MULTISCALE MODEL. SIMUL.
Vol. 17, No. 4, pp. 1274–1300

PROJECTION-BASED EMBEDDING THEORY FOR SOLVING KOHN–SHAM DENSITY FUNCTIONAL THEORY*

LIN LIN† AND LEONARDO ZEPEDA-NUNEZ‡

Abstract. Quantum embedding theories are playing an increasingly important role in bridging different levels of approximation to the many-body Schrödinger equation in physics, chemistry, and materials science. In this paper, we present a linear algebra perspective of the recently developed projection-based embedding theory (PET) [Manby et al., *J. Chem. Theory Comput.*, 8 (2012), pp. 2564–2268], restricted to the context of Kohn–Sham density functional theory. By partitioning the global degrees of freedom into a "system" part and a "bath" part and by choosing a proper projector from the bath, PET is an in principle exact formulation to confine the calculation to the system part only and hence can be performed with reduced computational cost. Viewed from the perspective of domain decomposition methods, one particularly interesting feature of PET is that it does not enforce a boundary condition explicitly, and it remains applicable even when the discretized Hamiltonian matrix is dense, such as in the context of the planewave discretization. In practice, the accuracy of PET depends on the accuracy of the bath projector. Based on the linear algebra reformulation, we develop a first-order perturbation correction to the projector from the bath to improve its accuracy. Numerical results for real chemical systems indicate that with a proper choice of reference system used to compute the bath projector, the perturbatively corrected PET can be sufficiently accurate even when strong perturbation is applied to very small systems, such as the computation of the ground state energy of a SiH_3F molecule, using a SiH_4 molecule as the reference system.

Key words. density functional theory, projection embedding theory, numerical linear algebra

AMS subject classifications. 65Z05, 65F15

DOI. 10.1137/18M1202670

1. Introduction.

Multiphysics simulation usually involves two or more physical scales. In the context of electronic structure theory, even though everything on the scale of electrons and molecules is described by the many-body Schrödinger equation, the concept behind multiphysics simulation remains valid. The direct solution to the many-body Schrödinger equation itself is prohibitively expensive, except for systems with a handful of electrons. This has led to the development of various theoretical tools in both quantum physics and quantum chemistry to find approximate solutions to the Schrödinger equation. These theories can be effectively treated as "different levels of physics" providing different levels of accuracy. However, depending on the accuracy required, the computational cost associated with such approximate theories can still be very high. So if a large quantum system can be partitioned into a "system" part containing the degrees of freedom that are of interest that need to be treated using a relatively accurate theory and a "bath" part containing the rest of the degrees

*Received by the editors July 24, 2018; accepted for publication (in revised form) August 27, 2019; published electronically December 12, 2019.
https://doi.org/10.1137/18M1202670

Funding: This work was partially supported by the Department of Energy under grants DE-SC0017867 amd DE-AC02-05CH11231, the SciDAC program, and the Air Force Office of Scientific Research under award FA9550-18-1-0095. The first author was supported by the National Science Foundation under grant DMS-1652330.

†Department of Mathematics, University of California, Berkeley, Berkeley, CA 94720, and Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720 (linlin@math.berkeley.edu).

‡Computational Research Division, Berkeley, CA 94720. Current address: Department of Mathematics, University of Wisconsin-Madision, Madison, WI 53706. (lzepeda@lbl.gov; zepedanunez@wisc.edu).

of freedom that can be treated using a less accurate theory, it becomes naturally desirable to have a numerical method that can bridge the two levels of theories. In quantum physics, such "multiscale" methods have been actively developed in the past few decades and are often called "quantum embedding theories" (see, e.g., [3, 37, 8, 20, 14, 16, 12, 5, 35, 19, 38, 18, 30, 6, 25] and [33] for a recent brief review).

The projection-based embedding theory (PET) [28] is a recently developed quantum embedding theory, which is a versatile method that can be used to couple a number of quantum theories together in a seamless fashion (also see recent works [26, 7]). This paper is a first step toward a mathematical understanding of PET. To make the discussions concrete, we assume that the system part is described by the widely used Kohn–Sham density functional theory (KSDFT) [15, 23] and that the bath part is also described by KSDFT but solved only approximately. Although this setup is simpler than the one presented in [28], it is already interesting from the perspective of approximate solution of large-scale eigenvalue problems, as will be detailed below.

After proper discretization, KSDFT can be written as the following nonlinear eigenvalue problem:

$$H[P]\Psi = \Psi\Lambda, \quad P = \Psi\Psi^*, \tag{1.1}$$

where the Hamiltonian $H[P] \in \mathbb{C}^{N\times N}$ is a Hermitian matrix and the diagonal matrix $\Lambda \in \mathbb{R}^{N_e\times N_e}$ encodes the algebraically lowest N_e eigenvalues ($N \gg N_e$). N is the number of degrees of freedom of the Hamiltonian operator after discretization, and N_e is the number of electrons in the system (spin degrees of freedom omitted). The eigenvectors associated with Λ are denoted by $\Psi = [\psi_1, \ldots, \psi_{N_e}] \in \mathbb{C}^{N\times N_e}$, and Ψ satisfies the orthonormality condition $\Psi^*\Psi = I_{N_e}$, where I_{N_e} is the identity of size N_e. The matrix P is a spectral projector, usually called the *density matrix*. The Hamiltonian $H[P]$ depends on the density matrix P in a nonlinear fashion, and (1.1) needs to be solved self-consistently.

Without loss of generality, the system part can be defined as the degrees of freedom associated with a set of indices $\mathcal{I}_s$ and the bath part with a set of indices $\mathcal{I}_b$ so that $\mathcal{I}_s \cup \mathcal{I}_b = \{1, \ldots, N\}$. Usually, $|\mathcal{I}_b| \gg |\mathcal{I}_s|$. We are mostly interested in the accurate computation of physical observables associated with the system part, i.e., the matrix block of the density matrix $P_{\mathcal{I}_s,\mathcal{I}_s}$. Since the eigenvalue problem (1.1) couples all degrees of freedom together, this task still requires a relatively accurate description of the rest of the density matrix.

In a nutshell, PET assumes the following decomposition of the density matrix:

$$P = P_s + P_{0,b}, \tag{1.2}$$

where P_s is the density matrix corresponding to the system part whose block corresponding to the bath part, $(P_s)_{\mathcal{I}_b,\mathcal{I}_b}$, approximately vanishes. Similarly, $P_{0,b}$, called the bath projector, is the density matrix from the bath part whose block corresponding to the system part, $(P_{0,b})_{\mathcal{I}_s,\mathcal{I}_s}$, approximately vanishes. The decomposition of the system and bath part is performed using projectors, thus leading to the name of PET. Such a decomposition can be in principle exact. The subscript 0 indicates that $P_{0,b}$ is computed from a *reference* system and thus is only obtained approximately. Furthermore, P_s is constrained by $P_{0,b}$ according to the orthogonality condition

$$P_{0,b}P_s = 0. \tag{1.3}$$

The condition (1.3) acts as a soft "boundary condition" for a modified Kohn–Sham problem, of which the number of eigenvectors to be computed can be much smaller

than N_e. Hence, PET reduces the computational cost compared to solving (1.1) by reducing the number of eigenvectors and eigenvalues to compute.

Related works:

From a practical perspective, PET can be seamlessly integrated into many electronic structure software packages, given its alluring matrix-free nature; i.e., PET only requires matrix-vector multiplication operation of the form $H\psi$. Thus, it can be applicable even when H is a dense matrix, such as in the planewave discretization, or when explicit access to H is not readily available.

This is in contrast to, e.g., the widely used Green's function embedding methods (see, e.g., [3, 37, 35, 19, 25]), where explicit access to H is usually required to compute Green's functions of the form $G(z) := (z - H)^{-1}$. In addition, when a large basis set such as a planewave discretization is used, even storing the Green's functions can be challenging. However, when a small basis set is used and H is a sparse matrix, Green's function embedding methods can be combined with fast algorithms [25] to yield a lower computational complexity than that of PET. It may also perform better for systems with small gaps.

Contribution:

The contribution of this paper is two fold. First, we provide a mathematical understanding of PET from a linear algebra perspective, which can be concisely stated as an energy minimization problem with extra orthogonality constraints. The corresponding Euler–Lagrange equation from the energy minimization problem gives rise to a modified Kohn–Sham problem, and the original PET formulation can be understood as a penalty method for implementing the extra orthogonality constraint (1.3). We then extend the formalism to the nonlinear case as in KSDFT.

Second, we found that the standard perturbation analysis cannot be applied directly to PET. However, through a proper choice of the basis set, it is possible to reformulate PET in a form suitable for such analysis, which allows us to compute a perturbative correction. In addition, we show that such correction only contributes to the bath part.[1]

Our numerical results for real chemical systems confirm the effectiveness of the method. In particular, we find that the method can reach below the chemical accuracy (1 kcal/mol, or 0.0016 au) even when applied to very small systems, such as the computation of the ground state energy of a SiH_3F molecule from the reference of a SiH_4 molecule. We also demonstrate the accuracy of the energy and the atomic force for the PET and the perturbatively corrected PET using other molecules, such as benzene and anthracene.

Organization:

This paper is organized as follows. We derive PET for linear problems in section 2 and introduce the first-order perturbative correction to PET in section 3. We then generalize the discussion to nonlinear problems in section 4. We discuss the strategy to evaluate the bath projector using localization methods in section 5. We then present the numerical results in section 6, followed by the conclusion and discussion in section 7.

2. PET for linear problems. We first introduce PET in the context of solving a linear eigenvalue problem. Let $H \in \mathbb{C}^{N\times N}$ be a Hermitian matrix, whose eigenvalues are ordered nondecreasingly as $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_{N_e} < \lambda_{N_e+1} \le \cdots \le \lambda_N$. Here we assume that there is a positive energy gap $\Delta_g = \lambda_{N_e+1} - \lambda_{N_e}$.

[1] We refer readers to the main text (in particular, section 3) for the formula of the aforementioned perturbation.

Consider the following energy minimization problem:

$$E = \inf_{\substack{P^2 = P,\, P^* = P \\ \mathrm{Tr}P = N_e}} \mathcal{E}[P], \tag{2.1}$$

where E is called the energy and the energy functional $\mathcal{E}[P]$ is defined as

$$\mathcal{E}[P] := \mathrm{Tr}[HP]. \tag{2.2}$$

Note that the condition $P = P^2$ requires P to be a projector, with eigenvalues being either 0 or 1. The trace condition ensures that there are precisely N_e eigenvalues that are equal to 1. Proposition 1 states that the minimizer is attained by solving a linear eigenvalue problem. This is a well-known result in linear algebra; nonetheless, we provide its proof here in order to motivate the derivation for PET later.

PROPOSITION 1. *Let $H \in \mathbb{C}^{N\times N}$ be a Hermitian matrix, and assume that there is a positive gap between the N_eth and (N_e+1)th eigenvalue of H. Then the variational problem* (2.1) *has a unique minimizer, denoted by P, which is given by the solution to the following linear eigenvalue problem:*

$$H\Psi = \Psi\Lambda, \quad P = \Psi\Psi^*. \tag{2.3}$$

Here (Ψ, Λ) are the lowest N_e eigenpairs of H.

Proof. Since H is a Hermitian matrix, it can be diagonalized as

$$H = \hat{\Psi}\hat{\Lambda}\hat{\Psi}^*. \tag{2.4}$$

Here $\hat{\Lambda} = \mathrm{diag}[\lambda_1, \ldots, \lambda_N] \in \mathbb{R}^{N\times N}$ is a diagonal matrix containing all the eigenvalues of H ordered nondecreasingly, and $\hat{\Psi} \in \mathbb{C}^{N\times N}$ is a unitary matrix with its first N_e columns given by Ψ. Then

$$\mathcal{E}[P] = \mathrm{Tr}[HP] = \mathrm{Tr}[\hat{\Psi}\hat{\Lambda}\hat{\Psi}^*P] = \mathrm{Tr}[\hat{\Lambda}\hat{\Psi}^*P\hat{\Psi}] = \mathrm{Tr}[\hat{\Lambda}\hat{P}] = \sum_{i=1}^{N} \lambda_i \hat{P}_{ii} =: \hat{\mathcal{E}}[\hat{P}], \tag{2.5}$$

where $\hat{P} = \hat{\Psi}^* P \hat{\Psi}$ is the density matrix with respect to the basis given by $\hat{\Psi}$. Thus, (2.1) is equivalent to

$$E = \inf_{\substack{\hat{P}^2 = \hat{P},\, \hat{P}^* = \hat{P} \\ \mathrm{Tr}\hat{P} = N_e}} \hat{\mathcal{E}}[\hat{P}]. \tag{2.6}$$

Since $\lambda_{N_e+1} - \lambda_{N_e} > 0$, the minimizer is achieved by setting

$$\hat{P}_{ii} = \begin{cases} 1, & \text{if } i \le N_e, \\ 0, & \text{if } i > N_e. \end{cases} \tag{2.7}$$

The last step in our proof is to show that $\hat{P}$ is a diagonal matrix, with ones and zeros at the main diagonal, i.e.,

$$\hat{P}_{ij} = \begin{cases} 1, & \text{if } i = j \text{ and } i \le N_e, \\ 0, & \text{otherwise}, \end{cases} \tag{2.8}$$

and that it is the unique minimizer. Consider the optimization problem

$$\max_{v^*v=1} v^*\hat{P}v. \tag{2.9}$$

The first-order optimality condition implies that any maximizer v must be an eigenvector of $\hat{P}$. Let e_i be the ith column of the identity matrix of size N. Since an eigenvalue of $\hat{P}$ can be only 0 or 1 and $\hat{P}_{ii} = e_i^* \hat{P} e_i = 1$, we find that e_i is a maximizer of (2.9) and thus is an eigenvector of $\hat{P}$ associated to the eigenvalue 1.

