0) \psi_i^\dagger(0) = \Psi(0) \sigma(0) \Psi^\dagger(0),$$

with the chemical potential  $\mu$  slightly adjusted so that the normalization condition (3.2) is still satisfied. Here  $\sigma_0 := \sigma(0) = \text{diag}[s_1(0), \dots, s_N(0)]$  is a diagonal matrix. Since the occupation number satisfies  $0 < s_i(0) < 1$ , we have  $N > N_e$ , and  $\rho^2(0) \prec \rho(0)$ . We also assume  $N_g \gg N$ . Note that different from the case of a pure state, the density matrix  $\rho$  for a mixed state is no longer a projection operator.

If  $\Psi(t)$  is the solution to the nonlinear Schrödinger equation (2.1), then

$$\rho(t) = \Psi(t) \sigma_0 \Psi^\dagger(t), \quad (3.3)$$

is the unique solution of Eq. (1.1) (viewed as a large ODE system) with the initial state  $\rho(0)$ . Hence in practice, we only need to solve Eq. (2.1) in the same way as for the pure initial state, but weigh the contribution of each time-dependent vector  $\psi_i(t)$  always by the *initial* occupation number  $\sigma_i(0)$ . This fact that the occupation number  $\sigma(t)$  remains as a constant matrix  $\sigma_0$  can also be derived directly (see Eq. (A.9) in Appendix A). However, similar to the case with a pure initial state, Eq. (2.1) can require a relatively small time step size.

### 3.2. PT dynamics with a mixed initial state

Note that we may still apply a gauge matrix  $U(t) \in \mathbb{C}^{N \times N}$  and define  $\Phi(t) = \Psi(t)U(t)$  with initial condition  $U(0) = I_N$ . In such a case, we must also redefine the occupation number matrix as

$$\sigma(t) = U^\dagger(t) \sigma_0 U(t), \quad (3.4)$$

so that

$$\rho(t) = \Phi(t) \sigma(t) \Phi^\dagger(t) \quad (3.5)$$

is satisfied. Here  $\sigma(t)$  is now a Hermitian matrix of size  $N$  and may no longer be diagonal for  $t > 0$ . We would like to solve again the optimization problem in Eq. (2.3) so that the gauge-transformed wavefunctions  $\Phi(t)$  vary as slowly as possible. For simplicity, we may also define a gauge-invariant projector

$$P(t) = \Psi(t) \Psi^\dagger(t) = \Phi(t) \Phi^\dagger(t),$$

so that

$$P\Phi = \Phi, \quad \dot{P}\Phi + P\dot{\Phi} = \dot{\Phi} \quad (3.6)$$

The objective in the optimization Eq. (2.3) becomes

$$\|\dot{\Phi}\|_F^2 = \|P\dot{\Phi}\|_F^2 + \|(I - P)\dot{\Phi}\|_F^2.$$

The second term is in fact independent of the choice of the gauge (or the wave function  $\Phi$ ), because

$$\|(I - P)\dot{\Phi}\|_F^2 = \|\dot{P}\Phi\|_F^2 = \text{Tr}(\Phi^\dagger \dot{P}^\dagger \dot{P}\Phi) = \text{Tr}(\dot{P}^\dagger \dot{P}\Phi\Phi^\dagger) = \text{Tr}(\dot{P}^\dagger \dot{P}P),$$

where we used the fact that  $\text{Tr}(AB) = \text{Tr}(BA)$ . Therefore, the optimization problem again leads to the PT condition

$$P\dot{\Phi} = 0. \quad (3.7)$$

Plugging back into Eq. (3.6) we obtain

$$i\dot{\Phi} = i\dot{P}\Phi = [H, P]\Phi = H\Phi - PH\Phi,$$

or equivalently,

$$i\partial_t\Phi(t) = (I - P(t))H(t, \rho(t))\Phi(t).$$

Here the identity matrix is given as  $I = I_{N_g}$ , and we have used that  $P(t)\Phi(t) = \Phi(t)$ .

In order to close the equation, it remains to identify the equation of motion of  $\sigma(t)$ . First, by differentiating the equation  $\Psi(t)U(t) = \Phi(t)$  and using (2.5), we may derive the dynamics of the gauge  $U(t)$ , i.e.

$$(i\partial_t\Psi(t))U(t) + \Psi(t)(i\partial_tU(t)) = H(t, \rho(t))\Psi(t)U(t) - \Psi(t)\Psi^\dagger(t)H(t, \rho(t))\Psi(t)U(t).$$

This gives

$$i\partial_t U(t) = -(\Psi^\dagger(t)H(t, \rho(t))\Psi(t))U(t). \quad (3.8)$$

By differentiating both sides of Eq. (3.4) and using Eq. (3.8), we have

$$\begin{aligned} i\partial_t \sigma(t) &= (i\partial_t U^\dagger(t))\sigma_0 U(t) + U^\dagger(t)\sigma_0(i\partial_t U(t)) \\ &= U^\dagger(t)(\Psi^\dagger(t)H(t, \rho(t))\Psi(t))\sigma_0 U(t) - U^\dagger(t)\sigma_0(\Psi^\dagger(t)H(t, \rho(t))\Psi(t))U(t) \\ &= \Phi^\dagger(t)H(t, \rho(t))\Phi(t)\sigma(t) - \sigma(t)\Phi^\dagger(t)H(t, \rho(t))\Phi(t) \\ &= [\Phi^\dagger(t)H(t, \rho(t))\Phi(t), \sigma(t)]. \end{aligned}$$

Therefore the equation of motion for  $\sigma(t)$  only depends on the slowly varying gauge-transformed wavefunctions  $\Phi(t)$ .

In summary, the parallel transport dynamics with a general initial state consists of the following set of equations

$$\begin{aligned} i\partial_t \Phi(t) &= (I - P(t))H(t, \rho(t))\Phi(t), \\ i\partial_t \sigma(t) &= [\Phi^\dagger(t)H(t, \rho(t))\Phi(t), \sigma(t)], \\ \rho(t) &= \Phi(t)\sigma(t)\Phi^\dagger(t), \quad P(t) = \Phi(t)\Phi^\dagger(t), \\ \Phi(0) &= \Psi_0, \quad \sigma(0) = \sigma_0. \end{aligned} \quad (3.9)$$

Eq. (3.9) gives a self-contained definition of the transformed wavefunctions  $\Phi(t)$  and the matrix  $\sigma(t)$  under the optimal gauge. Therefore, we can directly solve Eq. (3.9) to numerically approximate the state  $\rho(t)$  without computing the PT gauge matrix explicitly. Notice that, compared to the Schrödinger dynamics in which one can set  $\sigma(t) = \sigma(0)$ , the PT dynamics in Eq. (3.9) introduces one extra nonlinear term in the propagation of the wavefunctions, and enlarges the size of the ODE system via a non-trivial dynamics of the transformed  $\sigma(t)$ . This is different from the pure state setting where only an extra term in the equation of  $\Phi$  is added. However, due to the assumption that  $N_g \gg N$ , the increase of the number of variables by  $N^2$  due to  $\sigma(t)$  does not add too much overhead in the numerical simulation.

In Appendix A, we provide an alternative derivation of Eq. (3.9) using the tangent space formulation, which follows the derivation in [27,28] for Lindblad equations, and is more analogous to the derivation of the dynamical low-rank approximation. The derivation also provides an alternative perspective of the gauge choice in terms of an auxiliary Hamiltonian.

#### 4. Numerical propagation of the parallel transport dynamics

In order to solve Eq. (3.9) numerically, for simplicity we assume that a uniform time discretization  $t_n = nh$ , and  $h$  is the time step size. The numerical values of  $\Phi(t), \sigma(t), \rho(t), P(t)$  at time  $t = t_n$  are denoted by  $\Phi_n, \sigma_n, \rho_n, P_n$ , respectively, and we define  $H_n = H(t_n, \rho_n)$ . Previous studies suggested that when the spectral radius of  $H$  is large, the PT dynamics should be solved using implicit time integrators [18,1]. This allows one to use a time step much larger than  $\|H\|^{-1}$ , and the result from the PT dynamics can be much more accurate than that from the Schrödinger dynamics using the same step size.

In order to discretize the PT dynamics with a mixed initial state, we consider the implicit midpoint (IM) rule (also known as the Gauss-Legendre method of order 2). We introduce the shorthand notations

$$\Phi_{n+\frac{1}{2}} = \frac{1}{2}(\Phi_n + \Phi_{n+1}), \quad \sigma_{n+\frac{1}{2}} = \frac{1}{2}(\sigma_n + \sigma_{n+1}), \quad (4.1)$$

and accordingly

$$P_{n+\frac{1}{2}} = \Phi_{n+\frac{1}{2}}(\Phi_{n+\frac{1}{2}}^\dagger \Phi_{n+\frac{1}{2}})^{-1} \Phi_{n+\frac{1}{2}}^\dagger, \quad \rho_{n+\frac{1}{2}} = \Phi_{n+\frac{1}{2}} \sigma_{n+\frac{1}{2}} \Phi_{n+\frac{1}{2}}^\dagger, \quad H_{n+\frac{1}{2}} = H\left(t_{n+\frac{1}{2}}, \rho_{n+\frac{1}{2}}\right). \quad (4.2)$$

We remark that  $\rho_{n+\frac{1}{2}}$  is only a shorthand notation and may not be an admissible density matrix. In particular, even if  $\text{Tr}[\rho_n] = \text{Tr}[\rho_{n+1}] = N_e$  (see Proposition 1), we may not have  $\text{Tr}[\rho_{n+\frac{1}{2}}] = N_e$ . On the other hand,  $P_{n+\frac{1}{2}}$  is still a projector satisfying  $P_{n+\frac{1}{2}} \Phi_{n+\frac{1}{2}} = \Phi_{n+\frac{1}{2}}$ .

With these notations, the parallel transport-implicit midpoint scheme (PT-IM) reads

$$i\frac{\Phi_{n+1} - \Phi_n}{h} = (I - P_{n+\frac{1}{2}})H_{n+\frac{1}{2}}\Phi_{n+\frac{1}{2}}, \quad (4.3)$$

$$i\frac{\sigma_{n+1} - \sigma_n}{h} = \left[ \Phi_{n+\frac{1}{2}}^\dagger H_{n+\frac{1}{2}} \Phi_{n+\frac{1}{2}}, \sigma_{n+\frac{1}{2}} \right], \quad (4.4)$$

which form a set of nonlinear algebraic equations and need to be solved self-consistently. We can rewrite Eqs. (4.3) and (4.4) as

$$\begin{aligned} \Phi_{n+1} &= \Phi_n + \frac{h}{i}(I - P_{n+\frac{1}{2}})H_{n+\frac{1}{2}}\Phi_{n+\frac{1}{2}}, \\ \sigma_{n+1} &= \sigma_n + \frac{h}{i} \left[ \Phi_{n+\frac{1}{2}}^\dagger H_{n+\frac{1}{2}} \Phi_{n+\frac{1}{2}}, \sigma_{n+\frac{1}{2}} \right]. \end{aligned} \quad (4.5)$$

If we choose  $(\Phi_{n+1}, \sigma_{n+1})$  to be the unknowns and identify it with a vector  $x \in \mathbb{C}^{N_g N + N^2}$ , then Eqs. (4.3) and (4.4) can be viewed as a fixed point equation in the abstract form

$$x = T(x).$$

The structure of this fixed point problem resembles that of the self-consistent field iterations (SCF) in standard electronic structure calculations [33]. Here we use Anderson's mixing method [2] to solve this fixed problem (see Appendix B). Other numerical schemes for SCF iteration may also be applied [25,7,17,24,37] (also see [23] for general discussions on solving fixed point problems).

