

## Approximating a wavefunction as an unconstrained sum of Slater determinants

Gregory Beylkin,<sup>1,a)</sup> Martin J. Mohlenkamp,<sup>2,b)</sup> and Fernando Pérez<sup>1,c)</sup>

<sup>1</sup>*Department of Applied Mathematics, University of Colorado at Boulder, 256 UCB, Boulder, Colorado 80309-0526, USA*

<sup>2</sup>*Department of Mathematics, Ohio University, 321 Morton Hall, Athens, Ohio 45701, USA*

(Received 25 July 2007; accepted 25 January 2008; published online 14 March 2008)

The wavefunction for the multiparticle Schrödinger equation is a function of many variables and satisfies an antisymmetry condition, so it is natural to approximate it as a sum of Slater determinants. Many current methods do so, but they impose additional structural constraints on the determinants, such as orthogonality between orbitals or an excitation pattern. We present a method without any such constraints, by which we hope to obtain much more efficient expansions and insight into the inherent structure of the wavefunction. We use an integral formulation of the problem, a Green's function iteration, and a fitting procedure based on the computational paradigm of separated representations. The core procedure is the construction and solution of a matrix-integral system derived from antisymmetric inner products involving the potential operators. We show how to construct and solve this system with computational complexity competitive with current methods. © 2008 American Institute of Physics. [DOI: 10.1063/1.2873123]

### I. INTRODUCTION

Given the difficulties in solving the multiparticle Schrödinger equation, current numerical methods in quantum chemistry/physics are remarkably successful. Part of their success comes from efficiencies gained by imposing structural constraints on the wavefunction to match physical intuition. However, such methods scale poorly to high accuracy and are biased to only reveal structures that were part of their own construction. Our goal is to develop a method that scales well to high accuracy and allows an unbiased exploration of the structure of the wavefunction. In this paper, we take a step toward this goal by developing a method to approximate the wavefunction as an unconstrained sum of Slater determinants.

Since the multiparticle fermionic wavefunction is an antisymmetric function of many variables, it is natural to approximate it as a sum of Slater determinants, at least as a first step. Motivated by the physical intuition that electrons may be excited into higher energy states, the configuration interaction (CI) family of methods chooses a set of determinants with predetermined orbitals and then optimizes the coefficients used to combine them. When it is found insufficient, methods to optimize the orbitals, work with multiple reference states, etc., are introduced (along with an alphabet of acronyms). A common feature of all these methods is that they impose some structural constraints on the Slater determinants, such as orthogonality of orbitals or an excitation pattern. As the requested accuracy increases, these structural constraints trigger an explosion in the number of determinants used, making the computation intractable for high accuracy.

The *a priori* structural constraints present in CI-like methods also force the wavefunction to comply with such structure, whether or not it really is the case. For example, if you use a method

<sup>a)</sup>Electronic mail: beylkin@colorado.edu.

<sup>b)</sup>Electronic mail: mjm@math.ohiou.edu.

<sup>c)</sup>Electronic mail: fernando.perez@colorado.edu.

that approximates the wavefunction as a linear combination of a reference state and excited states, you could not learn that the wavefunction is better approximated as a linear combination of several nonorthogonal, near-reference states. Thus, the choice of numerical method is not just a computational issue; it can help or hinder our understanding of the wavefunction.

For these reasons, our goal is to construct an adaptive numerical method without imposing *a priori* structural constraints besides that of antisymmetry. In this paper, we derive and present an algorithm for approximating a wavefunction with an unconstrained sum of Slater determinants, with fully adaptive single-electron functions. In particular, we discard the notions of reference state and excitation of orbitals. The functions comprising the Slater determinants need not come from a particular basis set, be orthogonal, or follow some excitation pattern. They are computed so as to optimize the overall representation. In this respect, we follow the philosophy of separated representations,<sup>7,6</sup> which allow surprisingly accurate expansions with remarkably few terms.