We can repeat the same argument for $\{e_i\}_{i=1}^{N_e}$ showing that they are all eigenvectors of $\hat{P}$. We can use the same argument but using a min instead of a max in (2.9) to show that $\{e_i\}_{i=N_e+1}^{N}$ are eigenvectors of $\hat{P}$ associated with the eigenvalue 0. Therefore, $\hat{P}$ is diagonalized by the identity matrix, and its diagonal entries are given by the eigenvalues as in (2.7).

Hence, by construction, $\hat{P}$ of the form (2.8) is the unique minimizer of (2.6). Then

$$P = \hat{\Psi} \hat{P} \hat{\Psi}^* = \Psi \Psi^* \tag{2.10}$$

is the unique minimizer of (2.1), where Ψ is given by the first N_e columns of $\hat{\Psi}$. ∎

When N and N_e are large, the solution of the linear eigenvalue problem (2.3) can be expensive. However, suppose that we have already solved the eigenvalue problem for a reference matrix H_0 and we would like to solve the eigenvalue problem for another matrix H such that $H - H_0$ is approximately zero outside the matrix block given by the index set $\mathcal{I}_s$. In such a case, PET aims at reducing the computational cost by solving a modified eigenvalue problem that involves a much smaller number of eigenvectors.

More specifically, for a reference system $H_0 \in \mathbb{C}^{N \times N}$, let P_0 be the minimizer of the following problem:

$$E_0 = \inf_{\substack{P^2 = P,\, P^* = P \\ \mathrm{Tr} P = N_e^0}} \mathrm{Tr}[H_0 P]. \tag{2.11}$$

We split the minimizer as

$$P_0 = P_{0,b} + P_{0,s}. \tag{2.12}$$

Here $P_{0,b}$ and $P_{0,s}$ are called system and bath projector, respectively, and are projectors themselves, i.e.,

$$P_{0,b}^2 = P_{0,b}, \quad P_{0,s}^2 = P_{0,s}. \tag{2.13}$$

The rank of $P_{0,b}$ is denoted by $N_b := \mathrm{Tr} P_{0,b}$. Here we use the symbol N_b instead of $N_{0,b}$ to emphasize that the rank of the bath projector $P_{0,b}$ remains the same before and after the perturbation, as will be seen in the discussion later. We assume that $N_b \approx N_e^0$, and hence the rank of $P_{0,s}$ is much smaller than N_b. The splitting procedure (2.12) is by no means unique; we will discuss one possible method based on localization techniques to choose $\Psi_{0,b}$ in section 5.

Together with $P_0^2 = P_0$, we have

$$P_0^2 = (P_{0,b} + P_{0,s})^2 = P_{0,b}^2 + P_{0,s}^2 + P_{0,b} P_{0,s} + P_{0,s} P_{0,b} = P_{0,b} + P_{0,s}. \tag{2.14}$$

Using (2.13), we have $P_{0,b} P_{0,s} + P_{0,s} P_{0,b} = 0$. Then

$$P_{0,b} P_{0,s} P_{0,s} + P_{0,s} P_{0,b} P_{0,s} = P_{0,b} P_{0,s} (I + P_{0,s}) = 0.$$

Since $I + P_{0,s}$ is invertible, we arrive at the orthogonality condition $P_{0,b} P_{0,s} = 0$. It is also convenient to write

$$P_{0,b} = \Psi_{0,b} \Psi_{0,b}^*, \quad \Psi_{0,b}^* \Psi_{0,b} = I_{N_b}. \tag{2.15}$$

By proper rotation[2] of the $\Psi_{0,b}$ matrix, without loss of generality we may assume that

$$\Psi_{0,b}^* H_0 \Psi_{0,b} := \Lambda_{0,b} \tag{2.16}$$

is a diagonal matrix. We define $\mathcal{B}_0 := \mathrm{span}\{\Psi_{0,b}\}$ with its orthogonal complement denoted by $\mathcal{B}_0^\perp$.

The main ansatz in PET is that the density matrix P can be split as

$$P = P_{0,b} + P_s, \tag{2.17}$$

where $P_s^2 = P_s$ is also a projector and $P_{0,b}$ is a bath projector as in (2.12). Similar to the discussion above, we arrive at the orthogonality condition (1.3). Since the rank of $P_{0,b}$ is already N_b, the rank of P_s is thus equal to $N_s := N_e - N_b$, and we expect that $N_s \ll N_b$. Note that the dimension of H_0 and H must be the same, but N_e^0 and N_e can be different. Thus, the ranks of $P_{0,s}$ and P_s can also be different. This is necessary in the context of KSDFT, where the system part can involve different numbers and/or types of atoms from that in the reference system.

With the bath projector fixed, PET solves the following constrained minimization problem only with respect to P_s:

$$E^{\mathrm{PET}} = \inf_{\substack{P_s^2 = P_s,\, P_s^* = P_s \\ P_{0,b} P_s = 0,\, \mathrm{Tr} P_s = N_s}} \mathrm{Tr}[H(P_s + P_{0,b})]. \tag{2.18}$$

Compared to (2.1), we find that PET restrains the feasibility set of density matrices to those satisfying the ansatz (2.17). Hence, by the variational principle E^{PET} provides an upper bound of the energy, i.e., $E^{\mathrm{PET}} \geq E$. Parallel to Proposition 1, the minimizer of (2.18) is uniquely obtained by a modified linear eigenvalue problem. This is given in Proposition 2.

PROPOSITION 2 (projection-based embedding). *Let $H|_{\mathcal{B}_0^\perp}$ be the restriction of H to the subspace $\mathcal{B}_0^\perp$, and assume that there is a positive gap between the N_sth and (N_s+1)th eigenvalue of $H|_{\mathcal{B}_0^\perp}$. Then the variational problem* (2.18) *has a unique minimizer, denoted by P_s, which is given by the solution to the following linear eigenvalue problem:*

$$H|_{\mathcal{B}_0^\perp} \Psi_s = \Psi_s \Lambda_s, \quad P_s = \Psi_s \Psi_s^*. \tag{2.19}$$

Here (Ψ_s, Λ_s) are the lowest N_s eigenpairs of $H|_{\mathcal{B}_0^\perp}$.

Proof. First, the orthogonality condition $P_{0,b} P_s = 0$ implies that all columns of P_s should be in the subspace $\mathcal{B}_0^\perp$. Using the relation

$$\begin{aligned} \mathrm{Tr}[(I - P_{0,b}) H (I - P_{0,b}) P_s] &= \mathrm{Tr}[HP_s - HP_{0,b}P_s - P_{0,b}HP_s + P_{0,b}HP_{0,b}P_s], \\ &= \mathrm{Tr}[HP_s - HP_{0,b}P_s - HP_sP_{0,b} + HP_{0,b}P_sP_{0,b}], \\ &= \mathrm{Tr}[HP_s], \end{aligned}$$

we find that (2.18) is equivalent to the following minimization problem:

$$E^{\mathrm{PET}} = \inf_{\substack{P_s^2 = P_s,\, P_s^* = P_s \\ P_{0,b} P_s = 0,\, \mathrm{Tr} P_s = N_s}} \mathrm{Tr}[(I - P_{0,b}) H (I - P_{0,b}) P_s] + \mathrm{Tr}[HP_{0,b}]. \tag{2.20}$$

[2] This can be achieved by solving the eigenvalue problem $(\Psi_{0,b}^* H_0 \Psi_{0,b}) C_b = C_b \Lambda_{0,b}$ and redefining $\Psi_{0,b}$ to be $\Psi_{0,b} C_b$.

Given that $\mathrm{Tr}[HP_{0,b}]$ is a constant, we only need to focus on the first term $\mathrm{Tr}[(I-P_{0,b})H(I-P_{0,b})P_s]$. The matrix $(I-P_{0,b})H(I-P_{0,b})$, sometimes called the Huzinaga operator in the quantum chemistry literature [17], is Hermitian and is identical to H when restricted to the subspace $\mathcal{B}_0^\perp$. With some abuse of notation, $H|_{\mathcal{B}_0^\perp}$ can be diagonalized as

$$H|_{\mathcal{B}_0^\perp} = \hat{\Psi}\hat{\Lambda}\hat{\Psi}^*. \tag{2.21}$$

Since the dimension of $\mathcal{B}_0^\perp$ is $N-N_b$, $\hat{\Lambda} = \mathrm{diag}[\hat{\lambda}_1, \ldots, \hat{\lambda}_{N-N_b}] \in \mathbb{R}^{(N-N_b)\times(N-N_b)}$ is a diagonal matrix containing all the eigenvalues of $H|_{\mathcal{B}_0^\perp}$ ordered nondecreasingly, and $\hat{\Psi} \in \mathbb{C}^{N\times(N-N_b)}$ is given by orthogonal columns of an unitary matrix in the subspace $\mathcal{B}_0^\perp$. Then

$$\begin{aligned}\mathrm{Tr}[(I-P_{0,b})H(I-P_{0,b})P_s] &= \mathrm{Tr}[\hat{\Psi}\hat{\Lambda}\hat{\Psi}^*P_s] = \mathrm{Tr}[\hat{\Lambda}\hat{\Psi}^*P_s\hat{\Psi}] \\ &= \mathrm{Tr}[\hat{\Lambda}\hat{P}_s] = \sum_{i=1}^{N-N_b} \hat{\lambda}_i \hat{P}_{ii}.\end{aligned}$$

Here $\hat{P}_s$ is the matrix representation of P_s with respect to the basis $\hat{\Psi}$ of the subspace $\mathcal{B}_0^\perp$.

Thus, similar to the proof of Proposition 1, we arrive at the minimization problem

$$E^{\mathrm{PET}} = \inf_{\substack{\hat{P}_s^2=\hat{P}_s,\, \hat{P}_s^*=\hat{P}_s \\ \mathrm{Tr}\hat{P}_s = N_s}} \mathrm{Tr}[\hat{\Lambda}\hat{P}_s] + \mathrm{Tr}[HP_{0,b}], \tag{2.22}$$

whose minimizer is given by

$$(\hat{P}_s)_{ij} = \begin{cases} 1, & \text{if } i=j \text{ and } i \le N_s, \\ 0, & \text{otherwise,} \end{cases} \tag{2.23}$$

and

$$P_s = \hat{\Psi}\hat{P}_s\hat{\Psi}^* = \Psi_s\Psi_s^*. \tag{2.24}$$

Here Ψ_s are the first N_s columns of $\hat{\Psi}$ corresponding to the lowest N_s eigenvalues. ∎

We note that, even if H_0 and H have a positive energy gap, it may not be necessarily the case for $H|_{\mathcal{B}_0^\perp}$. Therefore, we need to explicitly make this assumption in the proposition.

Furthermore, we point out that $\{\hat{\lambda}_i\}$, the eigenvalues of $H|_{\mathcal{B}_0^\perp}$, are not, in general, a subset of $\{\lambda_i\}$, the eigenvalues of H. Nonetheless, according to (2.22), E^{PET} can be computed in terms of the trace

$$E^{\mathrm{PET}} = \mathrm{Tr}[H(P_s+P_{0,b})] = \sum_{i=1}^{N_s} \hat{\lambda}_i + \mathrm{Tr}[HP_{0,b}],$$

which yields an upper bound to the energy E.