It may seem that the implementation of the PT dynamics has a large overhead compared to the direct propagation of the SD dynamics using an explicit integrator. When  $N \ll N_g$  and  $N$  is not very large, the cost of rt-TDDFT calculations is mostly given by the number of operations of the form  $H\Phi$ . Due to the removal of the fast oscillating gauge degrees of freedom, the time step in the PT-dynamics can be chosen to be much larger than that in the SD dynamics, which effectively reduces the number of  $H\Phi$  operations. The computational advantage has been demonstrated in rt-TDDFT calculations for real chemical systems for semi-local exchange-correlation functionals [18,1], and the advantage is even more pronounced for the more expensive hybrid exchange-correlation functionals calculations [19,20]. These calculations have been performed with a pure initial state, and we expect that similar savings can be observed in rt-TDDFT calculations for real systems. The computational cost per iteration of the Anderson mixing method is given in Appendix B.

The following proposition shows that PT-IM preserves the orthogonality as well as the trace condition.

**Proposition 1.** Assume  $\Phi_n^\dagger \Phi_n = I_N$ ,  $\sigma_n = \sigma_n^\dagger$ ,  $\text{Tr}[\sigma_n] = N_e$ , and that Eqs. (4.3) and (4.4) have a unique solution  $(\Phi_{n+1}, \sigma_{n+1})$ , then the solution satisfies

$$\Phi_{n+1}^\dagger \Phi_{n+1} = I_N, \quad (4.6)$$

and

$$\sigma_{n+1}^\dagger = \sigma_{n+1}, \quad \text{Tr}[\sigma_{n+1}] = N_e, \quad \text{Tr}[\sigma_{n+1}^2] = \text{Tr}[\sigma_n^2]. \quad (4.7)$$

As a consequence, we have  $\text{Tr}[\rho_{n+1}] = \text{Tr}[\rho_n] = N_e$ .

**Proof.** First, use the definition in (4.1) and apply  $\Phi_{n+\frac{1}{2}}^\dagger$  to both sides of (4.3), we obtain

$$\frac{i}{2h}(\Phi_{n+1}^\dagger \Phi_{n+1} - \Phi_n^\dagger \Phi_n) - \frac{i}{2h}(\Phi_{n+1}^\dagger \Phi_n - \Phi_n^\dagger \Phi_{n+1}) = \Phi_{n+\frac{1}{2}}^\dagger (I - P_{n+\frac{1}{2}}) H_{n+\frac{1}{2}} \Phi_{n+\frac{1}{2}} = 0.$$

On the left-hand side, the first term is anti-Hermitian and the second term is Hermitian. So both terms must vanish, and

$$\Phi_{n+1}^\dagger \Phi_{n+1} = \Phi_n^\dagger \Phi_n = I_N.$$

This proves Eq. (4.6).

Second, denote by  $\tilde{H} := \Phi_{n+\frac{1}{2}}^\dagger H_{n+\frac{1}{2}} \Phi_{n+\frac{1}{2}}$ , we may solve the equation

$$\sigma_{n+1} - \sigma_n = -\frac{ih}{2}[\tilde{H}, \sigma_n] - \frac{ih}{2}[\tilde{H}, \sigma_{n+1}]$$

to obtain  $\sigma_{n+1}$ . Applying the Hermite conjugation to both sides and using that  $\tilde{H}, \sigma_n$  are Hermitian matrices, we have

$$\sigma_{n+1}^\dagger - \sigma_n = -\frac{ih}{2}[\tilde{H}, \sigma_n] - \frac{ih}{2}[\tilde{H}, \sigma_{n+1}^\dagger].$$

The uniqueness of  $\sigma_{n+1}$  implies  $\sigma_{n+1} = \sigma_{n+1}^\dagger$ . Moreover, since the right-hand side of Eq. (4.4) is traceless, we have  $\text{Tr}[\sigma_{n+1}] = \text{Tr}[\sigma_n]$ .

Finally, applying  $\sigma_{n+\frac{1}{2}}$  from the left to both sides of Eq. (4.4), we have

$$\frac{i}{2h}(\sigma_{n+1}^2 - \sigma_n^2 - \sigma_{n+1}\sigma_n + \sigma_n\sigma_{n+1}) = \sigma_{n+\frac{1}{2}}[\tilde{H}, \sigma_{n+\frac{1}{2}}].$$

Since the right-hand side is traceless, by taking trace of both sides we obtain

$$\text{Tr}[\sigma_{n+1}^2] = \text{Tr}[\sigma_n^2].$$

This finishes the proof of the equalities in (4.7).  $\square$

Eq. (4.6) can be viewed as a consequence of the general fact that the PT-IM method preserves quadratic invariants, and in particular orthogonality constraints (see e.g. [14, pp 132] for a more general description of orthogonality preserving Runge-Kutta methods). Proposition 1 confirms that the PT-IM scheme preserves orthogonality of  $\Phi(t)$ , as well as the number of electrons.

## 5. Error analysis

In this section, we consider the numerical error of the PT-IM scheme for concreteness, and compare the form of the error terms with those from the Schrödinger dynamics. The error analysis can also be extended to other Runge-Kutta methods and linear multistep methods.

### 5.1. Error analysis of the PT dynamics

Before proceeding with the detailed error analysis, we first provide some abstract perspectives. Let the local truncation error be defined as  $e_k(X) = X(t_k) - \tilde{X}_k$ , where  $t_k = kh$  and  $\tilde{X}_k$  represents the numerical solution of  $X$  at the  $k$ -th step with previous step to be exactly  $X(t_{k-1})$ , where  $X$  is the concatenation of  $\Phi$  and  $\sigma$ , namely  $\begin{pmatrix} \Phi \\ \sigma \end{pmatrix}$ . Since IM is a second order method, the local truncation error can be bounded in terms of the third order derivatives [15]

$$\|e_k(X)\| \leq C \max_{t \in [t_{k-1}, t_k]} \|\partial_t^3 X(t)\| h^3. \quad (5.1)$$

Here  $C$  is an absolute constant depending only on the choice of the numerical scheme.

Note that  $P$  is a rank- $N$  projector, and can be identified with the Grassmann manifold  $\text{Gr}(N_g, N; \mathbb{C})$ , i.e. the  $N$ -dimensional subspace of  $\mathbb{C}^{N_g}$ . On the other hand, the gauge-transformed wavefunctions  $\Phi$  belongs to the Stiefel manifold  $\text{St}(N_g, N; \mathbb{C})$ , which is the set of first  $N$  columns of an  $N_g$ -dimensional unitary matrices. The Grassmann manifold is the quotient space of  $\text{St}(N_g, N; \mathbb{C})$  by  $\text{U}(N)$ , denoted by

$$\text{Gr}(N_g, N; \mathbb{C}) = \text{St}(N_g, N; \mathbb{C})/\text{U}(N)$$

The projector  $P(t)$  can be identified with a curve in  $\text{Gr}(N_g, N; \mathbb{C})$ , obtained by solving the von Neumann equation. On the other hand, the wavefunctions  $\Psi(t)$ ,  $\Phi(t)$  in the Schrödinger and the parallel transport gauge are *lifts* of the curve  $P(t)$  from the quotient space to  $\text{St}(N_g, N; \mathbb{C})$ . In particular,  $\Phi(t)$  can be identified as the unique *horizontal lift* [34] of  $P(t)$ , starting from the initial condition  $\Psi_0$  (which fixes a gauge choice initially). We have demonstrated that the parallel transport gauge yields the slowest dynamics in the sense of minimizing  $\|\partial_t \Phi\|_F$ . For simplicity, in the following discussions we will consider the operator norm  $\|\cdot\|$  for  $\Phi$ ,  $P$  and their time derivative. We expect that the size of the  $k$ -th order time derivative  $\|\partial_t^k \Phi\|$  should also be bounded by that of  $\|\partial_t^k P\|$ . On the other hand,  $\|\partial_t^k \Psi\|$  may not be bounded by  $\|\partial_t^k P\|$  due to the gauge matrix.

Recall the relation

$$P\Phi = \Phi, \quad P\partial_t\Phi = 0,$$

and this gives

$$\partial_t\Phi = (\partial_t P)\Phi.$$

Keep differentiating and obtain

$$\partial_t^2\Phi = [\partial_t^2 P + (\partial_t P)^2]\Phi, \quad \partial_t^3\Phi = [\partial_t^3 P + 2(\partial_t^2 P)(\partial_t P) + (\partial_t P)(\partial_t^2 P) + (\partial_t P)^3]\Phi.$$

Using the fact that  $\|\Phi\| = 1$ , we have

$$\|\partial_t\Phi\| \leq \|\partial_t P\|.$$

Similarly

$$\|\partial_t^2\Phi\| \leq \|\partial_t^2 P + (\partial_t P)^2\| \leq \|\partial_t^2 P\| + \|(\partial_t P)\|^2,$$

and

$$\|\partial_t^3\Phi\| \leq \|\partial_t^3 P\| + 3\|\partial_t^2 P\|\|\partial_t P\| + \|\partial_t P\|^3.$$

This implies that  $\|\partial_t^k\Phi\|$  is controlled by  $\|\partial_t^\ell P\|$  with  $\ell \leq k$ . On the other hand,

$$\sigma = \Phi^\dagger \rho \Phi$$

implies that the time derivative  $\|\partial_t^k \sigma(t)\|$  is controlled by  $\|\partial_t^\ell \Phi\|$  and  $\|\partial_t^\ell \rho\|$  with  $\ell \leq k$ . To be specific, a direct computation gives

$$\begin{aligned}\partial_t^3 \sigma = & (\partial_t^3 \Phi^\dagger) \rho \Phi + \Phi^\dagger (\partial_t^3 \rho) \Phi + \Phi^\dagger \rho (\partial_t^3 \Phi) + 6(\partial_t \Phi^\dagger)(\partial_t \rho)(\partial_t \Phi) \\ & + 3(\partial_t^2 \Phi^\dagger)(\partial_t \rho) \Phi + 3(\partial_t^2 \Phi^\dagger) \rho (\partial_t \Phi) + 3(\partial_t \Phi^\dagger)(\partial_t^2 \rho) \Phi + 3(\partial_t \Phi^\dagger) \rho (\partial_t^2 \Phi) + 3\Phi^\dagger (\partial_t^2 \rho)(\partial_t \Phi) \\ & + 3\Phi^\dagger (\partial_t \rho)(\partial_t^2 \Phi),\end{aligned}$$

and hence

$$\begin{aligned}\|\partial_t^3 \sigma\| \leq & 2\|\partial_t^3 \Phi\| + \|\partial_t^3 \rho\| + 6\|\partial_t \Phi^\dagger\| \|\partial_t \rho\| \|\partial_t \Phi\| \\ & + 6\|\partial_t^2 \Phi\| \|\partial_t \rho\| + 6\|\partial_t^2 \Phi\| \|\partial_t \Phi\| + 6\|\partial_t \Phi\| \|\partial_t^2 \rho\|,\end{aligned}$$

where we used the facts that  $\|\Phi\| = 1$  and  $\|\rho\| \leq 1$ .