Our construction generates a solution using an iterative procedure based on nonlinear approximations via separated representations. To accomplish this nonlinear approximation, we derive a system of integral equations that describe the fully correlated many-particle problem. The computational core of the method is the repeated construction and solution of a matrix-integral system of equations.

Specifically, our approach has the following distinctive features:

- We use an adaptive representation for single-electron functions, but our method does not depend on its details.
- We use an integral formulation of the multiparticle Schrödinger equation and a Green's function iteration to converge to the ground-state wavefunction. The Green's function is decomposed and applied using separated approximations obtained by expanding the kernel into Gaussians.
- We use a variant of the so-called alternating least-squares algorithm to reduce the error of our approximation using a sum of a given number of Slater determinants.
- We compute antisymmetric inner products involving portions of the Hamiltonian operator by reducing them to formulas involving only combinations of standard integrals. In particular, we avoid the direct application of the electron-electron potential and instead compute convolutions with the Poisson kernel.

By doing this, we hope to represent the effects of correlations in the most natural and concise way possible, thus providing both computational efficiency and physical insight. We believe that this algorithm and the system of integral equations underlying it provide the foundation for a new approach to solving the multiparticle Schrödinger equation. We defer to the sequels several important issues, such as algorithmic size consistency/extensivity and the treatment of the interelectron cusp.

In Sec. II, we formulate the problem more carefully, make precise some of the statements that we made in this introduction, and give a high-level description of the method. We then present the derivations and proofs in the following sections.

## II. PROBLEM FORMULATION AND DESCRIPTION OF THE METHOD

### A. Formulation of the problem

We consider the time-independent, nonrelativistic, multiparticle Schrödinger equation and fix the nuclei according to the Born–Oppenheimer approximation, so the equation describes the steady state of an interacting system of electrons. For each of the  $N$  electrons in the system, there are three spatial variables  $\mathbf{r}=(x,y,z)$  and a discrete spin variable  $\sigma$  taking the values  $\{-\frac{1}{2}, \frac{1}{2}\}$ , which we combine and denote  $(\mathbf{r}, \sigma)$  by  $\gamma$ . The Hamiltonian operator  $\mathcal{H}$  is a sum of a kinetic energy operator  $\mathcal{T}$ , a nuclear potential operator  $\mathcal{V}$ , and an electron-electron interaction operator  $\mathcal{W}$ , defined in atomic units by

$$\mathcal{H} = T + \mathcal{V} + \mathcal{W} = -\frac{1}{2} \sum_{i=1}^N \Delta_i + \sum_{i=1}^N v(\mathbf{r}_i) + \frac{1}{2} \sum_{i=1}^N \sum_{j \neq i}^N \frac{1}{\|\mathbf{r}_i - \mathbf{r}_j\|}, \quad (1)$$

where  $\Delta_i$  is the three-dimensional Laplacian acting in the variable  $\mathbf{r}_i$  and  $v(\mathbf{r})$  is a sum of terms of the form  $-Z_a/\|\mathbf{r}-\mathbf{R}_a\|$  from a nucleus at position  $\mathbf{R}_a$  with charge  $Z_a$ . The antisymmetric eigenfunctions of  $\mathcal{H}$  represent electronic states of the system and are called wavefunctions. Antisymmetric means that under the exchange of any two coordinates, the wavefunction is odd, e.g.,  $\psi(\gamma_2, \gamma_1, \dots) = -\psi(\gamma_1, \gamma_2, \dots)$ . The bound-state wavefunctions have negative eigenvalues and are of greatest interest. We will focus on the ground-state wavefunction, which has the most negative eigenvalue, although the techniques can be used for other states. In summary, our goal is to find  $E$  and  $\psi$ , with  $E$  the most negative eigenvalue in

$$\mathcal{H}\psi = E\psi, \quad (2)$$

subject to the antisymmetry condition on  $\psi$ . Analytic methods can give qualitative results about the solutions and determine limiting cases, but most quantitative results must be obtained numerically. Although the equation is “just” an eigenvalue problem, its numerical solution presents several serious difficulties, among them the large number of variables and the antisymmetry condition on the solution. The simplest method that addresses these two difficulties is Hartree–Fock (HF) (see, e.g., Ref. 31), which uses the antisymmetrization of a single product, called a *Slater determinant*, to approximate the  $N$ -particle wavefunction, i.e.,