In addition, when computing Ψ_s, all the vectors $\Psi_{0,b}$ lie in the null space of $(I-P_{0,b})H(I-P_{0,b})$ and do not belong to the range of $H|_{\mathcal{B}_0^\perp}$; thus, they should be avoided in the computation. This issue becomes noticeable when $\hat{\lambda}_{N_s} > 0$, and it

would be incorrect to simply select the first N_s eigenpairs of $(I - P_{0,b})H(I - P_{0,b})$. One practical way to get around this problem is to add a negative shift c so that all the first N_s eigenvalues of the matrix $(I - P_{0,b})(H + cI)(I - P_{0,b})$ become negative.

This issue can also be automatically taken care of by applying the projector $I - P_{0,b}$ to the computed eigenvectors in an iterative solver so that the computation is restricted to the subspace of interest $\mathcal{B}_0^\perp$.

Remark 3. The original formulation of PET [28] can be understood as a penalty formulation to implement the orthogonality constraint, i.e.,

$$E^{\mathrm{PET},\mu} = \inf_{\substack{P_s^2 = P_s,\, P_s^* = P_s \\ \mathrm{Tr}P_s = N_s}} \mathrm{Tr}[H(P_s + P_{0,b})] + \mu \mathrm{Tr}[P_{0,b}P_s]. \tag{2.25}$$

This advantage of the penalty formulation is that the domain of P_s has the same form as that in Proposition 1 but with a modified energy functional. The corresponding Euler–Lagrange equation is given by the eigenvalue problem for the matrix $H + \mu P_{0,b}$, and P_s is the density matrix corresponding to the first N_s eigenpairs. Therefore, by selecting the penalty μ to be sufficiently large (in practice it is set to 10^6 or larger), the orthogonality condition is approximately enforced.

3. Perturbative correction to PET for linear problems. In the following, we define

$$\delta H := H - H_0 \tag{3.1}$$

and the PET projector

$$P^{\mathrm{PET}} = P_s + P_{0,b}, \tag{3.2}$$

where P_s is given by the solution of (2.19).

3.1. Consistency. First, we would like to verify that PET is a consistent theory: when $H = H_0$ and $N_e = N_e^0$, for any choice of the bath projector $P_{0,b}$, the minimizers from (2.1) and (2.18) should yield the same density matrix. This is ensured by Proposition 4.

PROPOSITION 4 (consistency of PET). *When $H = H_0$ and $N_e = N_e^0$, the solution to PET satisfies $P_s = P_{0,s}$.*

Proof. By the Courant–Fischer min-max theorem,

$$\hat{\lambda}_{N_s+1} = \max_{\substack{\dim(S)=N-N_e \\ S\subset\mathcal{B}_0^\perp}} \min_{u\in S\setminus\{0\}} \frac{u^* H|_{\mathcal{B}_0^\perp} u}{u^* u} \le \max_{\dim(S)=N-N_e} \min_{u\in S\setminus\{0\}} \frac{u^* H u}{u^* u} = \lambda_{N_e+1}.$$

Furthermore, when $H = H_0$, all eigenvectors of H corresponding to eigenvalues λ_{N_e+1} and above are in the subspace $\mathcal{B}_0^\perp$, and hence $\hat{\lambda}_{N_s+1} \ge \lambda_{N_e+1}$. Therefore,

$$\hat{\lambda}_{N_s+1} = \lambda_{N_e+1}.$$

Again using the Courant–Fischer min-max theorem, we have

$$\hat{\lambda}_{N_s} \le \lambda_{N_e}. \tag{3.3}$$

Hence, the gap condition of H, i.e., $\lambda_{N_e+1} - \lambda_{N_e} > 0$, implies that the gap condition for Proposition 2 holds, i.e.,

$$\hat{\lambda}_{N_s+1} - \hat{\lambda}_{N_s} > 0.$$

Since the minimizer of PET is obtained from a constrained domain of the density matrix, we have

$$E = \inf_{\substack{P^2 = P,\, P^* = P \\ \mathrm{Tr}P = N_e}} \mathrm{Tr}[H_0 P] \leq \inf_{\substack{P_s^2 = P_s,\, P_s^* = P_s \\ P_{0.b} P_s^* = 0,\, \mathrm{Tr}P_s = N_s}} \mathrm{Tr}[H_0(P_s + P_{0,b})], \tag{3.4}$$

where $P_s = P_{0,s}$ already achieves the minimum. By the uniqueness of the minimizer in Proposition 2, we have $P_s = P_{0,s}$. ∎

Remark 5. The proof of Proposition 4 is not entirely straightforward. This is mainly due to the fact that $P_{0,b}$ is obtained through some linear combination of eigenvectors of H_0 corresponding to the lowest N_e eigenvalues. Hence, H and $P_{0,b}$ generally do not commute even when $H = H_0$. Nonetheless, the consistency of PET implies that PET is an in principle exact theory, given the proper choice of the reference projector $P_{0,b}$.

Remark 6. Assume $N_e = N_e^0$; then we have that

$$\begin{aligned} \|P - P^{\mathrm{PET}}\| &\leq \|P - P_0 + P_0 - P_0^{\mathrm{PET}} + P_0^{\mathrm{PET}} - P^{\mathrm{PET}}\|, \\ &\leq \|P - P_0\| + \|P_0 - P_0^{\mathrm{PET}}\| + \|P_0^{\mathrm{PET}} - P^{\mathrm{PET}}\|, \\ &\leq \|P - P_0\| + \|P_0^{\mathrm{PET}} - P^{\mathrm{PET}}\|. \end{aligned}$$

Here we used $P_0 = P_0^{\mathrm{PET}}$ by Proposition 4, and $\|\cdot\|$ means the operator norm. In addition, given that H_0 has a positive energy gap and $\|\delta H\|$ is sufficiently small, we have, by continuity, that

$$\|P - P_0\| \sim \mathcal{O}(\|\delta H\|).$$

By the proof of Proposition 4, $H|_{\mathcal{B}_0^\perp}$ also has a positive gap. Using (3.2), we have that

$$\|P_0^{\mathrm{PET}} - P^{\mathrm{PET}}\| = \|P_s - P_{0,s}\| \sim \mathcal{O}(\|\delta H\|),$$

thus resulting in

$$\|P - P^{\mathrm{PET}}\| \sim \mathcal{O}(\|\delta H\|). \tag{3.5}$$

In addition, Proposition 4 states that when $\delta H = 0$, there is zeroth-order consistency for the energy $E = E^{\mathrm{PET}}$. Then we can use a standard perturbative argument coupled with the computation above and the fact that P^{PET} lies in the feasible set to obtain

$$|E - E^{\mathrm{PET}}| \sim \mathcal{O}(\|\delta H\|^2).$$

3.2. Perturbation. In the following discussion, we derive a perturbative correction to the density matrix when $H \approx H_0$. Unfortunately, standard perturbation analysis for eigenvalue problems does not apply directly given that PET depends on the solution of two separate eigenvalue problems: one from H_0, to determine the projector $P_{0,b}$, and other from $H|_{\mathcal{B}_0^\perp}$, to compute P_s.

In order to bypass this difficulty, through a proper choice of the basis set, the two eigenvalue problems can be formally combined into one. In this rotated basis, we use standard perturbative analysis to compute a perturbative correction. The results are then rotated back to the original basis.

Let us split the set of vectors $\hat{\Psi}$ from the eigendecomposition (2.21) as

$$\hat{\Psi} = [\Psi_s, \Psi_u],$$

where Ψ_s corresponds the projector P_s according to Proposition 2 and Ψ_u denotes the rest of the vectors. Here the subscript u stands for *unoccupied orbitals* following the terminology of KSDFT. Correspondingly, the diagonal matrix $\hat{\Lambda}$ is split into a block diagonal form as

$$\hat{\Lambda} = \begin{bmatrix} \Lambda_s & 0 \\ 0 & \Lambda_u \end{bmatrix}.$$

We combine $\hat{\Psi}$ and $\Psi_{0,b}$ from (2.16) to form a unitary $N \times N$ matrix

$$W := [\Psi_{0,b}, \Psi_s, \Psi_u], \tag{3.6}$$

and the matrix representation of H with respect to the basis W, denoted by H_W, can be written as

$$\begin{aligned} H_W = W^* H W &= \begin{bmatrix} \Psi_{0,b}^* H \Psi_{0,b} & \Psi_{0,b}^* H \Psi_s & \Psi_{0,b}^* H \Psi_u \\ \Psi_s^* H \Psi_{0,b} & \Psi_s^* H \Psi_s & \Psi_s^* H \Psi_u \\ \Psi_u^* H \Psi_{0,b} & \Psi_u^* H \Psi_s & \Psi_u^* H \Psi_u \end{bmatrix}, \\ &= \begin{bmatrix} \Psi_{0,b}^* H \Psi_{0,b} & \Psi_{0,b}^* H \Psi_s & \Psi_{0,b}^* H \Psi_u \\ \Psi_s^* H \Psi_{0,b} & \Lambda_s & 0 \\ \Psi_u^* H \Psi_{0,b} & 0 & \Lambda_u \end{bmatrix}. \end{aligned}$$

In the second equality, we have used the fact that all columns of Ψ_s, Ψ_u belong to $\mathcal{B}_0^\perp$ and consist of eigenvectors of $H|_{\mathcal{B}_0^\perp}$ associated with different sets of eigenvalues. Hence, the inner product $\Psi_s^* H \Psi_u$ vanishes. Again we note that not all off-diagonal matrix blocks vanish even when $H = H_0$.

From this perspective, we find that PET makes two approximations: First, it discards the off-diagonal matrix blocks so that

$$H_W \approx \begin{bmatrix} \Psi_{0,b}^* H \Psi_{0,b} & 0 & 0 \\ 0 & \Lambda_s & 0 \\ 0 & 0 & \Lambda_u \end{bmatrix},$$

and, second, it replaces the block corresponding to the interactions within the bath by $\Psi_{0,b}^* H_0 \Psi_{0,b}$, which is equal to $\Lambda_{0,b}$ following (2.16). The resulting matrix,

$$H_W^{\mathrm{PET}} = \begin{bmatrix} \Lambda_{0,b} & 0 & 0 \\ 0 & \Lambda_s & 0 \\ 0 & 0 & \Lambda_u \end{bmatrix},$$

is already diagonalized in the W-basis. Assume that the first N_e eigenvalues of H_W^{PET} include all the diagonal entries of $\Lambda_{0,b}$ (according to Proposition 4, this is at least valid when $H = H_0$ and $N_e = N_e^0$); we find that the density matrix in the W-basis takes the block diagonal form

$$P_W^{\mathrm{PET}} = \begin{bmatrix} I_{N_b} & 0 & 0 \\ 0 & I_{N_s} & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

When rotated back to the standard basis, the density matrix becomes

$$P^{\mathrm{PET}} = W P_W^{\mathrm{PET}} W^* = \Psi_{0,b} \Psi_{0,b}^* + \Psi_s \Psi_s^* = P_s + P_{0,b},$$

which is the PET solution.

One advantage of the representation in the W-basis is that the density matrix P_W^{PET} can be concisely written using the Cauchy contour integral formula as

$$P_W^{\mathrm{PET}} = \frac{1}{2\pi\imath}\oint_{\mathcal{C}} (zI - H_W^{\mathrm{PET}})^{-1}\,\mathrm{d}z. \tag{3.7}$$

Here $\mathcal{C}$ is a contour in the complex plane surrounding only the lowest N_e eigenvalues of H_W^{PET}.

The first-order perturbative correction to PET is then given by the neglected off-diagonal matrix blocks $\Psi_s^* H \Psi_{0,b}$ and $\Psi_u^* H \Psi_{0,b}$ and the diagonal term involving $\Psi_{0,b}^*(H - H_0)\Psi_{0,b}$. The formula for the first-order perturbation is given in Proposition 7.

PROPOSITION 7 (first-order perturbation). *The first-order perturbation to the density matrix from PET is given by*

$$\delta P = \delta\Psi_{0,b}\Psi_{0,b}^* + h.c., \tag{3.8}$$

where $\delta\Psi_{0,b} \in \mathbb{C}^{N\times N_b}$ satisfies the equation

$$Q\left(\lambda_{i;0,b} I - H\right) Q\delta\psi_{i;0,b} = Q(H\psi_{i;0,b}), \quad Q\delta\psi_{i;0,b} = \delta\psi_{i;0,b}. \tag{3.9}$$

Here the projector $Q = I - (P_s + P_{0,b}) = \Psi_u\Psi_u^*$; $\lambda_{i;0,b}$ is the ith diagonal element of $\Lambda_{0,b}$, and $\psi_{i;0,b}, \delta\psi_{i;0,b}$ are the ith column of $\Psi_{0,b}, \delta\Psi_{0,b}$, respectively, and h.c. stands for the Hermitian conjugate of the first term.