To summarize, the error analysis of the PT dynamics is reduced to the estimate of  $\|\partial_t^k P\|$  and  $\|\partial_t^k \rho\|$ . In particular, for the analysis of PT-IM, we need  $k \leq 3$ .

**Lemma 2.** Suppose  $H(t, \rho)$  is continuously differentiable in terms of  $t$  and  $\rho$  up to second order. Then the derivatives of  $P$  satisfy

$$\|\partial_t P\| \leq \|[H, P]\|, \quad (5.2)$$

$$\|\partial_t^2 P\| \leq \|[H_t, P]\| + \|H_\rho [H, \rho]\| + \|[H, [H, P]]\|, \quad (5.3)$$

$$\begin{aligned}\|\partial_t^3 P\| \leq & \|[H_{tt}, P]\| + 2\|(H_t)_\rho [H, \rho]\| + \|H_{\rho\rho} ([H, \rho])^2\| + \|H_\rho [H_t, \rho]\| \\ & + \|H_\rho [H_\rho [H, \rho], \rho]\| + \|H_\rho [H, [H, \rho]]\| + 2\|[H_t, [H, P]]\| + 2\|[H_\rho [H, \rho], [H, P]]\| \\ & + \|[H, [H_t, P]]\| + \|[H, [H_\rho [H, \rho], P]]\| + \|[H, [H, [H, P]]]\|,\end{aligned} \quad (5.4)$$

where the subscripts denote the partial derivatives.

**Proof.** The first inequality is trivial. To prepare for the differentiation of  $P$ , we start by computing the derivatives of  $H$ . For notational simplicity, we use the subscripts to denote the partial derivative and omit the explicit  $(t, \rho(t))$  dependence in  $H$ . The first order derivative of  $H$  reads

$$\dot{H} := \frac{d}{dt} H(t, \rho(t)) = H_t + H_\rho \rho_t = H_t - iH_\rho [H, \rho], \quad (5.5)$$

and the second order derivative of  $H$  is given by

$$\begin{aligned}\ddot{H} := & \frac{d^2}{dt^2} H(t, \rho(t)) = \frac{d}{dt} H_t - i \frac{d}{dt} (H_\rho [H, \rho]) \\ & = H_{tt} + (H_t)_\rho \rho_t - i(H_t)_\rho [H, \rho] - iH_{\rho\rho} \rho_t [H, \rho] - iH_\rho [\dot{H}, \rho] - iH_\rho [H, \rho_t].\end{aligned}$$

It follows from  $i\partial_t \rho = [H, \rho]$  that

$$\begin{aligned}\ddot{H} = & H_{tt} - 2i(H_t)_\rho [H, \rho] - H_{\rho\rho} ([H, \rho])^2 \\ & - iH_\rho [H_t, \rho] - H_\rho [H_\rho [H, \rho], \rho] - H_\rho [H, [H, \rho]].\end{aligned} \quad (5.6)$$

The second order derivative of  $P$  becomes

$$\begin{aligned}\partial_t^2 P = & -i \frac{d}{dt} ([H(t, \rho(t)), P(t)]) = -i[\dot{H}, P] - i[H, \partial_t P] \\ & = -i[H_t, P] - [H_\rho [H, \rho], P] - [H, [H, P]],\end{aligned} \quad (5.7)$$

together with the fact that  $\|P\| \leq 1$ , we obtain (5.3). Similarly, the third order derivative of  $P$  can be computed explicitly via

$$\partial_t^3 P = -i[\ddot{H}, P] - 2i[\dot{H}, \partial_t P] - i[H, \partial_t^2 P].$$

Plugging in (5.5), (5.6) and (5.7), one obtains

$$\begin{aligned}\partial_t^3 P = & -i[H_{tt}, P] - 2[(H_t)_\rho [H, \rho], P] + i[H_{\rho\rho} ([H, \rho])^2, P] - [H_\rho [H_t, \rho], P] \\ & + i[H_\rho [H_\rho [H, \rho], \rho], P] + i[H_\rho [H, [H, \rho]], P] - 2[H_t, [H, P]] \\ & + 2i[H_\rho [H, \rho], [H, P]] - [H, [H_t, P]] + i[H, [H_\rho [H, \rho], P]] + i[H, [H, [H, P]]].\end{aligned}$$

Taking the norm yields the desired result.  $\square$

**Lemma 3.** Suppose  $H(t, \rho)$  is continuously differentiable in terms of  $t$  and  $\rho$  up to second order. The derivatives of  $\rho$  satisfy

$$\|\partial_t \rho\| \leq \|[H, \rho]\|, \quad (5.8)$$

$$\|\partial_t^2 \rho\| \leq \|[H_t, \rho]\| + \|[H_\rho[H, \rho], \rho]\| + \|[H, [H, \rho]]\|, \quad (5.9)$$

$$\begin{aligned} \|\partial_t^3 \rho\| &\leq \|[H_{tt}, \rho]\| + 2 \|[H_t]_\rho[H, \rho], \rho]\| + \|[H_\rho([H, \rho])^2, \rho]\| + \|[H_\rho[H_t, \rho], \rho]\| \\ &\quad + \|[H_\rho[H_\rho[H, \rho], \rho], \rho]\| + \|[H_\rho[H, [H, \rho]], \rho]\| + 2 \|[H_t, [H, \rho]]\| + \|[H, [H_t, \rho]]\| \\ &\quad + 2 \|[H_\rho[H, \rho], [H, \rho]]\| + \|[H, [H_\rho[H, \rho], \rho]]\| + \|[H, [H, [H, \rho]]]\|, \end{aligned} \quad (5.10)$$

where the subscripts denote the partial derivatives.

**Proof.** The proof is similar as Lemma 2 since  $\rho$  satisfies the equation  $i\partial_t \rho = [H, \rho]$ , which has the same form of that for  $P$ .  $\square$

Therefore, the local truncation errors of the PT dynamics can be bounded by terms involving commutators of  $[H, P]$ ,  $[H, \rho]$ ,  $[H_t, \rho]$ ,  $[H_{tt}, \rho]$ ,  $[H_t, P]$ ,  $[H_{tt}, P]$ .

## 5.2. Comparison to the Schrödinger dynamics

In this section, we discuss the local truncation error of the Schrödinger dynamics and the global errors of the PT and Schrödinger dynamics. The local truncation error can be summarized in the following lemma. Note that in the bound, we keep the wavefunction  $\Psi$  for the terms without commutator structures, such as  $\|H^3\Psi\|$ , instead of replacing it by the operator norm  $\|H^3\|$ , because the latter could be significantly larger than the former.

**Lemma 4.** For the IM scheme, the local truncation errors of Schrödinger dynamics (2.1) can be bounded as

$$\begin{aligned} \|e_k(\Psi)\| \leq C & \left( \|H^3\Psi\| + \|HH_t\Psi\| + 2\|H_tH\Psi\| + \|H_{tt}\Psi\| \right. \\ & + \|[H_\rho[H, \rho]]\| + 2\|[H_\rho[H, \rho]H]\| + 2\|[H_t]_\rho[H, \rho]\| + \|[H_\rho([H, \rho])^2]\| \\ & \left. + \|[H_\rho[H_t, \rho]]\| + \|[H_\rho[H_\rho[H, \rho], \rho]]\| + \|[H_\rho[H, [H, \rho]]]\| \right), \end{aligned} \quad (5.11)$$

for some constant  $C$  that does not depend on  $t_k, h$ .

**Proof.** It suffices to calculate the derivatives of  $\Psi$ . The second order derivative is computed as

$$\partial_t^2 \Psi = -iH\partial_t \Psi - i\dot{H}\Psi = -H^2\Psi - i\dot{H}\Psi$$

and the third order derivative can be computed as

$$\begin{aligned} \partial_t^3 \Psi &= -iH\ddot{\Psi} - 2i\dot{H}\dot{\Psi} - i\ddot{H}\Psi \\ &= iH^3\Psi - H\dot{H}\Psi - 2\dot{H}H\Psi - i\ddot{H}\Psi \\ &= iH^3\Psi - HH_t\Psi + iHH_\rho[H, \rho]\Psi - 2H_tH\Psi \\ &\quad + 2iH_\rho[H, \rho]H\Psi - iH_{tt}\Psi + 2(H_t)_\rho[H, \rho]\Psi + iH_\rho([H, \rho])^2\Psi \\ &\quad - H_\rho[H_t, \rho]\Psi + iH_\rho[H_\rho[H, \rho], \rho]\Psi + iH_\rho[H, [H, \rho]]\Psi. \end{aligned}$$

Taking the norm and applying (5.1), we obtain the desired result.  $\square$

Lemma 2, Lemma 3 and Lemma 4 give the local truncation error errors of both PT and Schrödinger dynamics. Following the standard stability analysis [29], we obtain the global error bounds.

**Theorem 5 (Global error).** For the IM schemes of (2.1) and (3.9) up to the time  $t_n = T$ , there exists some constant  $C$  depending on  $T$  and  $\|H\|$  such that

1. The errors for the PT dynamics (3.9) satisfy

$$\|\Phi(t_n) - \Phi_n\| + \|\sigma(t_n) - \sigma_n\| \leq Cf_1(H, \rho, P)h^2, \quad (5.12)$$

where  $f_1$  is a function of  $H, \rho, P$  that is a linear combination of products of nested commutators up to three layers of the form

$$\|[A_4, A_3[A_2, A_1A_0]]\| \quad (5.13)$$

with  $A_0$  being one of the following

$$[H, P], \quad [H_t, P], \quad [H_{tt}, P], \quad [H, \rho], \quad [H_t, \rho], \quad [H_{tt}, \rho] \quad (5.14)$$

and  $A_i$  ( $i = 1, \dots, 4$ ) being the identity matrix  $I$ , functions of  $H, \rho, P$  or derivatives of  $H$ .

2. The error for the Schrödinger dynamics (2.1) satisfies

$$\|\Psi(t_n) - \Psi_n\| \leq C \left( f_2(H, \rho, P) + \|H^3 \Psi\| + \|HH_t \Psi\| + 2 \|H_t H \Psi\| + \|H_{tt} \Psi\| \right) h^2, \quad (5.15)$$

where  $f_2$  has the same form as  $f_1$ .

Theorem 5 shows that the error bound of PT dynamics exhibits commutator scaling while that of the Schrödinger equation does not. Intuitively, the commutator terms oscillate on the time scale of the density matrices and projectors, which are gauge-invariant. These terms can be much smaller than those on the right-hand-side of the Schrödinger estimate acting on the fast-varying wavefunctions. For example,  $[H, \rho] = i\rho$  oscillates on the scale of the density matrix, and  $H\Psi = i\Psi$  oscillates on the scale of the wavefunction. This observation is also illustrated numerically in Section 6. Moreover, in the next section, we show rigorously that in the near adiabatic regime, the PT dynamics gains one extra order of accuracy in a singularly perturbation parameter  $\epsilon$  compared to that of the Schrödinger dynamics.