$$\psi_{\text{HF}} = \mathcal{A} \prod_{i=1}^N \phi_i(\gamma_i) = \frac{1}{N!} \begin{vmatrix} \phi_1(\gamma_1) & \phi_1(\gamma_2) & \cdots & \phi_1(\gamma_N) \\ \phi_2(\gamma_1) & \phi_2(\gamma_2) & \cdots & \phi_2(\gamma_N) \\ \vdots & \vdots & & \vdots \\ \phi_N(\gamma_1) & \phi_N(\gamma_2) & \cdots & \phi_N(\gamma_N) \end{vmatrix}. \quad (3)$$

Any antisymmetric approximation  $\tilde{\psi}$  to the wavefunction  $\psi$  can be substituted into

$$\frac{\langle \mathcal{H}\tilde{\psi}, \tilde{\psi} \rangle}{\langle \tilde{\psi}, \tilde{\psi} \rangle}, \quad (4)$$

where  $\langle \cdot, \cdot \rangle$  is the usual inner product, to obtain an estimate for  $E$ . This estimate gives an upper bound on the lowest value of  $E$  that solves (2). Substituting (3) into (4), one can iteratively solve for  $\phi_i$  to minimize (4). The resulting  $\psi_{\text{HF}}$  will best approximate  $\psi$  in the sense of providing the best estimate (4).

To improve upon HF, it is natural to consider the antisymmetrization of a sum of products:

$$\psi_{(r)} = \mathcal{A} \sum_{l=1}^r s_l \prod_{i=1}^N \phi_i^l(\gamma_i), \quad (5)$$

which could also be written as a sum of Slater determinants. The coefficients  $s_l$  are introduced in order to have  $\|\phi_i^l\|=1$ . Many methods are based on this form, but they use it in different ways. The CI method (see, e.g., Ref. 58) chooses the functions  $\phi_i^l$  from a preselected master set of orthogonal functions and decides on a large number  $r$  of combinations to consider based on excitation level. Substituting (5) into (4) leads to a matrix eigenvalue problem that can be solved for the scalar coefficients  $s_l$ . The multiconfiguration self-consistent field method (e.g., Refs. 22 and 12) solves for the master set of orthogonal functions as well as the scalar coefficients. There are numerous variations and combinations of these methods, too many to describe here.

### 1. What is new here

In this work, we construct and demonstrate a method that also uses a wavefunction of the form (5) but without constraints on the  $\phi_i^l$ . We remove both structural constraints, such as an excitation pattern or orthogonality between single-electron functions, and representation constraints, such as those imposed by using a predetermined basis set.

Many methods (e.g., Refs. 56, 48, 40, 2, 21, 18, 20, 3, 17, 61, and 42) have loosened the constraints on the Slater determinants in one way or another, often with encouraging results. These works, however, only partially removed the constraints, and so, we claim, did not achieve the full potential of an unconstrained approximation. By removing these constraints, we hope to produce much better approximations at much smaller separation rank  $r$  than existing methods allow. We also hope to provide a new perspective from which to analyze and understand the wavefunction, free from the biases that physical intuition imposes.

Our hopes are based on our work in Refs. 6, 7, and 44 where we developed general methods to represent and compute with functions and operators in many dimensions. We used sums of separable functions, dubbed *separated representations*, similar to (5). We found rather natural examples where removing constraints produces expansions that are *exponentially* more efficient, i.e.,  $r=N$  instead of  $2^N$  or  $r=\log N$  instead of  $N$ . For example, in our approach we can have a two-term representation:

$$\psi = \mathcal{A} \prod_{i=1}^N \phi_i(\gamma_i) + \mathcal{A} \prod_{i=1}^N (\phi_i(\gamma_i) + \phi_{i+N}(\gamma_i)), \quad (6)$$

where  $\{\phi_j\}_{j=1}^{2N}$  form an orthogonal set. To represent the same function as (6) while imposing the constraint that factors come from a master orthogonal set would force one to multiply out the second term, and thus use a representation with  $2^N$  terms.