Remark 8. Following the previous notation, we may define the subspace $\mathcal{B} := \mathrm{span}\{\Psi_s, \Psi_{0,b}\}$, and Q is the projector on the orthogonal complement subspace $\mathcal{B}^\perp$. Since $\lambda_{0,b}$ is separated from the spectrum of $H|_{\mathcal{B}_0^\perp}$, (3.9) has a unique solution in $\mathcal{B}^\perp$. Equation (3.8) suggests that the first-order correction to the system part P_s vanishes, and the correction only comes from the bath part $P_{0,b}$. Furthermore, the correction is traceless due to the condition $\Psi_{0,b}^*\delta\Psi_{0,b} = 0$. This means that the density matrix after the first-order correction preserves the trace of the projector, which is N_e. In the context of KSDFT, this means that the first-order correction preserves the number of electrons in the system.

Remark 9. From (3.9), it may appear that the correction does not vanish even when $H = H_0$. However, note that $H_0\psi_{i;0,b} \in \mathcal{B} := \mathrm{span}\{\Psi_s, \Psi_{0,b}\}$, we have $QH_0\psi_{i;0,b} = 0$, and hence the first-order correction indeed vanishes. This is consistent with Proposition 4.

Remark 10. The perturbative correction requires the solution of N_b linear equations to correct the projector from the bath. It seems that this diminishes the purpose of PET, which reduces the number of eigenpairs to be computed from N_e to N_s from a practical perspective. Hence, the advantage of the perturbative correction becomes more apparent in the nonlinear setup in section 4, where the perturbation only needs to be applied once after the self-consistency is achieved.

Remark 11. We point out that (3.9) shares some similarities with the Sternheimer equation used in density functional perturbation theory [1]. The perturbative correction lies in the subspace orthogonal to the range of $P_s + P_{0,b}$. However, unlike the Sternheimer equation, in our case $\lambda_{i,0,b}$ is not necessarily an eigenvalue of H.

Proof. Our strategy is to derive the first-order perturbation in the W-basis, denoted by δP_W, and then obtain δP according to $\delta P = W\delta P_W W^*$.

Let us first denote by

$$\delta H_W^{\mathrm{PET}} = \begin{bmatrix} \Psi_{0,b}^* \delta H \Psi_{0,b} & \Psi_{0,b}^* H \Psi_s & \Psi_{0,b}^* H \Psi_u \\ \Psi_s^* H \Psi_{0,b} & 0 & 0 \\ \Psi_u^* H \Psi_{0,b} & 0 & 0 \end{bmatrix}$$

the neglected off-diagonal matrix blocks in PET; $\delta H_W^{\mathrm{PET}}$ may not be small even when $H = H_0$, but its contribution to the density matrix must vanish according to Proposition 4 and hence can be formally treated perturbatively.

Let $P_W = P_W^{\mathrm{PET}} + \delta P_W$. Setting $G_W(z) = (z - H_W)^{-1}$ and $G(z)_W^{\mathrm{PET}} = (z - H_W^{\mathrm{PET}})^{-1}$, we have the Dyson equation

$$G_W(z) = G_W^{\mathrm{PET}}(z) + G_W^{\mathrm{PET}}(z) \delta H_W^{\mathrm{PET}} G_W(z).$$

Thus, using the Cauchy integral formulation, we have that

$$\begin{aligned} \delta P_W &= P_W - P_W^{\mathrm{PET}}, \\ &= \frac{1}{2\pi\imath} \oint_{\mathcal{C}} G_W(z) - G_W^{\mathrm{PET}}(z) \, \mathrm{d}z, \\ &= \frac{1}{2\pi\imath} \oint_{\mathcal{C}} G_W^{\mathrm{PET}}(z) \delta H_W^{\mathrm{PET}} G_W(z) \, \mathrm{d}z, \\ &= \frac{1}{2\pi\imath} \oint_{\mathcal{C}} (zI - H_W^{\mathrm{PET}})^{-1} \delta H_W^{\mathrm{PET}} (zI - H_W)^{-1} \, \mathrm{d}z. \end{aligned}$$

By setting $H_W \approx H_W^{\mathrm{PET}}$, the first-order correction is

$$\delta P_W = \frac{1}{2\pi\imath} \oint_{\mathcal{C}} (zI - H_W^{\mathrm{PET}})^{-1} \delta H_W^{\mathrm{PET}} (zI - H_W^{\mathrm{PET}})^{-1} \, \mathrm{d}z.$$

Since H_W^{PET} is a diagonal matrix, δP_W should have the same matrix sparsity pattern as $\delta H_W^{\mathrm{PET}}$, i.e.,

$$\delta P_W = \begin{bmatrix} (\delta P_W)_{b,b} & (\delta P_W)_{b,s} & (\delta P_W)_{b,u} \\ (\delta P_W)_{s,b} & 0 & 0 \\ (\delta P_W)_{u,b} & 0 & 0 \end{bmatrix}.$$

First we compute

$$(\delta P_W)_{b,s} = \frac{1}{2\pi\imath} \oint_{\mathcal{C}} (zI - \Lambda_{0,b})^{-1} \Psi_{0,b}^* H \Psi_s (zI - \Lambda_s)^{-1} \, \mathrm{d}z.$$

Note that for any diagonal elements $\lambda_{i;0,b}, \lambda_{j;s}$ from $\Lambda_{0,b}, \Lambda_s$, respectively, they are both enclosed in the contour $\mathcal{C}$.

On the one hand, if $\lambda_{i;0,b} \neq \lambda_{j;s}$, then

$$\begin{aligned} \frac{1}{2\pi\imath} \oint_{\mathcal{C}} (z - \lambda_{i;0,b})^{-1} (z - \lambda_{j;s})^{-1} \, \mathrm{d}z &= \frac{1}{2\pi\imath} \oint_{\mathcal{C}} \frac{(z - \lambda_{i;0,b})^{-1} - (z - \lambda_{j;s})^{-1}}{\lambda_{i;0,b} - \lambda_{j;s}} \, \mathrm{d}z \\ &= \frac{1 - 1}{\lambda_{i;0,b} - \lambda_{j;s}} = 0. \end{aligned}$$

On the other hand, if $\lambda_{i;0,b} = \lambda_{j;s}$, then we would obtain an integral of the form

$$\frac{1}{2\pi\imath} \oint_{\mathcal{C}} (z - \lambda_{i;0,b})^{-2} \, \mathrm{d}z,$$

which vanishes since the residue for the integrand is zero.

For the term

$$(\delta P_W)_{b,b} = \frac{1}{2\pi\imath}\oint_{\mathcal{C}} (zI - \Lambda_{0,b})^{-1}\Psi_{0,b}^* \delta H \Psi_{0,b}(zI - \Lambda_{0,b})^{-1}\,\mathrm{d}z,$$

an analogous argument can be used to show that it vanishes.

This means that the matrix blocks $(\delta P_W)_{b,s}$, $(\delta P_W)_{s,b}$, and $(\delta P_W)_{b,b}$ vanish, and the only nonzero matrix blocks are $(\delta P_W)_{u,b}$ and its conjugate. Moreover,

$$\begin{aligned}(\delta P_W)_{u,b} &= \frac{1}{2\pi\imath}\oint_{\mathcal{C}} (zI - \Lambda_u)^{-1}\Psi_u^* H \Psi_{0,b}(zI - \Lambda_{0,b})^{-1}\,\mathrm{d}z,\\ &= \sum_i (\lambda_{i;0,b} - \Lambda_u)^{-1}\Psi_u^* H \psi_{i;0,b}.\end{aligned}$$

Back to the standard basis,

$$\begin{aligned}\delta P &= \Psi_u \sum_i (\lambda_{i;0,b} - \Lambda_u)^{-1}\Psi_u^* H \psi_{i;0,b}\psi_{i;0,b}^* + \text{h.c.}\\ &= \left(\sum_i \delta\psi_{i;0,b}\psi_{i;0,b}^*\right) + \text{h.c.}\end{aligned}$$

Here

$$\delta\psi_{i;0,b} = \Psi_u(\lambda_{i;0,b} - \Lambda_u)^{-1}\Psi_u^* H\psi_{i;0,b}.$$

Using the projector $Q = \Psi_u\Psi_u^*$, we find that $\psi_{i;0,b}$ satisfies (3.9), and we prove the proposition. ∎

We summarize the perturbatively corrected PET in Algorithm 3.1.

Algorithm 3.1 Perturbatively corrected PET for linear eigenvalue problems.

Input: H, H_0, $\Psi_{0,b}$, Ψ_s.
Output: $\delta\Psi_{0,b}, \delta P$.
1: Compute diagonal matrix $\Lambda_{0,b} = \Psi_{0,b}^* H_0 \Psi_{0,b}$.
2: Obtain the PET density matrix $P^{\mathrm{PET}} = \Psi_{0,b}\Psi_{0,b}^* + \Psi_s\Psi_s^*$.
3: Compute the right-hand side $R = (I - P^{\mathrm{PET}})H\Psi_{0,b}$.
4: Compute $\delta\Psi_{0,b}$ by solving (3.9).
5: Obtain the perturbation to the density matrix $\delta P = \delta\Psi_{0,b}\Psi_{0,b}^* + \Psi_{0,b}\delta\Psi_{0,b}^*$.

3.3. Comparison of PET with the Rayleigh–Schrödinger perturbation theory. Let us now have a more detailed comparison between the results from Proposition 7 with those from the standard Rayleigh–Schrödinger (RS) perturbation theory. For simplicity we consider the computation of the lowest, nondegenerate eigenpair (λ, ψ) corresponding to H. The perturbation is computed with respect to the lowest, nondegenerate eigenpair (λ_0, ψ_0) corresponding to the reference matrix H_0. We have

$$\lambda_0 = \psi_0^* H_0 \psi_0 = \mathrm{Tr}[H_0 P_0],$$

where the reference density matrix is $P_0 = \psi_0\psi_0^*$. The first-order correction to the eigenvalue is

$$\delta\lambda^{(1)} = \psi_0^*\delta H\psi_0 := \mathrm{Tr}[\delta H P_0],$$

and the first-order correction to the lowest eigenfunction can be computed as

$$\delta\psi^{(1)} = Q_0(\lambda_0 - H_0)^{-1}Q_0(\delta H\psi_0), \tag{3.10}$$

where $Q_0 = I - P_0$ projects to the subspace orthogonal to the range of P_0. This also gives the first-order correction of the density matrix as

$$\delta P^{(1)} = \delta\psi^{(1)}\psi_0^* + \psi_0(\delta\psi^{(1)})^*.$$

From (3.10), we have that $\delta\psi^{(1)}$ is orthogonal to ψ_0; thus, we can write

$$\delta P^{(1)}P_0 = \delta\psi^{(1)}\psi_0^*.$$

In addition, the first-order correction of the eigenfunction allows us to compute the second-order correction to the eigenvalue as

$$\delta\lambda^{(2)} = \psi_0^*\delta H\delta\psi^{(1)} = \mathrm{Tr}[P_0\delta H\delta P^{(1)}].$$

Let us then define

$$P^{(1)} := P_0 + \delta P^{(1)}, \quad \lambda^{(1)} := \lambda_0 + \delta\lambda^{(1)} = \mathrm{Tr}[HP_0],$$

and

$$\begin{aligned}\lambda^{(2)} &:= \lambda_0 + \delta\lambda^{(1)} + \delta\lambda^{(2)},\\ &= \mathrm{Tr}[P_0H] + \mathrm{Tr}[P_0\delta H\delta P^{(1)}],\\ &= \mathrm{Tr}[P_0HP_0] + \mathrm{Tr}[P_0H\delta P^{(1)}],\\ &= \mathrm{Tr}[P_0HP^{(1)}].\end{aligned} \tag{3.11}$$

Here we have used $\mathrm{Tr}[P_0H_0\delta P^{(1)}] = 0$.

To summarize, the RS perturbation theory states that

$$|\lambda - \lambda^{(1)}| \sim \mathcal{O}(\|\delta H\|^2),\ \|P - P^{(1)}\| \sim \mathcal{O}(\|\delta H\|^2),\ \text{and } |\lambda - \lambda^{(2)}| \sim \mathcal{O}(\|\delta H\|^3). \tag{3.12}$$

It is worth remarking that $\|P - P^{(1)}\| \sim \mathcal{O}(\|\delta H\|^2)$ does not imply $|\lambda - \lambda^{(2)}| \sim \mathcal{O}(\|\delta H\|^4)$. This is because the perturbed density matrix $P^{(1)}$ satisfies the symmetry and trace condition but not the idempotency condition as in the feasible set of the optimization problem (2.1). Therefore, the standard squared relation between the error of the eigenvalue and the error of the eigenfunction does not hold. In fact, (3.11) suggests that the eigenvalue computed to the correct order is not equal to $\mathrm{Tr}[HP^{(1)}]$ but to $\mathrm{Tr}[P_0HP^{(1)}]$.