We remark that the worst-case dependence of the constant  $C$  on the norm of  $H$  can be very pessimistic, which is due to the standard stability analysis through the Grönwall type estimates. However, the Schrödinger equation inherits a Hamiltonian structure and, together with the fact that IM is a symplectic scheme, this preconstant  $C$  may be dramatically improved such that it depends linearly on  $T$  and is even possibly independent of  $\|H\|$  [14]. In order to formally employ the symplectic properties, however, the PT-IM scheme needs to be slightly modified. This has been demonstrated in [1] for pure states. Numerical results in [1] also demonstrate that the performance of the schemes with and without the modification are almost the same, so the modification may only be of theoretical interest. For simplicity, we do not detail such modification here.

### 5.3. Near adiabatic regime

In the near adiabatic regime, we can use commutator structure to demonstrate provable advantage of the PT dynamics over the Schrödinger dynamics. Consider the singular perturbed Schrödinger equation:

$$i\epsilon \partial_t \Psi^\epsilon(t) = H(t, \rho^\epsilon(t)) \Psi^\epsilon(t), \quad \epsilon \ll 1. \quad (5.16)$$

Here  $\rho^\epsilon(0)$  is a pure state, and  $\Psi^\epsilon(0)$  consists of the eigenfunctions of  $H(0, \rho^\epsilon(0))$  corresponding to the algebraically lowest  $N$  eigenvalues.

Let  $\rho^\epsilon = P^\epsilon = \Psi^\epsilon \Psi^\epsilon \dagger$ . Then the PT dynamics become

$$i\epsilon \partial_t \Phi^\epsilon(t) = H(t, \rho^\epsilon(t)) \Phi^\epsilon(t) - \Phi^\epsilon(t) (\Phi^\epsilon \dagger(t) H(t, \rho^\epsilon(t)) \Phi^\epsilon(t)), \quad \Phi^\epsilon(0) = \Psi^\epsilon(0). \quad (5.17)$$

In the linear case ( $H(t, \rho(t)) = H(t)$  is independent of  $\rho$ ), if the gap condition is satisfied, i.e. there exists a positive gap between the  $N$ -th and  $(N+1)$ -th smallest eigenvalues of  $H(t)$  for all  $t \in [0, T]$ . The adiabatic theorem (see for example, [42,13,21,22]) for the Schrödinger dynamics (5.16) states that

$$\Psi^\epsilon(t) = \Psi_a(t) + \mathcal{O}(\epsilon), \quad (5.18)$$

where the columns of  $\Psi_a(t)$  are the eigenvectors of the Hamiltonian, namely, there exists a time-dependent diagonal matrix  $\Lambda(t)$  whose diagonal entries are eigenvalues of the Hamiltonian such that

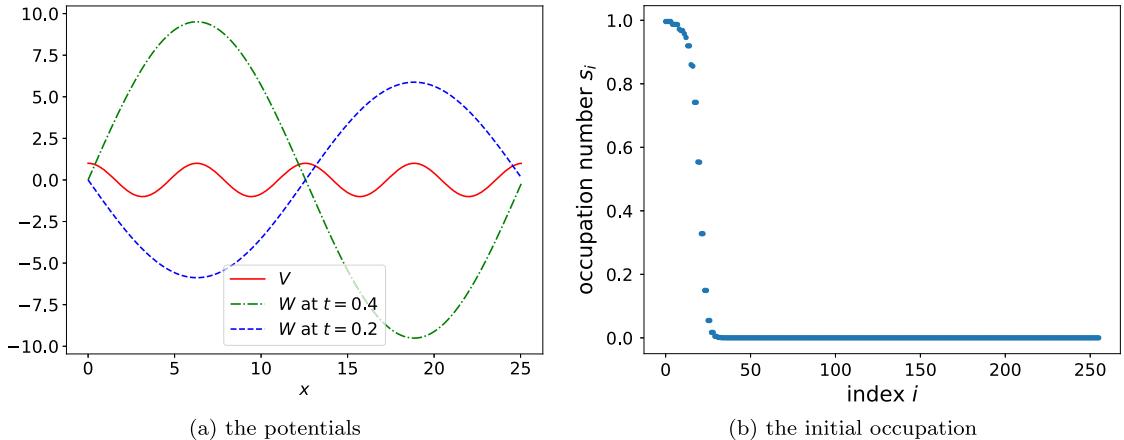
$$H(t) \Psi_a(t) = \Psi_a(t) \Lambda(t).$$

The adiabatic theorem can also be generalized to certain linear systems without a gap condition [3,41], and for some weakly nonlinear systems [9,11]. A detailed discussion of the technical conditions for the adiabatic approximation is beyond the scope of this paper. However, when such *a priori* estimate is available, we can evaluate the commutator as

$$[H, \rho^\epsilon] = H \Psi_a \Psi_a^\dagger - \Psi_a \Psi_a^\dagger H + \mathcal{O}(\epsilon) = \Psi_a \Lambda \Psi_a^\dagger - \Psi_a \Lambda \Psi_a^\dagger + \mathcal{O}(\epsilon) = \mathcal{O}(\epsilon). \quad (5.19)$$

We now examine the commutator terms in Lemma 3. Note that in the singular perturbed regime, one should replace the  $H$  in Lemma 3 by  $H/\epsilon$ , and hence the leading order terms in  $\epsilon$  are

$$\begin{aligned} & \epsilon^{-3} \|[H_{\rho^\epsilon} \rho^\epsilon ([H, \rho^\epsilon])^2, \rho^\epsilon]\| + \epsilon^{-3} \|[H_{\rho^\epsilon} [H_{\rho^\epsilon} [H, \rho^\epsilon], \rho^\epsilon], \rho^\epsilon]\| + \epsilon^{-3} \|[H_{\rho^\epsilon} [H, [H, \rho^\epsilon]], \rho^\epsilon]\| \\ & + \epsilon^{-3} \|[H_{\rho^\epsilon} [H, \rho^\epsilon], [H, \rho^\epsilon]]\| + \epsilon^{-3} \|[H, [H_{\rho^\epsilon} [H, \rho^\epsilon], \rho^\epsilon]]\| + \epsilon^{-3} \|[H, [H, [H, \rho^\epsilon]]]\| = \mathcal{O}(\epsilon^{-2}), \end{aligned}$$



**Fig. 1.** Left panel: The potentials  $V(x)$  (red solid) and  $W(t, x)$  at time  $t = 0.2$  (blue dashed) and  $t = 0.4$  (green dotted), respectively, where  $W$  is of time period  $1/8$ . Right panel: The initial occupation of the Fermi-Dirac statistics.  $L = 4$ ,  $\beta = 1.453$ , and the chemical potential is chosen such that the initial number of occupation  $N_e = \text{Tr}(\rho(0)) = 20$ .

thanks to (5.19). However, by replacing  $H$  in Lemma 4 by  $H/\epsilon$ , we obtain  $\|\partial_t^3 \Psi\| = \mathcal{O}(\epsilon^{-3})$  for the Schrödinger dynamics. Finally applying Theorem 5, we find that the numerical schemes for the PT dynamics can gain an order of magnitude in terms of the accuracy in  $\epsilon$ , which recovers the result in [1] for the linear case, and generalizes the result to the nonlinear case (provided that adiabatic theorems can be established).

## 6. Numerical results

In this section, we provide the numerical results of the parallel transport dynamics. Extensive numerical results and implementations on real chemical systems have been demonstrated for the PT dynamics with pure initial states [1,18–20]. Hence, we focus on the case of a mixed initial state in this section. In numerical examples, the relative numerical errors are computed by

$$\sup_{0 \leq k \leq n} \frac{\|X_k - X(t_k)\|}{\|X(t_k)\|},$$

where  $k = 0, 1, \dots, n$  and  $t_n$  is the final time, and  $X$  is the reference values of  $\Phi$ ,  $\sigma$  or  $\Psi$  with  $X_n$  represents the numerical results of the quantity  $X$  at the time  $t_n$ .

Our model system is defined by a periodic potential field given by a one-dimensional lattice structure with Hamiltonian

$$H(t) = -\frac{1}{2}\Delta + V(x) + W(x, t). \quad (6.1)$$

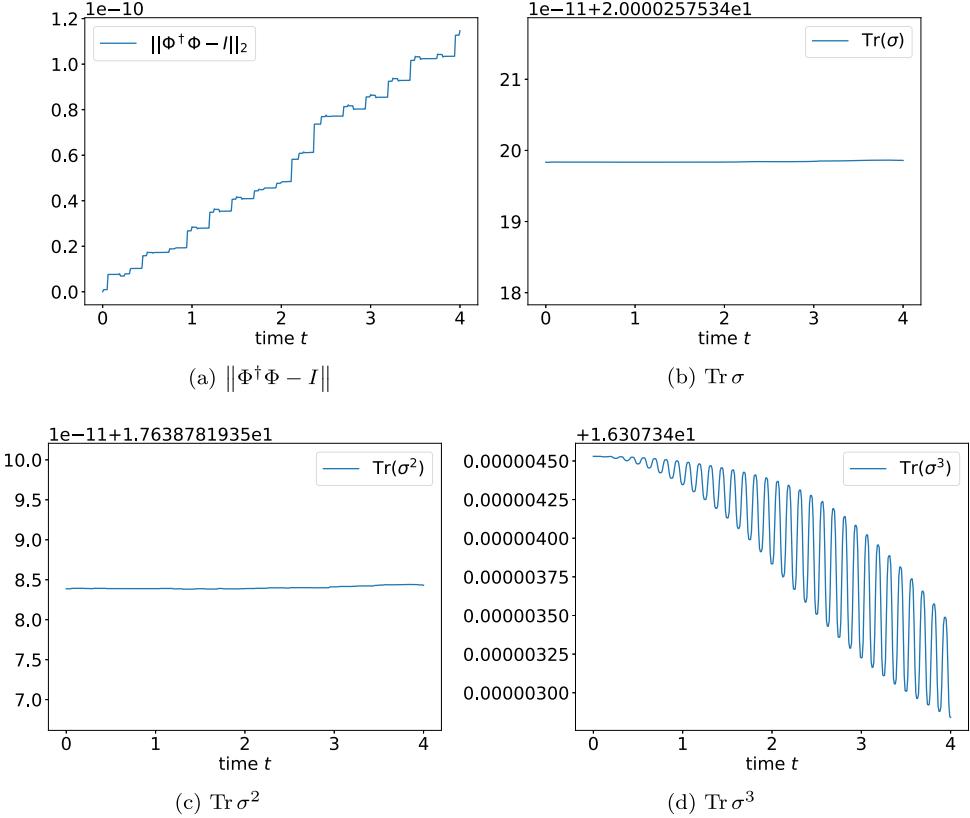
Here  $V(x) = \cos(x)$  is a static potential. The external time-dependent potential with frequency  $\omega$  is

$$W(x, t) = 10 \sin\left(\frac{x}{L}\right) \sin(\omega t), \quad (6.2)$$

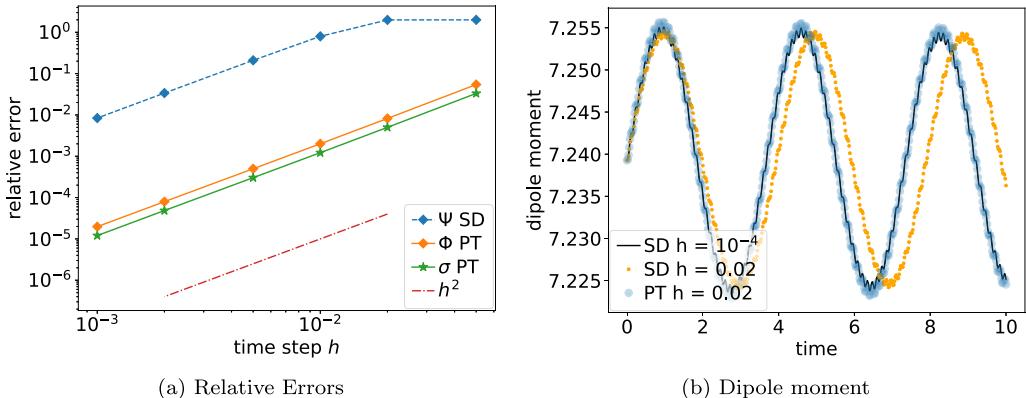
and  $L$  denotes the number of unit cells. The length of the lattice (the computational domain) is  $2\pi L$ . Fig. 1a shows a typical plot for the two potentials over the lattice cells. The parameters in the system are chosen as  $L = 4$ ,  $\beta = 1.453$ ,  $\omega = 16\pi$  and the chemical potential  $\mu = 3.299$ . The initial occupation number according to the Fermi-Dirac distribution is in Fig. 1b. Each unit cell is discretized via 64 equidistant grid points, and hence the total number of grid points is  $N_g = 64L$ .