At present, we have no proof that the wavefunction is well approximated by a structure that would benefit from the removal of constraints. The size  $r$  needed in practice and how it depends on the various parameters in the problem are thus still open questions. In Refs. 7, 6, and 44 the most interesting examples came from “reverse engineering” the numerical results to obtain formulas and proofs. We therefore expect that the tools we provide here will allow an exploration of the wavefunction, perhaps revealing an unexpected structure, and a strategy for a proof.

## B. Description of the algorithm

The removal of constraints in (5), and thus the basis sets, coefficients, and other structures that went along with them, also eliminates the conventional strategies for constructing (5) to minimize (4). It leads one to consider how one would compute the ground-state wavefunction if its numerical representation were not an issue. We choose to use an integral iteration, which we sketch in Sec. II B 1. In Appendix A, we sketch an alternative iteration based on gradient descent.

To use the form (5), we must choose some value of  $r$ , which determines the quality of the approximation. In Section II B 2, we show how to incorporate a nonlinear fitting step within the integral iteration in order to maintain fixed  $r$ . Accomplishing this fitting requires a significant amount of machinery, which makes up the body of the paper. Eventually, one would want to adaptively determine  $r$ , but we do not address that issue here.

### 1. A Green's function iteration

The eigenvalue equation (2) contains the differential operator  $\mathcal{H}$ , which has both the discrete negative eigenvalue(s) that we are interested in and unbounded, continuous, positive spectrum. In Refs. 33 and 34, this differential equation was reformulated as an integral equation. An iteration based on the integral formulation with Green's functions was also introduced in Refs. 33 and 34 and used in, e.g., Refs. 13 and 29.

In this section, we review this iteration, formulated at the continuous level. The actual implementation of the algorithm differs from this continuous formulation in three ways. First, and most

importantly, in Sec. II B 2, we modify the iteration to preserve our wavefunction representation (5). Second, we approximate the Green's function using Gaussians to some fixed but arbitrary accuracy in operator norm (Sec. V A). Third, we delegate the task of representing and operating on functions of the single-electron variable  $\gamma$  to an adaptive numerical method (see Ref. 5) with its own accuracy control. We note that an alternative to Ref. 5 may be used for this purpose as long as it is adaptive and has controlled accuracy. In this paper, we do not address some technical issues, such as finding a good initial guess or determining stopping criteria.

Define the Green's function

$$\mathcal{G}_\mu = (T - \mu\mathcal{I})^{-1} \quad (7)$$

for  $\mu < 0$  and consider the Lippmann–Schwinger integral equation

$$\lambda_\mu \psi_\mu = -\mathcal{G}_\mu[(\mathcal{V} + \mathcal{W})\psi_\mu]. \quad (8)$$

The subscript  $\mu$  on  $\lambda_\mu$  and  $\psi_\mu$  are to emphasize the dependence of the eigenvalues and eigenfunctions on  $\mu$ . The operator  $\mathcal{G}_\mu[(\mathcal{V} + \mathcal{W})]$  is bounded (see Refs. 35, 55, and 53) but without additional assumptions, it is not compact (see Ref. 54 Sec. XIII 5). Since we are interested in a bound state, we assume that the operators act on a bounded domain, as is justified by the exponential decay of the wavefunction.<sup>1</sup> Under this assumption,  $\mathcal{G}_\mu[(\mathcal{V} + \mathcal{W})]$  is compact, so (8) has only a discrete spectrum.

If  $\mu = E$ , then there is an eigenvalue  $\lambda_\mu = 1$  and the corresponding eigenfunction  $\psi_\mu$  of (8) is the desired ground-state eigenfunction of (2), as one can see by rearranging (8) into (2). One can show that  $\lambda_\mu = 1$  is the largest eigenvalue (see Ref. 45), so a simple iteration such as the power method yields the desired ground-state eigenfunction.