Motivated from (3.11), we may define the perturbed energy in the PET formulation as

$$E^{\mathrm{pert}} := \mathrm{Tr}[P^{\mathrm{PET}}HP^{\mathrm{pert}}], \tag{3.13}$$

where

$$P^{\mathrm{pert}} := P^{\mathrm{PET}} + \delta P. \tag{3.14}$$

However, the perturbation theory used in Proposition 7 differs form the RS perturbation theory in the sense that the perturbation is performed with respect to

$\delta H_W^{\mathrm{PET}} = H_W - H_W^{\mathrm{PET}}$ rather than δH. In particular, $\delta H_W^{\mathrm{PET}}$ may not vanish even when $H = H_0$ unless

$$\Psi_{0,b}^* H_0 \Psi_{0,s} = 0. \tag{3.15}$$

In fact, (3.15) will be satisfied if the columns of $\Psi_{0,b}$ are eigenvectors of H_0. In such a case, the results of the perturbation theory of PET agree with those from the RS perturbation theory:

$$\|P - P^{\mathrm{pert}}\| \sim \mathcal{O}(\|\delta H\|^2), \quad |E - E^{\mathrm{pert}}| \sim \mathcal{O}(\|\delta H\|^3).$$

This will be confirmed by the numerical results.

However, when (3.15) is violated, $\|\delta H_W^{\mathrm{PET}}\|$ may not be small even when $\|\delta H\|$ is small, and the perturbation theory developed in Proposition 7 holds only formally. In such a case, the perturbation theory of PET does not improve the asymptotic convergence rate, and we have

$$\|P - P^{\mathrm{pert}}\| \sim \mathcal{O}(\|\delta H\|), \quad |E - E^{\mathrm{pert}}| \sim \mathcal{O}(\|\delta H\|^2).$$

Interestingly, our numerical results indicate that even when the perturbative correction is formal, the preconstant can be much reduced after the perturbation correction.

4. PET for nonlinear problems. In this section, we generalize PET and the perturbative expansion to the nonlinear case as in KSDFT. First, define the energy functional

$$\mathcal{E}[P] = \mathrm{Tr}[PH_L] + E_{\mathrm{Hxc}}[P], \tag{4.1}$$

where H_L is the linear part of the Hamiltonian and is a given matrix derived from the discretized Laplacian operator and the electron-nuclei interaction potential. $E_{\mathrm{Hxc}}[P]$ consists of the Hartree, and exchange correlation energy; it is a nonlinear functional of the density matrix P. Moreover, all the information of the quantum system, including the atomic types and positions, is given by the electron-nuclei interaction in H_L. The ground state energy of KSDFT can be obtained from the following variational problem:

$$E = \inf_{\substack{P^2 = P,\, P^* = P \\ \mathrm{Tr}P = N_e}} \mathcal{E}[P]. \tag{4.2}$$

Analogous to Proposition 1, the corresponding Euler–Lagrange equation is

$$H[P]\Psi = (H_L + V_{\mathrm{Hxc}}[P])\Psi = \Psi\Lambda, \quad P = \Psi\Psi^*, \tag{4.3}$$

where (Ψ, Λ) are the lowest N_e eigenpairs of the nonlinear Hamiltonian $H[P]$ and the functional derivative $V_{\mathrm{Hxc}}[P] = \frac{\delta E_{\mathrm{Hxc}}[P]}{\delta P}$ is called the exchange correlation potential. This is precisely (1.1). However, we remark that the procedure of taking the lowest N_e eigenpairs, which is called the *aufbau* principle in electronic structure theories, is not always valid. The *aufbau* principle has been found to be violated for certain model energy functionals [27], but numerical experience indicates that it generally holds in the context of KSDFT calculations for real materials. In the discussion below, we always assume the counterpart to Proposition 1 holds for the nonlinear problems under consideration.

According to the discussion in section 2, the key ansatz of the PET is that for some reference system with a different linear part of the Hamiltonian $H_{0,L}$, we have evaluated the density matrix and computed the projector $P_{0,b}$. Then for the system of interest, PET evaluates the modified variational problem by restricting the feasible set of the density matrix as

$$E^{\mathrm{PET}} = \inf_{\substack{P_s^2 = P_s,\, P_s^* = P_s \\ P_{0,b} P_s = 0,\, \mathrm{Tr} P_s = N_s}} \mathcal{E}[P_s + P_{0,b}]. \tag{4.4}$$

Analogous to Proposition 2, by assuming the corresponding *aufbau* principle, PET can be solved by the following nonlinear eigenvalue problem:

$$H[P]|_{\mathcal{B}_0^\perp} \Psi_s := \Psi_s \Lambda_s, \quad P_s = \Psi_s \Psi_s^*, \quad P = P_s + P_{0,b}. \tag{4.5}$$

Here (Ψ_s, Λ_s) are the lowest N_s eigenpairs of the self-consistent Hamiltonian $H[P]|_{\mathcal{B}_0^\perp}$.

The first-order perturbative correction to PET is entirely analogous to Proposition 7. According to Remark 10, the effectiveness of the perturbative approach mainly lies in the fact that it only needs to be applied once after (4.5) reaches self-consistency.

Once P^{pert} is obtained, we define the energy as

$$E^{\mathrm{pert}} := E^{\mathrm{PET}} + \mathrm{Tr}[P^{\mathrm{PET}} H[P^{\mathrm{PET}}] P^{\mathrm{pert}}]; \tag{4.6}$$

i.e., our correction of the energy is only at the linear level. We point out that (4.6) is only correct in the spinless or spin-unrestricted case. For spin-restricted calculations, a factor $1/2$ needs to the included in the correction.

In addition, we note that we can compute the atomic forces for the PET solution using the Hellmann–Feynman formula, which is due to the fact that the solution satisfies a variational principle. However, for the perturbation, the resulting approximation does not satisfy any variational principle; thus, we use an expensive finite difference approach to compute the forces. For the sake of consistency, we use a standard second-order finite difference scheme to approximate the force for both PET and the corrected approximation.

Remark 12. In [28], the Euler–Lagrange equation takes a slightly different form from (4.5). The connection with the present formulation can be established by noting that the energy functional satisfies the identity

$$\mathcal{E}[P_s + P_{0,b}] = \mathcal{E}[P_s] + \left(\mathcal{E}[P_s + P_{0,b}] - \mathcal{E}[P_s]\right).$$

Then the Euler–Lagrange equation gives the Hamiltonian

$$H_L + V_{\mathrm{Hxc}}[P_s] + (V_{\mathrm{Hxc}}[P_s + P_{0,b}] - V_{\mathrm{Hxc}}[P_s])$$

restricted to the subspace $\mathcal{B}_0^\perp$. The term in the parentheses, $V_{\mathrm{emb}}(P_s) := (V_{\mathrm{Hxc}}[P_s + P_{0,b}] - V_{\mathrm{Hxc}}[P_s])$, is called the "embedding potential," which can be interpreted as an external potential imposed onto the system part from the bath. For instance, in the absence of the exchange correlation, $V_{\mathrm{Hxc}} \equiv V_{\mathrm{H}}$ is a linear mapping. Then $V_{\mathrm{emb}} = V_{\mathrm{H}}[P_{0,b}]$ is the Coulomb interaction solely due to the projector from the bath.

5. Evaluation of the bath projector. The success of PET relies on a proper choice of the reference projector $P_{0,b}$. The suggestion from [28] is to compute a set of localized functions within the subspace $\mathrm{span}\{\Psi_0\}$ to evaluate $P_{0,b}$. For simplicity, we

use the notation from the linear problem, but the procedure can be directly generalized to the nonlinear setup as well.

Simply speaking, for a class of matrices H satisfying the gap condition, we may expect that the matrix elements of the density matrix P decays rapidly along the off-diagonal direction. In the physics literature, this is referred to as the "nearsightedness" principle [22, 31], and there is a rich literature studying the validity of such decay property (see, e.g., [4, 2]). We further expect that there exists a unitary matrix $U \in \mathbb{C}^{N_e \times N_e}$, called a gauge matrix, so that each column of the rotated matrix $\Phi = \Psi U$ is localized; i.e., it concentrates on a small number of elements compared to the size of the vector N. We point out that efficient numerical algorithms have been developed to compute such gauge and the corresponding localized functions (see, e.g., [11, 29, 9]). Once the localized functions are obtained, we may find localized functions associated with the index set for the bath $\mathcal{I}_b$ denoted by $\Psi_{0,b}$. To make the discussion self-contained, we briefly introduce the recently developed selected columns of the density matrix (SCDM) method [9] below as a simple and robust localization method to generate $P_{0,b}$. Other localization techniques can certainly be used as well.

The main idea of the SCDM procedure is that the localized function Φ is obtained directly from columns of the density matrix $P = \Psi\Psi^*$. However, picking N_e random columns of P may result in a poorly conditioned basis. In order to choose a well-conditioned set of columns, denoted $\mathcal{C} = \{c_1, c_2, \ldots, c_{N_e}\}$, we may use a QR factorization with column pivoting (QRCP) procedure [13]. More specifically, we compute

$$\Psi^*\Pi = U \begin{bmatrix} R_1 & R_2 \end{bmatrix}, \tag{5.1}$$

where Π is a permutation matrix so that R_1 is a well-conditioned matrix. The set $\mathcal{C}$ is given by the union of the nonzero row indices of the first N_e columns of the permutation matrix Π. The unitary matrix U is the desired gauge matrix [9, 10], and $\Phi = \Psi U$ is a localized matrix. It can be seen that (5.1) directly leads to a QRCP factorization of P as

$$P\Pi = \Psi\Psi^*\Pi = (\Psi U) \begin{bmatrix} R_1 & R_2 \end{bmatrix},$$

where ΨU is a matrix with orthogonal columns.

Let us apply the SCDM procedure to H_0 and its eigenfunctions Ψ_0. With some abuse of notation, from a predefined bath index set $\mathcal{I}_b \subset \{1, \ldots, N\}$, we may associate the ith column of Φ_0 to the bath degrees of freedom if the ith element of $\mathcal{C}$ is in $\mathcal{I}_b$. These selected vectors, denoted by $\Phi_{0,b}$ form the bath projector $P_{0,b}$. Finally, the condition (2.16) can be satisfied by solving the eigenvalue problem

$$\Phi_{0,b}^* H_0 \Phi_{0,b} C_{0,b} = C_{0,b} \Lambda_{0,b}, \tag{5.2}$$

and then $\Psi_{0,b} = \Phi_{0,b} C_{0,b}$. We summarize the procedure for computing the $\Psi_{0,b}$ in Algorithm 5.1.

Remark 13. We point out that after performing the localization in Algorithm 5.1, the vectors in the resulting bath orbitals, $\Psi_{0,b}$, are not eigenvalues of H_0. Thus, as shown in the following, the perturbative correction does not improve the asymptotic convergence rate; however, the preconstants are greatly reduced. In fact, as it will be shown in the numerical experiments, when the perturbation is relatively large, the perturbative correction associated with the rotated vectors $\Psi_{0,b}$ has a considerably smaller error than the one associated to the eigenvectors of H_0.

Algorithm 5.1 Using the SCDM algorithm for constructing the bath projector.

Input: H_0, Ψ_0, $\mathcal{I}_b$, N_e^0.

Output: $\Psi_{0,b}$.

1: Perform QRCP for Ψ_0^*: $\Psi_0^*\Pi = U\begin{bmatrix} R_1 & R_2 \end{bmatrix}$. The set $\mathcal{C}$ is given by the union of the nonzero row indices of the first N_e^0 columns of the permutation matrix Π.

2: Compute $\Phi_0 = \Psi_0 U$. Form a submatrix $\Phi_{0,b} := [\varphi_{i;0}]_{\mathcal{C}_i \in \mathcal{I}_b}$, where $\varphi_{i;0}$ is the ith column of Φ_0.

3: Solve the eigenvalue problem (5.2), and compute $\Psi_{0,b} = \Phi_{0,b} C_{0,b}$.

6. Numerical examples. We present several examples to demonstrate the effectiveness of the PET method and the perturbation scheme. The numerical tests were coded in MATLAB 2017b. For the solution of KSDFT in the nonlinear case, PET and the perturbative correction are implemented within the KSSOLV [36] software package. All calculations are performed in a dual socket server with Intel Xeon E5-2670 CPUs and 386 Gb of RAM.

6.1. Linear case. We first consider a simple Hamiltonian in one dimension with zero Dirichlet boundary conditions:

$$H_0 = -\frac{1}{2}\frac{d^2}{dx^2} + V_0(x), \quad V_0(x) := \sum_{i=1}^{3} -40e^{-100(x-\tilde{x}_i)^2}, \quad x \in [-1,1]. \tag{6.1}$$

Here the centers of the Gaussians $\tilde{x} = (-0.5, 0, 0.5)^T$. The resulting potential is shown in Figure 6.1. The one-dimensional Laplacian is discretized with a standard 3-point stencil finite difference scheme with 512 grid points.