We first verify the Proposition 1 numerically by simulating the PT dynamics to  $T_{\text{final}} = 4$  using the PT-IM scheme with a step size  $h = 0.01$ . We set  $N_e = 20$ , and  $N = 64$ . The norm of  $\Phi_{n+1}^\dagger \Phi_{n+1}$  and values of  $\text{Tr} \sigma_n$  and  $\text{Tr} \sigma_n^2$  are plotted for the simulation time in Fig. 2. It can be seen that the values of all three quantities are constant throughout the simulation, which agrees with Proposition 1. In comparison, we also plot the higher order trace  $\text{Tr} \sigma_n^3$ , which is not a conserved quantity. Nonetheless, the fluctuation of  $\text{Tr} \sigma_n^3$  is still very small and on the order of  $10^{-6}$ .

Next, we compare the numerical errors in simulating the Schrödinger dynamics (SD) and PT dynamics. Both dynamics are simulated using IM schemes to  $T_{\text{final}} = 1$ . We set  $\mu = 26.893$  (corresponding to  $N_e = 60$ ) and  $N = 80$ . In order to verify the convergence rate numerically, we set the time steps to be 0.05, 0.02, 0.01, 0.005, 0.002, 0.001. The reference solution is computed using a fine time step of  $2 \times 10^{-5}$ . Fig. 3a shows that both SD-IM and PT-IM are second order methods, but the preconstant of PT-IM is much smaller. The accuracy of the PT dynamics can also be shown in terms of physical observables, e.g. the dipole moment:



**Fig. 2.** Numerical verification (time step  $h = 0.01$ ) of the orthogonality of  $\Phi$  and the trace preservation of  $\sigma$  and  $\sigma^2$ , as shown in Proposition 1. On the other hand, the trace of higher powers of  $\sigma$  (e.g.  $\sigma^3$ ) may not be preserved in the PT-IM scheme.

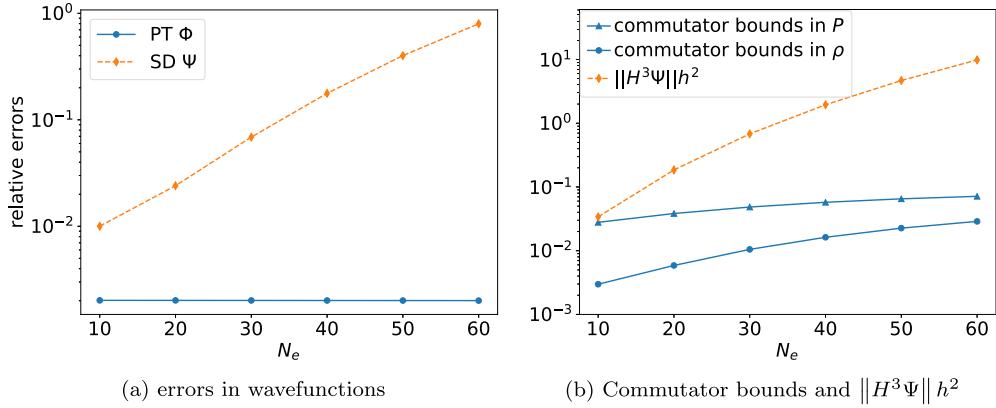


**Fig. 3.** Left Panel: Log-log plot of the relative errors of  $\Phi$ ,  $\sigma$ ,  $\Psi$  and the density matrix  $\rho$  computed via both PT and Schrödinger dynamics (SD). Right Panel: Evolution of the dipole moment for PT-IM with  $h = 0.02$ , SD-IM with  $h = 0.02$ , and SD-IM with  $h = 0.0001$  (reference solution).

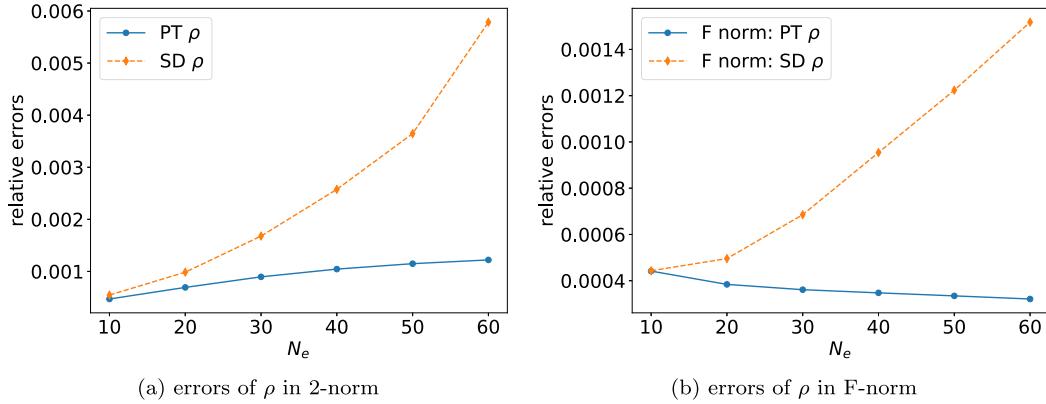
$$\langle x(t) \rangle := \text{Tr}(x\rho(t)).$$

Fig. 3b compares the dipole moment computed in three different scenarios: PT-IM with  $h = 0.02$ , SD-IM with  $h = 0.02$ , and SD-IM using a very small time step  $h = 0.0001$ . We find that the difference between the time-dependent dipole moment obtained from PT-IM with a large time step  $h = 0.02$  is almost the same as that from the reference solution. However, SD-IM with the same time step size is only accurate for a short periodic of time, and its accuracy significantly deteriorates as  $t$  increases.

In order to demonstrate that the commutator scaling in Theorem 5, we now vary the number of electrons  $N_e$ , and compare the results of PT-IM and SD-IM. The chemical potential  $\mu$  is set to 3.299, 7.028, 12.291, 18.951, 26.893, and the



**Fig. 4.** Relative errors versus the number of occupation  $N_e$  in semi-log scale. Left panel: relative errors for the wavefunctions  $\Phi$  in PT gauge and  $\Psi$  in the Schrödinger gauge. Right panel: the commutator bounds on the right-hand-sides of  $\partial_t^3 P$  in Lemma 2 and  $\partial_t^3 \rho$  in Lemma 3 versus  $\|H^3 \Psi\| h^2$  that appears in Lemma 4. The commutator bounds (as in PT) are significant smaller than  $\|H^3 \Psi\| h^2$  term (as in the Schrödinger gauge).



**Fig. 5.** Plots of the relative errors of  $\rho$  versus the number of occupation  $N_e$ . Left panel: relative errors in 2-norm. Right panel: the relative errors in the Frobenius norm (F-norm).

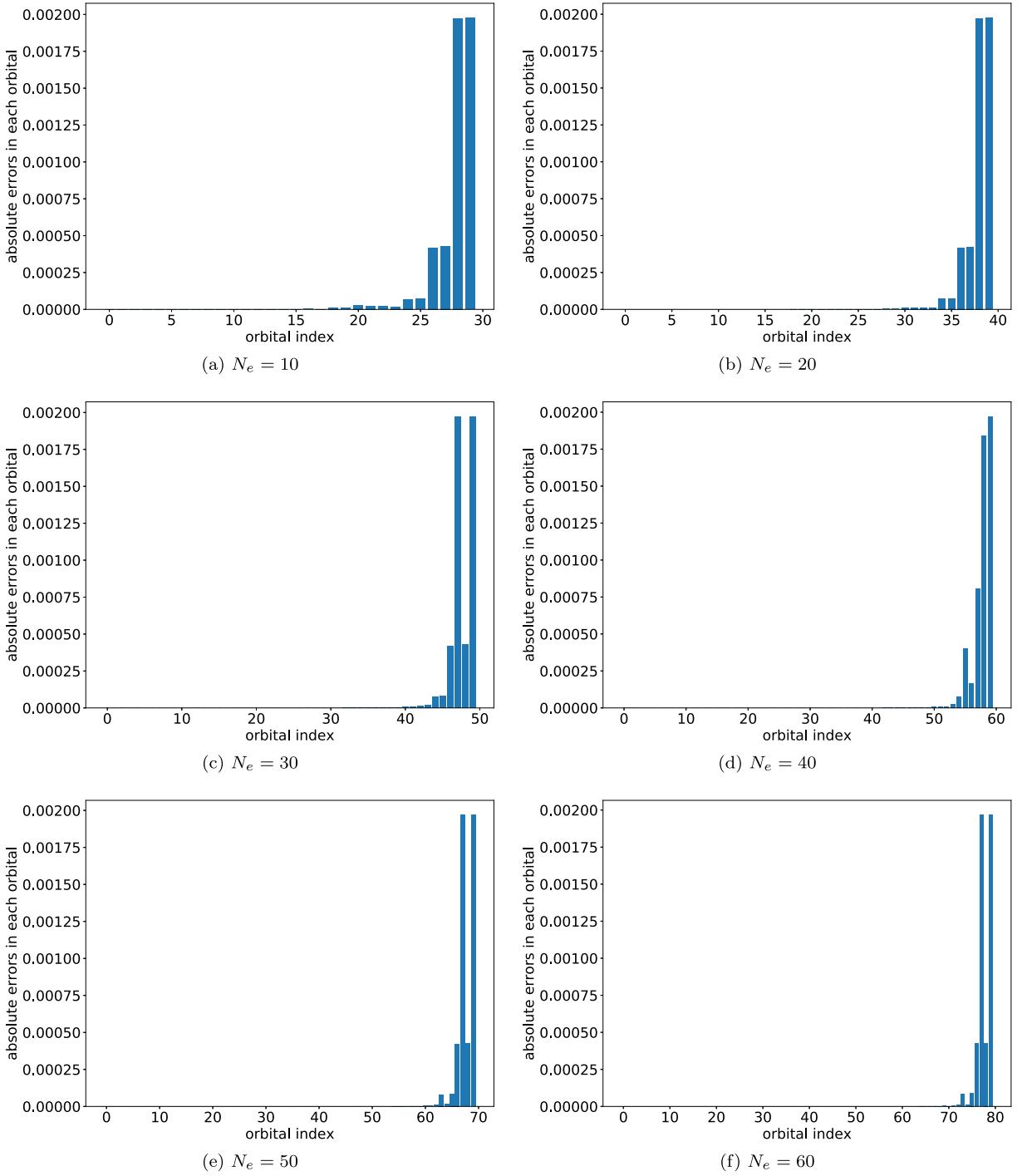
corresponding  $N_e$  are 10, 20, 30, 40, 50, 60, respectively. We also set  $N = N_e + 20$ ,  $h = 0.01$ , and  $T_{\text{final}} = 1$ . The reference solution is computed using a very small step size  $h = 2 \times 10^{-5}$ .