The eigenvalues  $\lambda_\mu$  depend analytically on  $\mu$ , so when  $\mu$  is sufficiently close to  $E$ , the power method will still yield an eigenfunction of (8) with energy near the minimum of (4). The convergence rate of the power method to produce  $\psi_\mu$  and  $\lambda_\mu$  is linear, and depends, as usual, on the gap between the eigenvalues in (8). From  $\psi_\mu$  and  $\lambda_\mu$ , one can then compute an improved estimate  $\mu$  for  $E$ . In the practical use of this approach, one does not wait for the power method to converge at each step but instead intertwines it with the update of  $\mu$ . Beginning with an approximation to the energy  $\mu_0 \approx E$  and an approximate wavefunction  $\psi_0$ , one converts (8) to an iteration

$$\tilde{\psi}_n = -\mathcal{G}_{\mu_n}[(\mathcal{V} + \mathcal{W})\psi_n]. \quad (9)$$

After each iteration, one normalizes by setting

$$\psi_{n+1} = \tilde{\psi}_n / \|\tilde{\psi}_n\|. \quad (10)$$

Following the approach of Ref. 29, we can use the update rule

$$\mu_{n+1} = \mu_n - \langle (\mathcal{V} + \mathcal{W})\psi_n, \psi_n - \tilde{\psi}_n \rangle / \|\tilde{\psi}_n\|^2. \quad (11)$$

*Remark 1:* An eigenfunction of (8) is an eigenfunction of (2) only when  $\mu$  is an eigenvalue of (2) and the eigenfunction of (8) has eigenvalue  $\lambda = 1$ . To obtain the next eigenfunction of (2) above the ground state, it appears that one needs to compute the first two eigenfunctions of (8), the first with eigenvalue  $\lambda > 1$  and the second with eigenvalue  $\lambda = 1$ . We will develop such deflation procedure elsewhere.

## 2. Approximating with fixed separation rank $r$

We restrict the method to approximate wavefunctions of the form (5), with  $r$  fixed, by replacing the definition of  $\tilde{\psi}_n$  in (9). We would like to redefine  $\tilde{\psi}_n$  to be the function of the form (5) that minimizes the (least-squares) error

$$\|\tilde{\psi}_n - (-\mathcal{G}_{\mu_n}[(\mathcal{V} + \mathcal{W})\psi_n])\|. \quad (12)$$

In general, such minimization problems may be ill posed in the sense that the infimum may occur at a limit point (see Ref. 16 and the references therein). In Ref. 6, we describe a method to balance the least-squares error with the loss-of-precision error due to a large condition number, so that the problem becomes well posed. However, even then, there are no known algorithms that assure convergence to the global minimum. Instead, we settle for an algorithm that, at each iteration, constructs a  $\tilde{\psi}_n$  with lower value of (12), unless it is already at a minimum.

Since fitting using (12) instead of directly using (9) introduces an error, the update rule (11) may no longer give a quadratic convergence and, in any case, is not expected to converge to the true energy. One may choose to replace the update rule (11) with the more robust but slower converging rule

$$\mu_{n+1} = \frac{\langle \mathcal{H}\psi_{n+1}, \psi_{n+1} \rangle}{\|\psi_{n+1}\|^2}, \quad (13)$$

which is based on (4). Other rules may be possible as well. At present, we do not have enough numerical experience to decide which rule to prefer.

The Green's function iteration itself does not enforce the antisymmetry condition. In order to assure convergence to an antisymmetric solution, we use the pseudonorm induced by the pseudo-inner-product  $\langle \cdot, \cdot \rangle_{\mathcal{A}} = \langle \mathcal{A}(\cdot), \mathcal{A}(\cdot) \rangle$ , as we did in Ref. 6.