For the reference problem, we evaluate the 3 eigenfunctions corresponding to the lowest 3 eigenvalues. As shown in Figure 6.2, the eigenvectors Ψ_0 are indeed delocalized across the entire interval $[-1, 1]$. After applying the SCDM algorithm (see Algorithm 5.1), the resulting orbitals Φ_0 become much more localized, as shown in Figure 6.2.

We define the new Hamiltonian by changing the height of the last Gaussian function as

$$H = -\frac{1}{2}\frac{d^2}{dx^2} + V(x), \quad V(x) := \sum_{i=1}^{2} -40e^{-100(x-\tilde{x}_i)^2} - 100e^{-100(x-\tilde{x}_3)^2}. \tag{6.2}$$

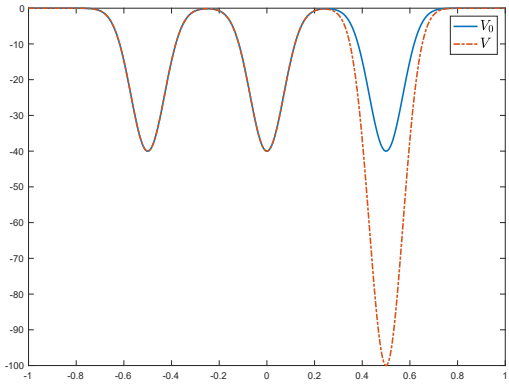
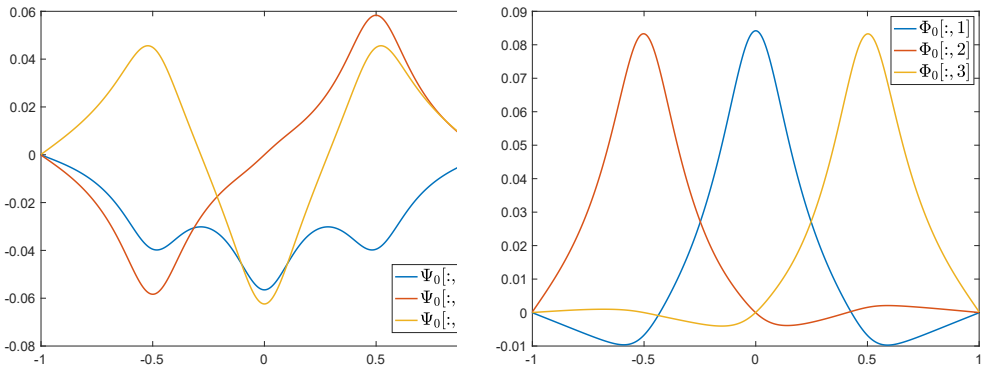
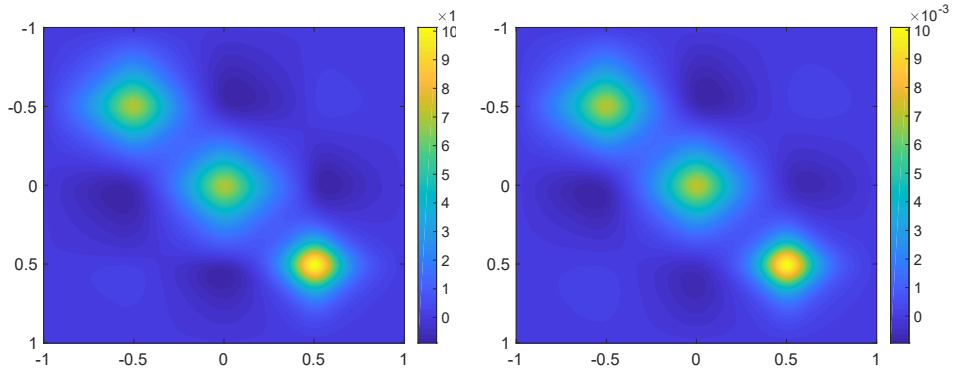


FIG. 6.1. *Potential for both H_0 and H.*



(a) (b)

FIG. 6.2. (a) *First* 3 *delocalized orbitals (columns of* Ψ_0*);* (b) *first* 3 *localized orbitals (columns of* Φ_0*).*



(a) (b)

FIG. 6.3. (a) *Exact density matrix;* (b) *PET density matrix.*

We observe that two columns in Φ_0 are localized far from the modified Gaussian, and we consider them as the bath orbitals. We set $\mathcal{I}_b = \{1, \ldots, 340\}$; this ensures that $\Phi_{0,b} = [\Phi_0[:,1], \Phi_0[:,2]]$. We then compute the linear PET problem to obtain Ψ_s, and we build the PET density matrix as shown in Figure 6.3, which is accurate up to 3 digits in relative error. Furthermore, the error is mostly localized around the third Gaussian function as one would expect. The relative error of the energy is 1.42×10^{-3}. We find it remarkable that for such a small system, the solution from PET is already very accurate despite the strong overlap of the system and bath orbitals.

Finally, we use Algorithm 3.1 to compute the perturbed density matrix,P^{pert}, which is more accurate than the PET density matrix without the perturbation, P^{PET}, as depicted in Figure 6.4. We can observe how the perturbation decreases the error in the density matrix by taking a look at the electron density, $\rho = \text{diag}(P)$, in Figure 6.5. In addition, the accuracy of the energy is improved, with its relative error reduced from 1.42×10^{-3} to 1.01×10^{-4}. If we increase the bath size from 1 to 2, the accuracy of the energy is improved further to 7.15×10^{-5} and 2.12×10^{-5}, without and with the perturbative correction, respectively.

In order to showcase the asymptotic convergence of PET and the first-order perturbation discussed at the end of section 3, we introduce a family of perturbed Hamil-

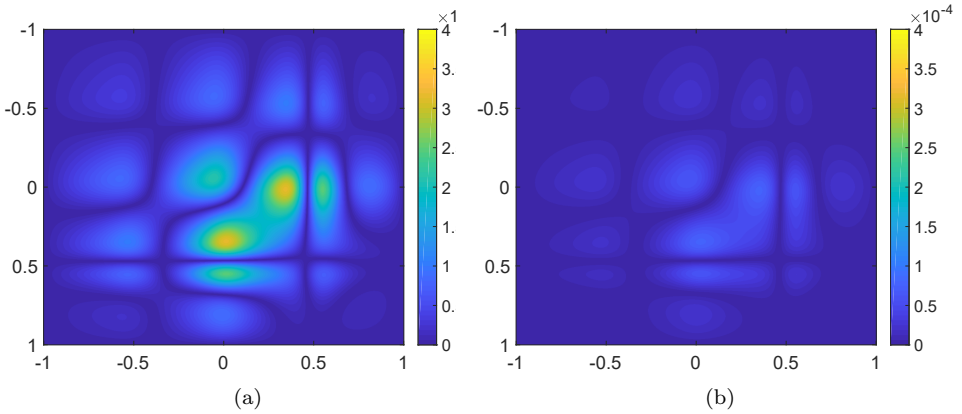


FIG. 6.4. *Error of the density matrix with respect to the reference answer:* (a) *PET density matrix;* (b) *PET density matrix plus the first-order perturbative correction.*

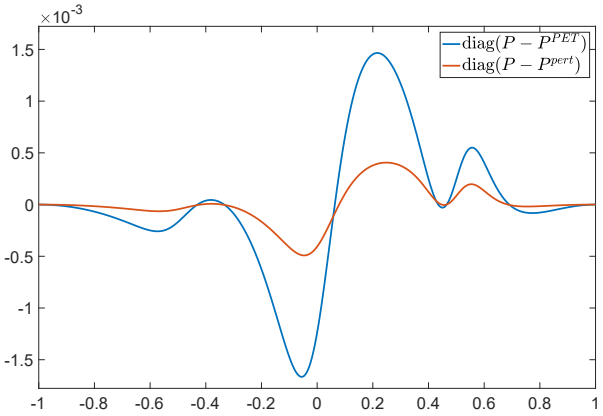


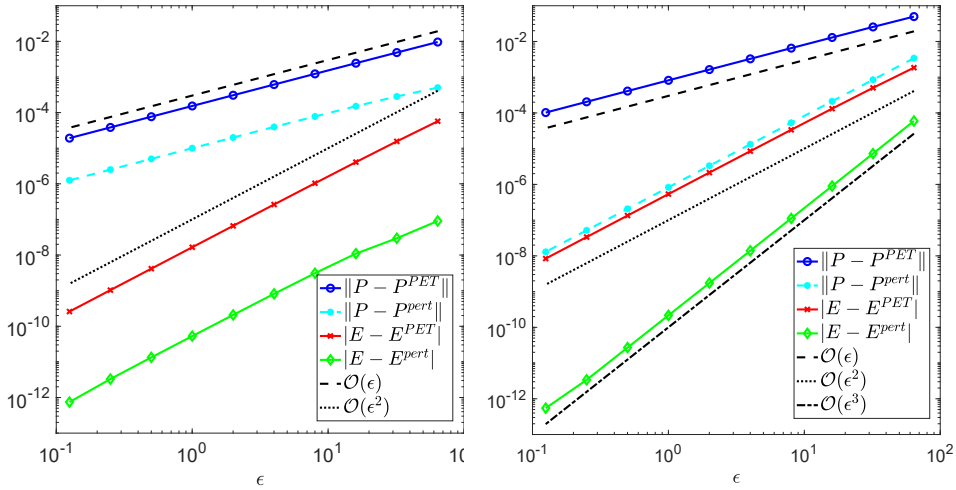
FIG. 6.5. *Relative error of the electron density using PET density matrix and the perturbed PET density matrix.*

tonians as

$$H_\epsilon = -\frac{1}{2}\frac{d^2}{dx^2} + V_\epsilon(x), \quad V_\epsilon(x) := \sum_{i=1}^{2} -40e^{-100(x-\tilde{x}_i)^2} - (40+\epsilon)e^{-100(x-\tilde{x}_3)^2}. \tag{6.3}$$

Then, in an analogous fashion as above, we compute the PET approximation and the associated perturbative correction for each Hamiltonian as $\epsilon \to 0$. Figure 6.6(a) shows the error of approximation of the density matrix and the energy as the ϵ tends to zero. As discussed in section 3, our first-order perturbation is computed with respect to $\delta H_W^{\mathrm{PET}}$, which can remain to be of $\mathcal{O}(1)$ even if $\delta H = 0$.

On the one hand, when we use localized orbitals to define the system and bath orbitals, (3.15) is not satisfied. In this case, the decay of the error of the PET density and energy is $\mathcal{O}(\epsilon)$ and $\mathcal{O}(\epsilon^2)$, respectively, as shown by Figure 6.6(a). Although the asymptotic convergence after first-order correction remains unchanged, the preconstants are significantly reduced by one to two orders of magnitude compared to the results of the PET.



(a) (b)

FIG. 6.6. *Error of the approximate density matrix and energy, with bath orbitals defined using* (a) *localized orbitals and* (b) *delocalized eigenfunctions.*

On the other hand, when we use the delocalized eigenfunctions to define the system and bath orbitals, (3.15) is satisfied. In such a case, Figure 6.6(b) shows that the error of the approximate density matrix and energy after the perturbation correction decay is $\mathcal{O}(\epsilon^2)$ and $\mathcal{O}(\epsilon^3)$, which agrees with results from the standard RS perturbation theory. However, the preconstants are larger than those in Figure 6.6(a). In particular, we can observe from Figure 6.6 that when the perturbation is relatively large, partitioning the system with spatially localized orbitals indeed improves the accuracy of PET, especially when the perturbative correction is used.

6.2. Nonlinear case. For KSDFT calculations, we modified the KSSOLV software package [36] to solve the PET equations (4.5) and to obtain the perturbation correction. KSSOLV uses a pseudospectral discretization with the plane wave set. All the operators, including Hamiltonian and projection operators, are efficiently implemented in a matrix-free fashion. Within each self-consistent field iteration, we use the locally optimal block preconditioned conjugate gradient method [21] to solve the linear eigenvalue problems. For the perturbative correction, we use the generalized minimal residual algorithm [32] method with a preconditioner [34] implemented via fast Fourier transforms (FFTs).