We plot the relative errors of both the PT and the Schrödinger dynamics in comparison with our theoretical bounds. It can be seen in Fig. 4 that as  $N_e$  increases, the relative error of the wavefunction in the Schrödinger dynamics grows much faster than that in the PT dynamics. Fig. 4 also plots the terms in the error bounds with or without the commutator structures, respectively. We find that the term without commutator structures can be much larger in magnitude, and the qualitative trend of the growth of the error bound with respect to  $N_e$  matches that of the error from the numerical simulation.

We also plot the relative errors in 2-norm of the density matrix  $\rho$  in Fig. 5a and Fig. 5b. The errors (measured in both the operator norm and the Frobenius norm) from the PT dynamics is smaller than that from the Schrödinger dynamics. Furthermore, as  $N_e$  increases, the relative error in the Frobenius norm from the PT dynamics in fact decreases. This phenomenon can be intuitively explained as follows. Note that the initial  $\sigma_0$  is a diagonal matrix of the following form

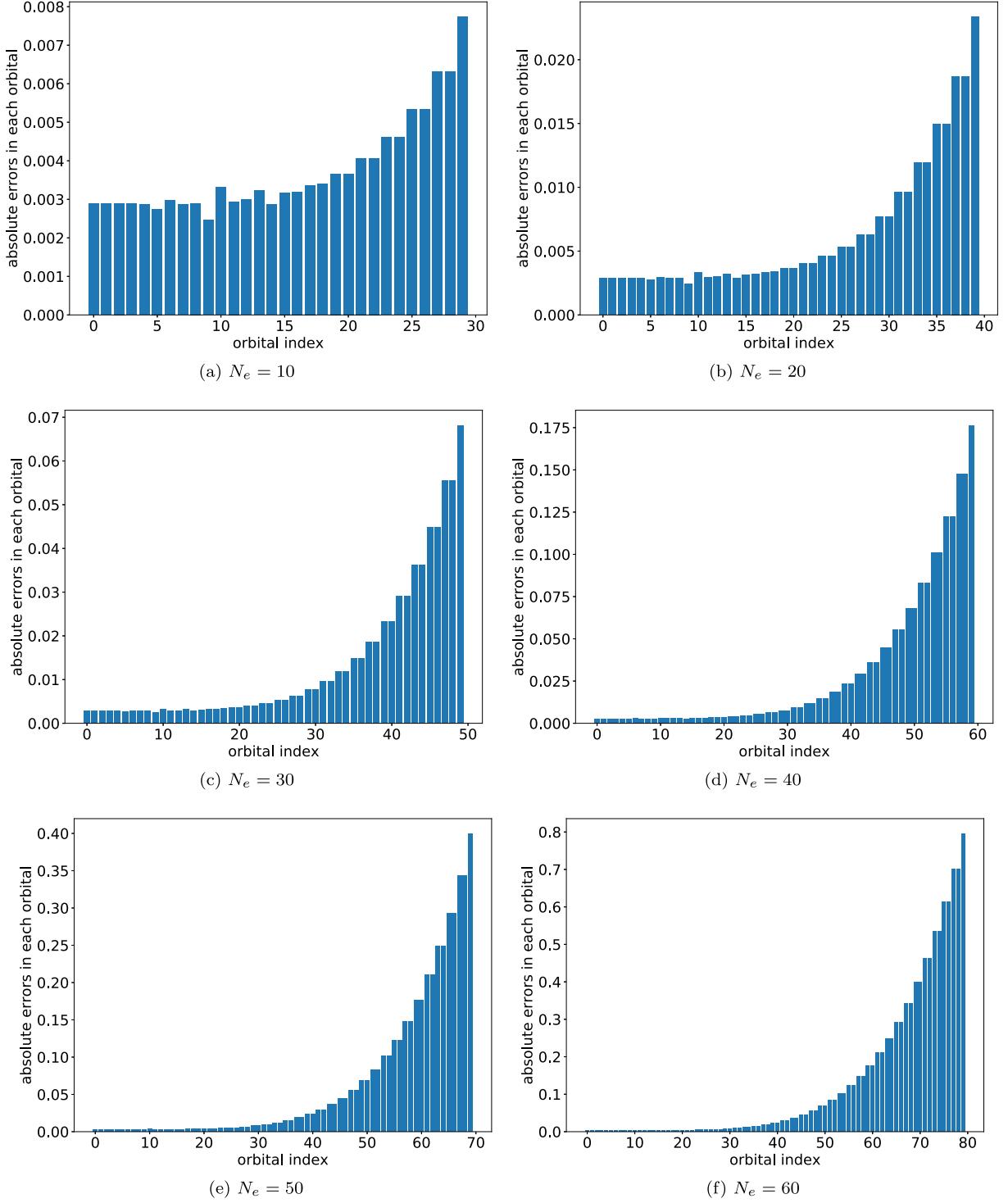
$$\begin{pmatrix} I_{m_0} & 0 & 0 \\ 0 & \sigma_* & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

where  $m_0$  is the number of fully occupied states and  $\sigma_*$  is a diagonal matrix representing the fractional states whose diagonal elements has values in  $(0, 1)$ . Then we expect that the fully occupied states are approximately in the near adiabatic regime, and their contribution to the error is much smaller than those from the fractionally occupied ones according to the commutator bound. In this example,  $m_0$  increases with  $N_e$ , but the size of  $\sigma_*$  does not change much with respect to  $N_e$ . Therefore, the error of the density matrix should be dominated by a small number of orbitals near the Fermi surface. To verify this statement, we plot in Fig. 6 the histogram of the errors in the vector 2-norm for all orbitals. Indeed, as  $N_e$  increases, the errors are dominant by only a few orbitals near the corresponding chemical potential  $\mu$ , and the number



**Fig. 6.** Plot of the error histogram of all orbitals for various  $N_e$  in the PT dynamics.

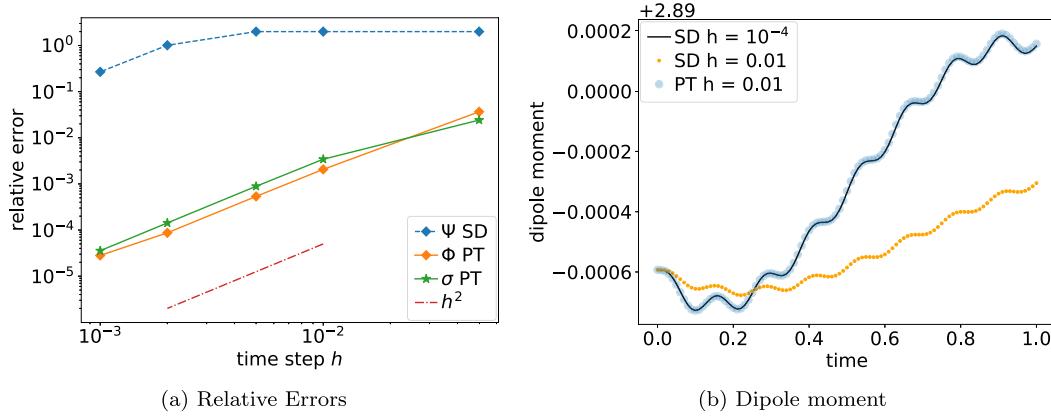
of the orbitals with significant errors does not increase with  $N_e$ . On the other hand, the Frobenius norm of the density matrix  $\|\rho\|_F = \mathcal{O}(\sqrt{N_e})$ . This explains the decay of the relative error of  $\rho$  in the PT-dynamics in Fig. 5b. By comparison, the histogram of the errors in the vector 2-norm for all orbitals in the Schrödinger gauge is provided in Fig. 7. We find that in the Schrödinger dynamics, the errors are propagated much more widely along the energy spectrum among a larger number of orbitals. It is also interesting to note that the maximal magnitude of the error increases significantly with respect to  $N_e$  in the Schrödinger dynamics, but the maximal error is nearly a constant and is much smaller in the PT dynamics.



**Fig. 7.** Plot of the error histogram of all orbitals for various  $N_e$  in the Schrödinger dynamics.

Finally, we demonstrate that the PT dynamics remains equally effective in the nonlinear regime. The rt-TDDFT Hamiltonian takes the following general form

$$H(t, \rho(t)) = -\frac{1}{2}\Delta + V_{\text{ext}}(x, t) + V_{\text{Hxc}}[\rho(t)] + V_X[\rho(t)], \quad (6.3)$$



**Fig. 8.** Left Panel: The log-log plot of the relative errors of  $\Phi$ ,  $\sigma$  in the model nonlinear rt-TDDFT calculation with Eq. (6.4), computed via PT-IM and SD-IM. Right Panel: a comparison of the dipole moment.

where  $V_{\text{ext}}$  represents the electron-ion interaction and when the external field changes with respect to time,  $V_{\text{ext}}$  may also depend on time  $t$ .  $V_{\text{Hxc}}$  is the Hartree and local exchange-correlation contribution and depends only on the diagonal part of  $\rho(t)$ , and  $V_X$  is the Fock exchange operator depending on the entire  $\rho(t)$ . More specifically,  $V_X$  is an integral operator defined by

$$[(V_X[\rho])\phi](x) = - \int K(x, y)\rho(x, y)\phi(y) dy$$

with the kernel  $K(x, y)$  represents the electron-electron interaction.

Following Eq. (6.3), we consider the following model problem

$$H(t, \rho) = -\frac{1}{2}\Delta + V(x) + W(x, t) + U[\rho], \quad (6.4)$$

where  $V = x^2$ , and  $W$  is as defined in (6.2) and the nonlinear term  $U[\rho]$  models  $V_X[\rho]$  with the Yukawa kernel

$$K(x, y) = \frac{2\pi}{\kappa\epsilon_0} e^{-\kappa|x-y|}.$$

Note that as  $\kappa \rightarrow 0$ , the Yukawa kernel approaches to the Coulomb interaction that diverges in one dimension and hence is typically used in place of the bare Coulomb interaction for one-dimensional problems. The parameters are chosen as  $\epsilon_0 = 100$  and  $\kappa = 0.01$  so that the range of the electrostatic interaction is sufficiently long. Here  $\mu = 148.99$  so that  $N_e = 60$  and we choose  $N = 80$ . We simulate the system using PT-IM and SD-IM up to  $T_{\text{final}} = 0.5$  and compare the relative errors. As shown in Fig. 8a, the errors from PT-IM are significantly smaller. A comparison of the dipole moment is presented in Fig. 8b. We also compute the dipole moment using  $h = 0.01$  and compare the results with the reference solution obtained using SD-IM with a very fine time step  $h = 0.0001$ . It can be seen that the result using the PT dynamics agrees well with the reference, which is not the case for the Schrödinger dynamics with the same step size.