We now describe our algorithm for constructing  $\tilde{\psi}_n$  of the form (5) to reduce, in an attempt to minimize, the least-squares error (12). To simplify notation in the description of our method, we now suppress the index  $n$  in (12) and consider a single problem of that form. We begin by setting  $\tilde{\psi} = \psi$  and then iteratively improve  $\tilde{\psi}$  to reduce (12). Although we can see several strategies for improving  $\tilde{\psi}$ , for concreteness, we will restrict our description to the strategy most similar to Ref. 6. To improve the approximation  $\tilde{\psi}$ , we loop through the variables (electrons). The functions in variables other than the current variable are fixed, and the functions in the current variable are modified to minimize the overall error (12). The error (12) depends linearly on the functions in a single variable, so the minimization becomes much easier. This general alternating least-squares approach is well known (see, e.g., Refs. 30, 38, 39, 11, 15, and 59). Although one could loop through the variables multiple times to try to minimize (12), it appears to be more cost effective to loop only once and then do the next Green's function iteration. We alternate through the directions, but for ease of exposition, we describe the  $k=1$  case. So,  $\tilde{\phi}_k^l$  is fixed for  $k > 1$ , and we will solve for the values of  $\tilde{\phi}_1^l$  for all  $l$ .

To refine in the current variable, we set up and solve a linear least-squares problem. The normal equations for a least-squares problem are derived by taking a gradient with respect to the free parameters and setting the result equal to zero. As long as the approximating function is linear and not degenerate in these parameters, the resulting equations are linear and have a unique solution, which minimizes the error with respect to these parameters. Usually, these free parameters are coefficients of the representation in some fixed basis. For example, to find the coefficients  $\{c_i\}$  to minimize

$$\left\| f - \sum_i c_i g_i \right\|^2 = \left\langle f - \sum_i c_i g_i, f - \sum_i c_i g_i \right\rangle, \quad (14)$$

construct the normal equations

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \quad (15)$$

with

$$A(k, i) = \langle g_k, g_i \rangle \text{ and } b(k) = \langle g_k, f \rangle, \quad (16)$$

solve them, and set  $c_i = x(i)$ . Instead of using coefficients in some basis as our parameters, we take the parameters to be the point values of our functions  $\tilde{\phi}_1^l$ , so that the gradient becomes a variational derivative. Formally, we consider a basis of delta functions  $\delta(\gamma - \cdot)$  and let their coefficients be our parameters. We still obtain linear normal equations (15), but now  $\mathbf{b}$  and  $\mathbf{x}$  are vectors of functions, and  $\mathbb{A}$  is a matrix of integral operators. Specifically,  $b(l)$  is a function of  $\gamma$ ,  $x(l')$  is a function of  $\gamma'$ , and  $A(l, l')$  is an integral operator mapping functions of  $\gamma'$  to functions of  $\gamma$ . The kernels in  $\mathbb{A}$  are formally defined by

$$A(l, l')(\gamma, \gamma') = \tilde{s}_l \tilde{s}_{l'} \left\langle \delta(\gamma - \gamma_1) \prod_{i=2}^N \tilde{\phi}_i^l(\gamma_i), \delta(\gamma' - \gamma_1) \prod_{i=2}^N \tilde{\phi}_i^{l'}(\gamma_i) \right\rangle_{\mathcal{A}}, \quad (17)$$

and the functions in  $\mathbf{b}$  are defined by

$$b(l)(\gamma) = \tilde{s}_l \sum_m^r s_m \left\langle \delta(\gamma - \gamma_1) \prod_{i=2}^N \tilde{\phi}_i^l(\gamma_i), -\mathcal{G}_\mu[\mathcal{V} + \mathcal{W}] \prod_{i=1}^N \phi_i^m(\gamma_i) \right\rangle_{\mathcal{A}}. \quad (18)$$

Once we solve (15), we set  $\tilde{\phi}_1^l = x(l)$ . To enforce the normalization convention  $\|\tilde{\phi}_1^l\| = 1$ , we can divide  $\tilde{\phi}_1^l$  by its norm and incorporate the norm into  $\tilde{s}_l$ .