6.2.1. Silane. We first consider a simple molecule, silane (SiH_4), whose electron density is shown in Figure 6.7, and we performed three different numerical experiments to showcase the accuracy of the method. Our reference system is the silane molecule from an equilibrium configuration. The bath-system partition is shown in Figure 6.7, in which we can observe that we fixed three orbitals as the bath, induced by $\mathcal{I}_b$, and the system part, composed by the fourth orbital induced by $\mathcal{I}_s$ (delimited by the segmented red line). We performed three different modifications to the atom associated with the fourth orbital:

- We elongate one hydrogen bond by 25%.
- We replace a hydrogen atom by a chlorine atom (Cl).
- We replace a hydrogen atom by a fluorine atom (F).

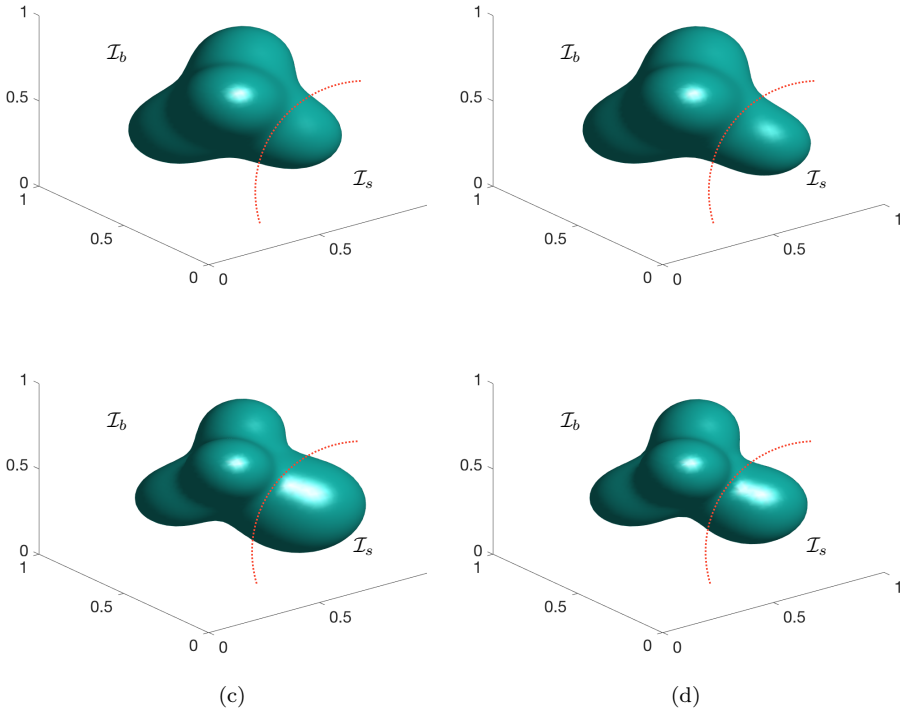


FIG. 6.7. (a) *Electron density of the silane molecule,* (b) *electron density of the* SiH_4 *molecule with one hydrogen bond elongated,* (c) *electron density of* SiH_3Cl*, and,* (d) *electron density of* SiH_3F*.*

Note that in the last two examples, the number of valence orbitals in the reference system is 4, while the number of valence orbitals in the perturbed systems are both 7. Hence, the perturbation introduced by the atom substitution is very large, especially compared to the small size of the molecule under study here.

We compare the results from PET and the perturbed PET against a reference solution obtained directly by solving the system in KSSOLV. In particular, we examine the relative error of the density matrices, the relative error of the electron density, the absolute error of the energy, and the absolute error of the atomic force at the modified location. All results are reported in atomic units. In particular, the unit of the energy is hartree, and the unit of the atomic force is hartree / bohr. In this case, the energy for PET was computed using the functional in (4.4). For the perturbed solution, we used (4.6). We used a second-order finite difference scheme to compute the forces at the perturbed atom.

The results for each of the experiments are shown in Tables 6.1 and 6.2. We can observe that the perturbation effectively reduces the error of the density matrix, the electron density, the energy, and atomic force. The only exception is the force of SiH_3F, which becomes coincidentally accurate for the PET, but the error after applying the perturbation theory is still around 10^{-3} au. Even for such a small system, after applying the perturbation formula, the error of the energy and force already reaches chemical accuracy.

6.2.2. Benzene. In this example, we show the performance of the method for a benzene molecule (C_6H_6) whose electron density is shown in Figure 6.8(a). We substitute one of the hydrogen atoms by a fluorine atom whose electron density is shown in

TABLE 6.1
Errors of the density matrices and electron densities for the different perturbation of the SiH_4 molecule.

Experiment	$\frac{\|\|P-P^{\mathrm{PET}}\|\|}{\|\|P\|\|}$	$\frac{\|\|P-P^{\mathrm{pert}}\|\|}{\|\|P\|\|}$	$\frac{\|\|\rho^{\mathrm{PET}}-\rho\|\|}{\|\|\rho\|\|}$	$\frac{\|\|\rho^{\mathrm{pert}}-\rho\|\|}{\|\|\rho\|\|}$
Elongated	6.03×10^{-2}	1.40×10^{-2}	1.24×10^{-2}	8.00×10^{-3}
SiH_3Cl	7.70×10^{-2}	1.74×10^{-2}	1.51×10^{-2}	6.71×10^{-3}
SiH_3F	9.12×10^{-2}	2.09×10^{-2}	6.64×10^{-3}	4.72×10^{-3}

TABLE 6.2
Errors for the different perturbation of the SiH_4 molecule.

Experiment	$E - E^{\mathrm{PET}}$	$E - E^{\mathrm{pert}}$	$F - F^{\mathrm{PET}}$	$F - F^{\mathrm{pert}}$
Elongated	5.98×10^{-3}	2.29×10^{-4}	1.51×10^{-2}	1.33×10^{-3}
SiH_3Cl	1.84×10^{-2}	1.94×10^{-3}	1.89×10^{-2}	2.72×10^{-3}
SiH_3F	1.66×10^{-2}	1.33×10^{-3}	9.29×10^{-5}	1.13×10^{-3}

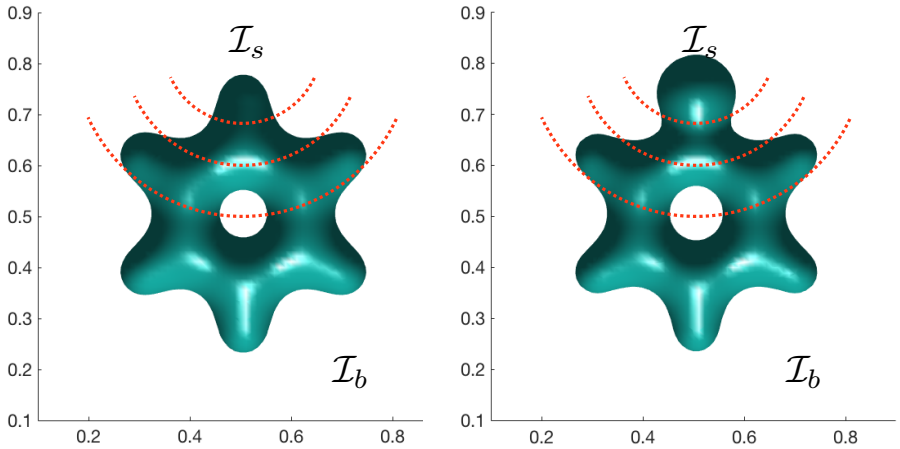


FIG. 6.8. (a) *Electron density for the benzene molecule and* (b) *the benzene molecule with an hydrogen atom replace by a fluorine.*

Figure 6.8(b). The benzene molecule has a total of 15 valence orbitals. To determine the partitions, we created a sphere centered at the replaced atom, and we performed the localization using Algorithm 5.1, where we labeled the different localized orbitals depending on the position of their associated pivots (from Algorithm 5.1). In particular, we labeled the orbitals whose pivots were within the sphere as system orbitals and the rest as bath orbitals. Tables 6.3 and 6.4 were generated by incrementally increasing the radius of the sphere until obtaining N_s system orbitals. The different partitions are depicted in Figure 6.8, for $N_s = 1, 4$ and 6, in which the segmented red line indicates the boundary between the bath and system partitions. Tables 6.3 and 6.4 show the errors of the density matrix and the electron density as well as the energy and the atomic force. We can observe a systematic decrease on the errors as the bath size decreases; i.e., N_s, the system size increases. When the system size is 7, the error of the energy and force after perturbative correction is already below the chemical accuracy and is as small as 1.02×10^{-4} and 4.17×10^{-4} au, respectively.

TABLE 6.3
Errors of the density matrix and electron density for the benzene molecule for different bath (and system) sizes.

N_s	$\frac{\|\|P-P^{\text{PET}}\|\|}{\|\|P\|\|}$	$\frac{\|\|P-P^{\text{pert}}\|\|}{\|\|P\|\|}$	$\frac{\|\|\rho^{\text{PET}}-\rho\|\|}{\|\|\rho\|\|}$	$\frac{\|\|\rho^{\text{pert}}-\rho\|\|}{\|\|\rho\|\|}$
1	9.41×10^{-2}	5.65×10^{-2}	1.36×10^{-2}	2.04×10^{-2}
3	6.32×10^{-2}	1.93×10^{-2}	5.56×10^{-3}	6.70×10^{-3}
5	6.46×10^{-2}	1.61×10^{-2}	4.41×10^{-3}	3.27×10^{-3}
7	5.04×10^{-2}	1.13×10^{-2}	2.93×10^{-3}	1.72×10^{-3}
9	2.98×10^{-2}	4.12×10^{-3}	1.56×10^{-3}	1.32×10^{-3}

TABLE 6.4
Errors of the energy and forces for the benzene molecule for different bath (and system) sizes.

N_s	$E - E^{\text{PET}}$	$E - E^{\text{pert}}$	$F - F^{\text{PET}}$	$F - F^{\text{pert}}$
1	4.05×10^{-2}	1.33×10^{-2}	2.69×10^{-2}	3.16×10^{-2}
3	1.87×10^{-2}	3.42×10^{-3}	1.73×10^{-2}	9.98×10^{-3}
5	1.20×10^{-2}	1.78×10^{-3}	4.24×10^{-3}	6.78×10^{-3}
7	7.89×10^{-3}	1.02×10^{-4}	3.73×10^{-3}	4.17×10^{-4}
9	3.02×10^{-3}	3.16×10^{-5}	4.05×10^{-3}	3.31×10^{-4}
11	2.81×10^{-3}	6.71×10^{-5}	3.68×10^{-3}	3.50×10^{-4}

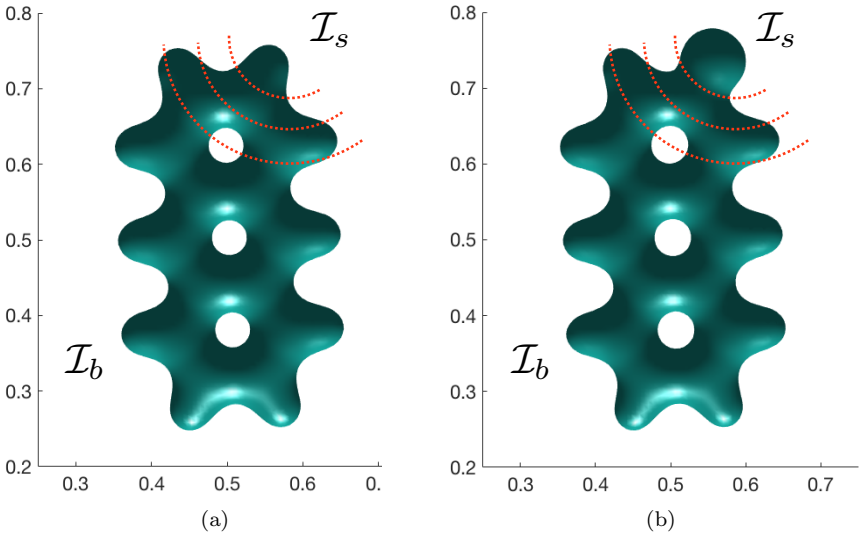


FIG. 6.9. (a) *Electron density for the anthracene molecule and* (b) *the anthracene molecule with an hydrogen atom replaced by a fluorine.*

6.2.3. Anthracene. Finally, we test our algorithm with the anthracene molecule ($C_{14}H_{10}$), which is composed of 3 benzene rings positioned longitudinally. Following the same procedure as with the benzene molecule, we compute the solution to the Kohn–Sham equations, whose electron density is shown in Figure 6.9, and we replace one hydrogen atom in one of the extremal rings by a fluorine atom (Figure 6.9(b)). From the total 33 orbitals for the anthracene, we define the bath orbitals and systems orbitals following the same procedure as for the benzene molecule. The partitions for $N_s = 1, 4$, and 6 are depicted in Figure 6.9, where the different segmented red lines indicate the boundary between the two partitions, in which they are denoted $\mathcal{I}_b$ and $\mathcal{I}_s$ for the bath and for the system, respectively.