## 7. Conclusion

In this paper, we have introduced the PT dynamics for mixed quantum states, which generalizes the PT dynamics for pure states presented in [1]. Both the PT and Schrödinger dynamics employ the low-rank structure of the density matrix, and can produce the same density matrix and all derived physical observables (such as dipole moments) as those from the von Neumann equation in the continuous time limit. The PT dynamics differ from the Schrödinger dynamics in terms of the choice of the gauge. In particular, the PT gauge yields the slowest possible dynamics for the wavefunctions. This allows us to significantly increase the time step size in the numerical simulation while maintaining accuracy. This is particularly useful when  $N_g \gg N$  and we only have access to the matrix-vector multiplication operations in the form of  $H\Phi$ . This is the case when, e.g. planewaves and finite difference methods are used for the spatial discretization.

For the temporal discretization, the PT dynamics can be combined with any Runge-Kutta and linear multistep integrators. As a concrete example, we propose the parallel transport-implicit midpoint (PT-IM) scheme, which is an implicit method suitable for treating Hamiltonians with a large spectral radius. It also preserves certain trace conditions and the orthogonality of the wavefunctions. We establish a new error bound for the PT dynamics, where all terms in the error bounds involve either the commutator of the Hamiltonian (and its derivatives) and the density matrix (or the associated spectral projector). As a comparison, the error analysis of the Schrödinger dynamics is also provided, which does not exhibit such commutator scaling. This new error bound, together with various numerical experiments, justifies the advantage of the PT dynamics for

the general mixed states, where the dynamics can be nonlinear, and beyond the near adiabatic regime. Implementation of PT-IM for rt-TDDFT calculations of large scale real metallic systems is under progress.

### CRediT authorship contribution statement

A.D., D.F. and L.L. designed research, performed research, and wrote the paper.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgement

This work was partially supported by the NSF Quantum Leap Challenge Institute (QLCI) program through grant number OMA-2016245 (D.F.), by the Department of Energy under Grant No. DE-SC0017867 and No. FWP-NQISCCAWL (D.A., L.L.), and by the National Science Foundation under No. DMS-1652330, OMA-2016245 (L.L.). D.A. acknowledges the support by the Department of Defense through the Hartree Postdoctoral Fellowship at QuICS. L.L. is a Simons Investigator.

### Appendix A. Alternative derivation of the parallel transport dynamics using the tangent space formulation

In this section, we provide an alternative derivation of the PT dynamics using the tangent space formulation. The presentation of this derivation consists of three parts: we first write down Eq. (3.5) as well as the domain of the low-rank factors. We then derive the equation of motion of the low-rank factors of the rank- $N$  density matrix. Finally, we introduce the optimal gauge (i.e. PT gauge).

Let  $\rho$  be of rank  $N$  ( $N_e \leq N < N_g$ ). Recall Eq. (3.5):

$$\rho(t) = \Phi(t)\sigma(t)\Phi^\dagger(t), \quad (\text{A.1})$$

where  $\sigma(t)$  is an  $N \times N$  positive semidefinite Hermitian matrix, and  $\Phi$  is an  $N_g \times N$  complex-valued matrix that satisfies  $\Phi^\dagger\Phi = I_N$ , in other words,  $\Phi$  belongs to the Stiefel manifold  $\text{St}(N, N_g)$  defined as

$$\text{St}(N, N_g) = \{\Phi \in \mathbb{R}^{N_g \times N} : \Phi^\dagger\Phi = I_N\}.$$

As is explained in Section 3, this decomposition (A.1) in fact admits an equivalence relationship  $(\Phi, \sigma) \equiv (\Phi U, U^\dagger\sigma U)$ , namely, for any  $N \times N$  unitary matrix  $U$ ,

$$\rho = \Phi\sigma\Phi^\dagger = (\Phi U)(U^\dagger\sigma U)(U^\dagger\Phi^\dagger).$$

Consider the infinitesimal variation of the tangent map of  $(\Phi, \sigma) \mapsto \Phi\sigma\Phi^\dagger$ . As is shown in [5, Lemma 4], the tangent space of the Stiefel manifold admits the parametrization  $i\omega\Phi$ , where  $\omega$  is a Hermitian matrix of size  $N_g$ . Denote

$$\dot{\Phi} = i\omega\Phi, \quad \dot{\sigma} = \xi,$$

where  $\xi$  is a traceless Hermitian matrix of size  $N$ . The infinitesimal variation of  $\rho$  can thus be represented as

$$i[\omega, \rho] + \Phi\xi\Phi^\dagger = \Phi(i[\Phi^\dagger\omega\Phi, \sigma] + \xi)\Phi^\dagger.$$

We then project  $\dot{\rho}$  onto this tangent space by minimizing the distance between them, namely,

$$\min \left\| -i[H(t, \rho(t)), \rho(t)] - i[\omega, \rho] - \Phi\xi\Phi^\dagger \right\|_F.$$

Hereafter, we drop the  $(t, \rho)$  in  $H$  for simplicity. The two stationary conditions in  $\omega$  and  $\xi$  read

$$[-i[H, \rho] - i[\omega, \rho] - \Phi\xi\Phi^\dagger, \rho] = 0, \quad (\text{A.2})$$

$$\Phi^\dagger(-i[H, \rho] - i[\omega, \rho] - \Phi\xi\Phi^\dagger)\Phi = \lambda I_N, \quad (\text{A.3})$$

where  $\lambda$  is the Lagrange multiplier introduced to satisfy the traceless condition of  $\xi$ . By taking trace on both sides of (A.3), we find that

$$\lambda = \frac{1}{N} \text{Tr} \left[ \Phi^\dagger (-i[H, \rho] - i[\omega, \rho]) \Phi \right] = 0,$$

because for any  $X$ ,

$$\text{Tr}(\Phi^\dagger [X, \rho] \Phi) = \text{Tr}(\Phi^\dagger X \Phi \sigma \Phi^\dagger \Phi - \Phi^\dagger \Phi \sigma \Phi^\dagger X \Phi) = \text{Tr}(\Phi^\dagger X \Phi \sigma - \sigma \Phi^\dagger X \Phi) = 0.$$

Define projection operators

$$P := \Phi \Phi^\dagger, \quad Q = I - P,$$

and one can express  $\Phi \xi \Phi^\dagger$  using (A.3) as

$$\xi = \Phi^\dagger (-i[H, \rho] - i[\omega, \rho]) \Phi, \quad \Phi \xi \Phi^\dagger = P (-i[H, \rho] - i[\omega, \rho]) P. \quad (\text{A.4})$$

Together with Eq. (A.2), we obtain

$$Q (-iH\rho - i\omega\rho) \rho = \rho (i\rho H + i\rho\omega) Q.$$

Note that  $P\rho = \rho$  and hence the left-hand side stays in the range of  $Q$  while the right-hand side remains in the range of  $P$ . By orthogonality, both sides of the equation vanish, which imposes some constraints on  $Q\omega P$  and  $P\omega Q$ . To be specific, one has

$$Q\omega\rho^2 = -QH\rho^2 \iff Q\omega\Phi\sigma^2\Phi^\dagger = -QH\Phi\sigma^2\Phi^\dagger.$$

Right multiply  $\Phi(\sigma^{-1})^2\Phi^\dagger$ , one finds that  $Q\omega P = -QHP$ . Similarly, we obtain  $P\omega Q = -PHQ$ . The general solution of this system of matrix equations is

$$\omega = -QHP - PHQ - PGP - QGQ,$$

where  $G$  is any Hermitian matrix due to the hermiticity of  $\omega$ . Since  $\dot{\Phi} = i\omega\Phi$ , the equation for  $\Phi$  can be written as

$$\begin{aligned} i\dot{\Phi} &= -\omega\Phi = QHP\Phi + PHQ\Phi + PG\Phi + QGQ\Phi \\ &= QH\Phi + PG\Phi, \end{aligned}$$

where the fact that  $P\Phi = \Phi$  and  $Q\Phi = 0$  is used. For the equation of  $\xi$ , (A.4) yields

$$\begin{aligned} i\dot{\sigma} &= i\xi = i\Phi^\dagger (-i[H, \rho] - i[\omega, \rho]) \Phi \\ &= [\Phi^\dagger H\Phi, \sigma] - [\Phi^\dagger G\Phi, \sigma]. \end{aligned}$$

Finally, we arrive at the dynamics for  $\Phi$  and  $\sigma$  in a closed form

$$i\partial_t \Phi = (I - \Phi\Phi^\dagger)H(t, \Phi\sigma\Phi^\dagger)\Phi + \epsilon\Phi\Phi^\dagger G\Phi, \quad (\text{A.5})$$

$$i\partial_t \sigma = [\Phi^\dagger (H(t, \Phi\sigma\Phi^\dagger) - G)\Phi, \sigma], \quad (\text{A.6})$$

where the Hermitian matrix  $G$  is an extra degree of freedom to be chosen.

The next step is to find the optimal choice of  $G$  such that the dynamics of  $\Phi$  changes the slowest, i.e. to find  $G$  such that

$$\min \|\dot{\Phi}\|_F^2 = \min \text{Tr}(\dot{\Phi}^\dagger \dot{\Phi}). \quad (\text{A.7})$$

The norm can be split into two parts

$$\begin{aligned} \|\dot{\Phi}\|_F^2 &= \|P\dot{\Phi}\|_F^2 + \|Q\dot{\Phi}\|_F^2 \\ &= \|\Phi\Phi^\dagger G\Phi\|_F^2 + \|QH\Phi\|_F^2. \end{aligned}$$

The Frobenius norm of the second term reads

$$\text{Tr}(\Phi^\dagger H^\dagger Q^\dagger QH\Phi) = \text{Tr}(Q^\dagger QH\Phi\Phi^\dagger H^\dagger) = \text{Tr}(Q^\dagger QH(t, \rho)PH^\dagger(t, \rho)),$$

which is independent of the gauge choice. Therefore, to optimize (A.7) one can choose  $G = 0$ . Now we arrive at the parallel transport dynamics

$$i\partial_t \Phi = (I - \Phi\Phi^\dagger)H(t, \Phi\sigma\Phi^\dagger)\Phi, \quad (\text{A.8})$$

$$i\partial_t \sigma = [\Phi^\dagger H(t, \Phi\sigma\Phi^\dagger)\Phi, \sigma],$$

which is equivalent to the PT dynamics (3.9) derived in Section 3.

It is worth pointing out that the Schrödinger gauge in fact corresponds to the choice  $G = H$  in (A.5) that gives rise to

$$\begin{aligned} i\partial_t \Phi &= H(t, \Phi \sigma \Phi^\dagger) \Phi, \\ \partial_t \sigma &= 0. \end{aligned} \tag{A.9}$$

This immediately implies that the number of occupied orbitals remains unchanged throughout the evolution, which verifies the validity of the solution in the Schrödinger dynamics in Eq. (3.3).