To solve the matrix-integral system (15), we need an iterative method for solving linear systems that uses only operations compatible with integral operators, such as matrix-vector products, vector scales and additions, and vector inner products. Typically, the matrix  $\mathbb{A}$  in normal equations is positive definite. Our operator  $\mathbb{A}$  is only semidefinite due to the null space in the antisymmetric pseudonorm. Fortunately,  $\mathbf{b}$  was computed with the same pseudonorm and has no component in the null space of  $\mathbb{A}$ , so we can still use methods for positive-definite matrices. Based on these considerations, we choose to use the conjugate gradient iterative method (see, e.g., Ref. 23) to solve (15). One initializes with  $\mathbf{r} = \mathbf{b} - \mathbb{A}\mathbf{x}$ ,  $\mathbf{v} = \mathbf{r}$ , and  $c = \langle \mathbf{r}, \mathbf{r} \rangle$ , and then the core of the method is the sequence of assignments  $\mathbf{z} \leftarrow \mathbb{A}\mathbf{v}$ ,  $t \leftarrow c / \langle \mathbf{v}, \mathbf{z} \rangle$ ,  $\mathbf{x} \leftarrow \mathbf{x} + t\mathbf{v}$ ,  $\mathbf{r} \leftarrow \mathbf{r} - t\mathbf{z}$ ,  $d \leftarrow \langle \mathbf{r}, \mathbf{r} \rangle$ ,  $\mathbf{v} \leftarrow \mathbf{r} + (d/c)\mathbf{v}$ , and  $c \leftarrow d$ , applied iteratively.

One advantage of using this iterative method with integral operators is that our algorithm does not rely on any particular basis. The representation of  $\mathbf{x}$  can naturally be adaptive in  $\gamma$ , for example, refining near the nuclei as indicated by the refinement in  $\mathbf{b}$ . We assume the availability of some adaptive, high-accuracy representation for single-electron functions, such as the polynomial multiwavelet representation demonstrated in Refs. 28 and 29, which effectively eliminates the basis-set error. For the estimates of computational cost, we use  $M$  to denote the cost to represent a function of  $\gamma$  or integrate such a function. The antisymmetry constraint requires  $N \leq M$ , and in general, we expect  $M$  to be much larger than  $N$ .

### 3. Summary of the remainder of the paper

The core of the paper is the development of the methods needed to construct  $\mathbb{A}$  in (17) and  $\mathbf{b}$  in (18). First, in Sec. III, we develop the machinery and algorithms for computing antisymmetric inner products involving the operators  $\mathcal{T}$ ,  $\mathcal{V}$ , and  $\mathcal{W}$ . Our formulation uses low-rank perturbations of matrices, thus avoiding cofactor expansions. We also avoid explicit construction of  $\mathcal{W}$  by incorporating its effect via spatial convolutions with the Poisson kernel in three dimensions. Second, in Sec. IV, we show how to compute antisymmetric inner products involving these operators and the delta function  $\delta(\gamma - \gamma_1)$ . Again, the key is to use low-rank perturbations of matrices.

In Sec. V, we assemble all our tools to demonstrate how to perform our main algorithm and, in particular, how to construct  $\mathbb{A}$  in (17) and  $\mathbf{b}$  in (18). We also gather the computational cost for the whole method. This cost depends on the following parameters:

- $N$ , the number of electrons.