TABLE 6.5
Errors for the anthracene molecule for different bath (and system) sizes.

N_s	$\frac{\|\|P-P^{\mathrm{PET}}\|\|}{\|\|P\|\|}$	$\frac{\|\|P-P^{\mathrm{pert}}\|\|}{\|\|P\|\|}$	$\frac{\|\|\rho^{\mathrm{PET}}-\rho\|\|}{\|\|\rho\|\|}$	$\frac{\|\|\rho^{\mathrm{pert}}-\rho\|\|}{\|\|\rho\|\|}$
1	8.65×10^{-2}	4.85×10^{-2}	1.59×10^{-2}	1.50×10^{-2}
3	6.10×10^{-2}	2.07×10^{-2}	5.47×10^{-3}	7.18×10^{-3}
5	5.01×10^{-2}	3.33×10^{-2}	3.24×10^{-3}	3.25×10^{-3}
7	4.42×10^{-2}	1.61×10^{-2}	2.47×10^{-3}	1.91×10^{-3}
9	3.31×10^{-2}	6.33×10^{-3}	1.13×10^{-3}	9.75×10^{-4}
11	2.88×10^{-2}	6.31×10^{-3}	1.12×10^{-3}	9.04×10^{-4}
13	2.86×10^{-2}	6.63×10^{-3}	1.03×10^{-3}	7.90×10^{-4}
15	1.83×10^{-2}	4.40×10^{-3}	6.73×10^{-4}	5.68×10^{-4}
19	1.73×10^{-2}	3.12×10^{-3}	5.23×10^{-4}	4.13×10^{-4}

TABLE 6.6
Errors of the energy and forces for the anthracene molecule for different bath (and system) sizes.

N_s	$E - E^{\mathrm{PET}}$	$E - E^{\mathrm{pert}}$	$F - F^{\mathrm{PET}}$	$F - F^{\mathrm{pert}}$
1	4.06×10^{-2}	1.33×10^{-2}	2.69×10^{-2}	3.16×10^{-2}
3	1.87×10^{-2}	3.42×10^{-3}	1.73×10^{-2}	9.98×10^{-3}
5	1.10×10^{-2}	1.97×10^{-3}	7.72×10^{-3}	6.78×10^{-3}
7	8.18×10^{-3}	2.02×10^{-4}	3.73×10^{-3}	9.79×10^{-4}
9	4.02×10^{-3}	3.16×10^{-5}	4.05×10^{-3}	3.31×10^{-4}
11	2.74×10^{-3}	2.53×10^{-5}	4.42×10^{-3}	2.64×10^{-4}
13	2.26×10^{-3}	6.35×10^{-5}	4.53×10^{-3}	2.15×10^{-4}
15	9.70×10^{-4}	1.15×10^{-5}	1.84×10^{-3}	2.83×10^{-4}
19	7.34×10^{-4}	5.16×10^{-6}	1.62×10^{-3}	1.02×10^{-4}

We compute the PET approximation and its perturbative correction for several different bath sizes, as shown in Tables 6.5 and 6.6. From Tables 6.5 and 6.6, we can clearly observe that the errors of all quantities decrease systematically with respect to the increase of the system size, and the perturbation method significantly increases the accuracy over the PET results. In particular, when the system size is 7, chemical accuracy is achieved after the perturbative correction is applied.

7. Conclusion. We have studied the recently developed PET from a mathematical perspective. Viewed as a method to approximately solve eigenvalue problems, PET solves a deflated eigenvalue problem by taking into account the knowledge from a related reference system. This deflated eigenvalue problem can be derived from the Euler–Lagrange equation of a standard energy minimization procedure with respect to the density matrices, with a non-standard constraint on the feasible set. From this perspective, the original formulation of PET can be seen as a penalty method for imposing the constraint. Numerical examples for linear problems as well as nonlinear problems from Kohn–Sham density functional theory calculations indicate that PET can yield accurate approximation to the density matrix, energy, and atomic forces. In order to further improve the accuracy of PET, we developed a first-order perturbation formula. We find that with the help of the perturbative treatment, PET can achieve chemical accuracy even for systems of relatively small sizes.

There are several immediate directions for future work. First, we have studied PET when the system and bath are treated using the same level of theory. From a physics perspective, it is more attractive to consider the case when the system part is treated with a more accurate theory than KSDFT with semilocal exchange correlation functionals. In particular, it would be interesting to understand PET when the system part is treated using KSDFT with nonlocal functionals, such as hybrid

functionals, or wavefunction theories, such as the coupled cluster method. It is also interesting to explore the PET in the context of solving time-dependent problems. Second, PET provides a size-consistent alternative for many methods in quantum physics and chemistry to be applied to solid-state systems. Some directions have already been pursued recently for using PET in the context of periodic systems [7, 26]. Third, the computation of the atomic force in PET is currently performed using the finite difference formula, which is expensive in practice. It would be desirable to develop a method with cost comparable to the Hellmann–Feynman method but without significant sacrifice of the accuracy. We note that there has been recent progress in this direction [24]. Finally, we believe that the asymptotic convergence property of PET is still dictated by the nearsightedness principle for systems satisfying the gap condition, but numerical results indicate that PET already achieves high accuracy even for system sizes that are well below the prediction from localization theories. Therefore, it is worthwhile to further study the convergence properties of PET as well as to perform further comparison with linear scaling–type methods.

Acknowledgments. We thank the Berkeley Research Computing (BRC) program at the University of California, Berkeley, for making computational resources available. We thank Garnet Chan and Frederick Manby for discussions and Joonho Lee for valuable suggestions and careful reading of the manuscript.

REFERENCES

[1] S. Baroni, S. de Gironcoli, A. Dal Corso, and P. Giannozzi, *Phonons and related crystal properties -+ from density-functional perturbation theory*, Rev. Modern Phys., 73 (2001), pp. 515–562.

[2] M. Benzi, P. Boito, and N. Razouk, *Decay properties of spectral projectors with applications to electronic structure*, SIAM Rev., 55 (2013), pp. 3–64.

[3] J. Bernholc, N. O. Lipari, and S. T. Pantelides, *Self-consistent method for point defects in semiconductors: Application to the vacancy in silicon*, Phys. Rev. Lett., 41 (1978), 895.

[4] C. Brouder, G. Panati, M. Calandra, C. Mourougane, and N. Marzari, *Exponential localization of Wannier functions in insulators*, Phys. Rev. Lett., 98 (2007), 046402.

[5] H. Chen and C. Ortner, *QM/MM Methods for Crystalline Defects. Part* 2: *Consistent Energy and Force-Mixing*, preprint, arXiv:1509.06627, 2015.

[6] W. Chibani, X. Ren, M. Scheffler, and P. Rinke, *Self-consistent Green's function embedding for advanced electronic structure methods based on a dynamical mean-field concept*, Phys. Rev. B, 93 (2016), 165106.

[7] D. V. Chulhai and J. D. Goodpaster, *Projection-based correlated wave function in density functional theory embedding for periodic systems*, J. Chem. Theory Comput., 14 (2018), pp. 1928–1942.

[8] P. Cortona, *Self-consistently determined properties of solids without band-structure calculations*, Phys. Rev. B, 44 (1991), 8454.

[9] A. Damle, L. Lin, and L. Ying, *Compressed representation of Kohn–Sham orbitals via selected columns of the density matrix*, J. Chem. Theory Comput., 11 (2015), pp. 1463–1469.

[10] A. Damle, L. Lin, and L. Ying, *Accelerating selected columns of the density matrix computations via approximate column selection*, SIAM J. Sci. Comput., 39 (2017), 1178-1198.

[11] J. M. Foster and S. F. Boys, *Canonical configurational interaction procedure*, Rev. Modern Phys., 32 (1960), 300.

[12] C.J. García-Cervera, J. Lu, and W. E, *Asymptotics-based sub-linear scaling algorithms and application to the study of the electronic structure of materials*, Commun. Math. Sci., 5 (2007), pp. 999–1024.

[13] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th ed., Johns Hopkins University Press, Baltimore, MD, 2013.

[14] J. D. Goodpaster, N. Ananth, F. R. Manby, and T. F. Miller III, *Exact nonadditive kinetic potentials for embedded density functional theory*, J. Chem. Phys., 133 (2010), 084103.

[15] P. Hohenberg and W. Kohn, *Inhomogeneous electron gas*, Phys. Rev., 136 (1964), pp. B864–B871.
[16] C. Huang, M. Pavone, and E. A. Carter, *Quantum mechanical embedding theory based on a unique embedding potential*, J. Chem. Phys., 134 (2011), 154110.
[17] S. Huzinaga and A. A. Cantu, *Theory of separability of many electron systems*, J. Chem. Phys., 55 (1971), pp. 5543–5549.
[18] A. A. Kananenka, E. Gull, and D. Zgid, *Systematically improvable multiscale solver for correlated electron systems*, Phys. Rev. B, 91 (2015), 121111.
[19] P. J. Kelly and R. Car, *Green's-matrix calculation of total energies of point defects in silicon*, Phys. Rev. B, 45 (1992), 6543.
[20] G. Knizia and G. K.-L. Chan, *Density matrix embedding: A strong-coupling quantum embedding theory*, J. Chem. Theory Comput., 9 (2013), pp. 1428–1432.
[21] A. V. Knyazev, *Toward the optimal preconditioned eigensolver: Locally optimal block preconditioned conjugate gradient method*, SIAM J. Sci. Comput., 23 (2001), pp. 517–541.
[22] W. Kohn, *Density functional and density matrix method scaling linearly with the number of atoms*, Phys. Rev. Lett., 76 (1996), pp. 3168–3171.
[23] W. Kohn and L. Sham, *Self-consistent equations including exchange and correlation effects*, Phys. Rev., 140 (1965), pp. A1133–A1138.
[24] S. J. R. Lee, F. Ding, F. R. Manby, and T. F. Miller III, *Analytical Gradients for Projection-Based Wavefunction-in-DFT Embedding*, preprint, arXiv:1903.05830, 2019.
[25] X. Li, L. Lin, and J. Lu, *PEXSI-σ: A Green's function embedding method for Kohn-Sham density functional theory*, Ann. Math. Sci. Appl., 3 (2018), 411.
[26] F. Libisch, M. Marsman, J. Burgdörfer, and G. Kresse, *Embedding for bulk systems using localized atomic orbitals*, J. Chem. Phys., 147 (2017), 034110.
[27] X. Liu, Z. Wen, X. Wang, M. Ulbrich, and Y. Yuan, *On the analysis of the discretized Kohn–Sham density functional theory*, SIAM J. Numer. Anal., 53 (2015), pp. 1758–1785.
[28] F. R. Manby, M. Stella, J. D. Goodpaster, and T. F. Miller III, *A simple, exact density-functional-theory embedding scheme*, J. Chem. Theory Comput., 8 (2012), pp. 2564–2568.
[29] N. Marzari and D. Vanderbilt, *Maximally localized generalized Wannier functions for composite energy bands*, Phys. Rev. B, 56 (1997), 12847.
[30] T. Nguyen, A. A. Kananenka, and D. Zgid, *Rigorous ab initio quantum embedding for quantum chemistry using Green's function theory: Screened interaction, nonlocal self-energy relaxation, orbital basis, and chemical accuracy*, J. Chem. Theory Comput., 12 (2016), pp. 4856–4870.
[31] E. Prodan and W. Kohn, *Nearsightedness of electronic matter*, Proc. Natl. Acad. Sci., 102 (2005), pp. 11635–11638.
[32] Y. Saad and M. H. Schultz, *GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Stat. Comput., 7 (1986), pp. 856–869.
[33] Q. Sun and G. K.-L. Chan, *Quantum embedding theories*, Accounts Chem. Res., 49 (2016), pp. 2705–2712.
[34] M. P. Teter, M. C. Payne, and D. C. Allan, *Solution of Schrödinger's equation for large systems*, Phys. Rev. B, 40 (1989), 12255.
[35] A. R. Williams, P. J. Feibelman, and N. D. Lang, *Green's-function methods for electronic-structure calculations*, Phys. Rev. B, 26 (1982), 5433.
[36] C. Yang, J. C. Meza, B. Lee, and L. W. Wang, *KSSOLV—A MATLAB toolbox for solving the Kohn–Sham equations*, ACM Trans. Math. Softw., 36 (2009), 10.
[37] R. Zeller and P. H. Dederichs, *Electronic structure of impurities in Cu, calculated self-consistently by Korringa-Kohn-Rostoker Green's-function method*, Phys. Rev. Lett., 42 (1979), 1713.
[38] D. Zgid and G. K.-L. Chan, *Dynamical mean-field theory from a quantum chemical perspective*, J. Chem. Phys., 134 (2011), 094115.