## Appendix B. Anderson's mixing method for solving nonlinear equations and computational cost

Let  $x_k$  be the value of  $(\Phi_{n+1}, \sigma_{n+1})$  at the  $k$ -th iteration, and define

$$r_k = x_k - T(x_k), \quad s_k = x_k - x_{k-1}, \quad y_k = r_k - r_{k-1},$$

then Anderson's mixing method obtains an update  $x_{k+1}$  according to

$$x_{k+1} = x_k - C_k r_k.$$

The matrix  $C_k$  can be viewed as an approximation to the Jacobian of the mapping  $x \mapsto x - T(x)$ . Define

$$S_k = (s_k, s_{k-1}, \dots, s_{k-\ell}), \quad Y_k = (y_k, y_{k-1}, \dots, y_{k-\ell}),$$

where  $\ell$  is called the mixing dimension. The matrix  $C_k$  is obtained by solving the minimization problem

$$\begin{aligned} \min_C \quad & \frac{1}{2} \|C - C_0\|_F^2 \\ \text{s.t.} \quad & S_k = CY_k, \end{aligned} \tag{B.1}$$

where  $C_0$  is an initial guess to the Jacobian. In the absence of further information, we may choose  $C_0 = \alpha I$  for some constant  $\alpha$ . The solution to the minimization problem

$$C_k = C_0 + (S_k - C_{k-1} Y_k) Y_k^+,$$

where  $Y_k^+ = (Y_k^\dagger Y_k)^{-1} Y_k^\dagger$  is the Moore-Penrose pseudoinverse. Therefore, the update rule for  $x_{k+1}$  is

$$x_{k+1} = x_k - C_0 (I - Y_k Y_k^+) r_k - S_k Y_k^+ r_k. \tag{B.2}$$

For the initial step, we may simply take  $x_0$  to be the vectorized form of  $(\Phi_n, \sigma_n)$  at the previous time step.

In the numerical solution of the fixed point problem, other than operations of the form  $H\Phi$ , the cost of the remaining linear algebra operations from  $x_k$  to  $x_{k+1}$  is  $\mathcal{O}(N_g N^2 + N^3) = \mathcal{O}(N_g N^2)$  under the assumption that  $N_g \gg N$ . The cost of the  $H\Phi$  operation is more involved and depends on the choice of the exchange-correlation functionals and basis sets. In particular, when hybrid exchange-correlation functionals are used, the cost of  $H\Phi$  is generally  $\mathcal{O}(N_g N^2)$  with a large prefactor (also possibly with a  $\log N_g$  factor when planewave basis functions are used, due to the fast Fourier transform). Therefore the computational cost per iteration step is  $\mathcal{O}(N_g N^2)$ , and is mostly dominated by the  $H\Phi$  operation.

## References

- [1] D. An, L. Lin, Quantum dynamics with the parallel transport gauge, *Multiscale Model. Simul.* 18 (2020) 612.
- [2] D.G. Anderson, Iterative procedures for nonlinear integral equations, *J. Assoc. Comput. Mach.* 12 (1965) 547–560.
- [3] J.E. Avron, A. Elgart, Adiabatic theorem without a gap condition, *Commun. Math. Phys.* 203 (2) (1999) 445–463.
- [4] A. Bruner, S. Hernandez, F. Mauger, P.M. Abanador, D.J. LaMaster, M.B. Gaarde, K.J. Schafer, L. Polata, Attosecond charge migration with TDDFT: accurate dynamics from a well-defined initial state, *J. Phys. Chem. Lett.* 8 (17) (2017) 3991–3996.
- [5] Y. Cao, J. Lu, Stochastic dynamical low-rank approximation method, *J. Comput. Phys.* 372 (2018) 564–586.
- [6] A. Castro, M. Marques, A. Rubio, Propagators for the time-dependent Kohn-Sham equations, *J. Chem. Phys.* 121 (2004) 3425–3433.
- [7] P.H. Dederichs, R. Zeller, Self-consistency iterations in electronic-structure calculations, *Phys. Rev. B* 28 (1983) 5462–5472.
- [8] P. Elliott, N.T. Maitra, Propagation of initially excited states in time-dependent density-functional theory, *Phys. Rev. A* 85 (5) (2012) 052510.
- [9] C. Fermanian-Kammerer, A. Joye, A nonlinear quantum adiabatic approximation, *Nonlinearity* 33 (9) (2020) 4715–4751.
- [10] S.A. Fischer, C.J. Cramer, N. Govind, Excited state absorption from real-time time-dependent density functional theory, *J. Chem. Theory Comput.* 11 (9) (2015) 4294–4303.
- [11] Z. Gang, P. Grech, Adiabatic theorem for the Gross-Pitaevskii equation, *Commun. Partial Differ. Equ.* 42 (5) (May 2017) 731–756.
- [12] A. Gómez Pueyo, M.A. Marques, A. Rubio, A. Castro, Propagators for the time-dependent Kohn-Sham equations: multistep, Runge-Kutta, exponential Runge-Kutta, and commutator free Magnus methods, *J. Chem. Theory Comput.* (2018).
- [13] G.A. Hagedorn, A. Joye, Elementary exponential error estimates for the adiabatic approximation, *J. Math. Anal. Appl.* 267 (1) (2002) 235–246.
- [14] E. Hairer, C. Lubich, G. Wanner, *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*, vol. 31, Springer, 2006.
- [15] E. Hairer, S.P. Nørsett, G. Wanner, *Solving Ordinary Differential Equation I: Nonstiff Problems*, vol. 8, Springer, 1987.
- [16] J. Heyd, G.E. Scuseria, M. Ernzerhof, Hybrid functionals based on a screened Coulomb potential, *J. Chem. Phys.* 118 (18) (2003) 8207–8215.

- [17] K.M. Ho, J. Ihm, J.D. Joannopoulos, Dielectric matrix scheme for fast convergence in self-consistent electronic-structure calculations, *Phys. Rev. B* 25 (1982) 4260–4262.
- [18] W. Jia, D. An, L.-W. Wang, L. Lin, Fast real-time time-dependent density functional theory calculations with the parallel transport gauge, *J. Chem. Theory Comput.* 14 (2018) 5645.
- [19] W. Jia, L. Lin, Fast real-time time-dependent hybrid functional calculations with the parallel transport gauge and the adaptively compressed exchange formulation, *Comput. Phys. Commun.* 240 (21) (2019).
- [20] W. Jia, L.-W. Wang, L. Lin, Parallel transport time-dependent density functional theory calculations with hybrid functional on summit, in: SC '19 Proceedings of the International Conference for High Performance Computing, number 79, 2019, pp. 1–23.
- [21] A. Joye, An adiabatic theorem for singularly perturbed Hamiltonians, *Ann. Inst. Henri Poincaré, Phys. Théor.* 63 (2) (1995) 231–250.
- [22] A. Joye, General adiabatic evolution with a gap condition, *Commun. Math. Phys.* 275 (1) (2007) 139–162.
- [23] C.T. Kelley, *Iterative Methods for Optimization*, vol. 18, SIAM, 1999.
- [24] G.P. Kerker, Efficient iteration scheme for self-consistent pseudopotential calculations, *Phys. Rev. B* 23 (1981) 3082–3084.
- [25] D.A. Knoll, D.E. Keyes, Jacobian-free Newton–Krylov methods: a survey of approaches and applications, *J. Comput. Phys.* 193 (2004) 357–397.
- [26] O. Koch, C. Lubich, Dynamical low-rank approximation, *SIAM J. Matrix Anal. Appl.* 29 (2007) 434–454.
- [27] C. Le Bris, P. Rouson, Low-rank numerical approximations for high-dimensional lindblad equations, *Phys. Rev. A* 87 (2) (2013) 1.
- [28] C. Le Bris, P. Rouson, J. Roussel, Adaptive low-rank approximation and denoised Monte Carlo approach for high-dimensional Lindblad equations, *Phys. Rev. A* 92 (6) (2015) 1.
- [29] R.J. LeVeque, *Finite Difference Methods for Ordinary and Partial Differential Equations: Steady-State and Time-Dependent Problems*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2007.
- [30] C. Lubich, *From Quantum to Classical Molecular Dynamics: Reduced Models and Numerical Analysis*, European Mathematical Society, 2008.
- [31] C. Lubich, I.V. Oseledets, A projector-splitting integrator for dynamical low-rank approximation, *BIT Numer. Math.* 54 (1) (2014) 171–188.
- [32] J. Ma, Z. Wang, L.-W. Wang, Interplay between plasmon and single-particle excitations in a metal nanocluster, *Nat. Commun.* 6 (1) (2015) 10107.
- [33] R. Martin, *Electronic Structure: Basic Theory and Practical Methods*, Cambridge Univ. Press, 2008.
- [34] M. Nakahara, *Geometry, Topology and Physics*, CRC Press, 2003.
- [35] G. Onida, L. Reining, A. Rubio, Electronic excitations: density-functional versus many-body Green's-function approaches, *Rev. Mod. Phys.* 74 (2002) 601.
- [36] J.P. Perdew, K. Burke, M. Ernzerhof, Generalized gradient approximation made simple, *Phys. Rev. Lett.* 77 (1996) 3865–3868.
- [37] P. Pulay, Convergence acceleration of iterative sequences: the case of SCF iteration, *Chem. Phys. Lett.* 73 (1980) 393–398.
- [38] D.A. Rehn, Y. Shen, M.E. Buchholz, M. Dubey, R. Namburu, E.J. Reed, ODE integration schemes for plane-wave real-time time-dependent density functional theory, *J. Chem. Phys.* 150 (1) (2019) 014101.
- [39] E. Runge, E.K.U. Gross, Density-functional theory for time-dependent systems, *Phys. Rev. Lett.* 52 (1984) 997.
- [40] A. Schleife, E.W. Draeger, Y. Kanai, A.A. Correa, Plane-wave pseudopotential implementation of explicit integrators for time-dependent Kohn-Sham equations in large-scale simulations, *J. Chem. Phys.* 137 (2012) 22.
- [41] S. Teufel, A note on the adiabatic theorem without gap condition, *Lett. Math. Phys.* 58 (3) (2002) 261–266, 2001.
- [42] S. Teufel, *Adiabatic Perturbation Theory in Quantum Dynamics*, Lecture Notes in Mathematics, vol. 1821, Springer-Verlag, Berlin, 2003.
- [43] C.A. Ullrich, *Time-Dependent Density-Functional Theory: Concepts and Applications*, Oxford Univ. Press, 2011.
- [44] Z. Wang, S.-S. Li, L.-W. Wang, Efficient real-time time-dependent density functional theory method and its application to a collision of an ion with a 2D material, *Phys. Rev. Lett.* 114 (2015) 1.
- [45] K. Yabana, G.F. Bertsch, Time-dependent local-density approximation in real time, *Phys. Rev. B* 54 (1996) 4484–4487.
- [46] S.-H. Yeh, A. Manjanath, Y.-C. Cheng, J.-D. Chai, C.-P. Hsu, Excitation energies from thermally assisted-occupation density functional theory: theory and computational implementation, *J. Chem. Phys.* 153 (8) (2020) 084120.
- [47] D.C. Yost, Y. Yao, Y. Kanai, Propagation of maximally localized Wannier functions in real-time TDDFT, *J. Chem. Phys.* 150 (19) (2019) 194113.