- $r$ , the separation rank used in (5).
- $M$ , the cost to represent a function of  $\gamma$ , as discussed in Sec. II B 2.
- $M_p$ , the cost to convolve a function of  $\gamma$  with the Poisson kernel  $1/\|\mathbf{r}\|$ . A Fourier-based Poisson solver on a uniform grid would achieve  $M \log M$ . For adaptive methods such as that we used,  $M_p$  may be considered proportional to  $M \log M$  as well, but since the complexity depends on the type of function to which the operator is applied, the estimate of the cost is complicated (see the discussion in Refs. 5 and 19).
- $I$ , the number of Green's function iterations, as discussed in Sec. II B 1. If we used more than one alternating least-squares pass (Sec. II B 2) per iteration, then  $I$  would include a factor of the number of passes.
- $S$ , the number of conjugate gradient iterations used to solve the system in Sec. II B 2. Although  $S$  in theory could be as many as the number of degrees of freedom  $rM$ , we generally have a very good starting point and so expect only a very small number to be needed.
- $L$ , the number of terms used to approximate the Green's function to relative error  $\epsilon$  with Gaussians. In Sec. V A, we prove that  $L = \mathcal{O}((\ln \epsilon)^2)$  independent of  $\mu$  and  $N$ .

In terms of these parameters, the cost to store the representation (5) is

$$\mathcal{O}(rNM) \quad (19)$$

and the computational cost to perform the algorithm is

$$\mathcal{O}(Ir^2N^2[L(N + M_p) + S(N + M)]). \quad (20)$$

For comparison, the cost to evaluate a single antisymmetric inner product via Löwdin's rules is  $\mathcal{O}(N^2(N+M))$ .

### C. Further considerations

We have implemented the method developed here and tested it sufficiently to verify the correctness of the algorithm as presented. The numerical results are too preliminary to allow us to make any particular claims at this point, however, so we will present them separately. The linear algebra accelerations based on Appendix B have not yet been implemented.

We develop the method in terms of the total variable  $\gamma$  without specifying the spin states. If a specific spin state is imposed on our initial trial wavefunction  $\psi_0$ , the iteration will preserve this state.

The representation (5) does not account for the interelectron cusp (see, e.g., Refs. 57, 47, 37, 50, 51, 36, and 27), and thus we cannot hope to achieve a small error  $\epsilon$  in the wavefunction with small  $r$ . As with CI methods, we may still be able to achieve a small error in the energy difference of two systems, which is often the quantity of interest in chemistry. For the current work, we fix  $r$  and adapt  $\phi_i'(\gamma_i)$  and  $s_i$  to minimize the error  $\epsilon$  rather than fix  $\epsilon$  and adaptively determine  $r$ . We are developing an extension to (5) that incorporates the cusp and hope to achieve a small error  $\epsilon$  through it.

Similarly, (5) is not size consistent/extensive. For example, if one applied it to a long line of identical, noninteracting subsystems, then  $r$  is expected to grow exponentially in the number of subsystems. We are developing a hierarchical extension to (5) suitable for such extended systems and hope to achieve linear scaling through it.

Although we have focused on the multiparticle Schrödinger equation, the tools that we have developed are another step toward general-purpose, automatically adaptive methods for solving high-dimensional problems.



### III. ANTISYMMETRIC INNER PRODUCTS

In this section, we develop methods for computing antisymmetric inner products involving  $\mathcal{W}$ ,  $\mathcal{V}$ , and  $\mathcal{T}$ . For this purpose, after setting notation, we develop methods for computing with low-rank perturbations of matrices, review the antisymmetry constraint, and define a notion of maximum coincidence. With these tools we then derive the main formulas.

#### A. Notation

We denote a column *vector* with suppressed indices by  $\mathbf{F}$  and with explicit indices by  $F(i)$ . We denote its conjugate transpose by  $\mathbf{F}^*$ . We use  $\mathbf{e}_i$  to denote the column vector that is 1 in coordinate  $i$  and zero otherwise. A linear *operator* is written  $\mathcal{L}$ . We denote a *matrix* with suppressed indices by  $\mathbf{L}$  and with explicit indices by  $L(i,j)$ . Recalling that  $\mathbf{r}=(x,y,z) \in \mathbf{R}^3$ , we combine spatial integration with summation over spins and define the integral

$$\int f(\gamma)d\gamma = \sum_{\sigma \in \{-1/2, 1/2\}} \int f(\mathbf{r}, \sigma) d\mathbf{r}. \quad (21)$$

We define the action of the single-electron kinetic and nuclear potential operators by

$$(\mathcal{T}_* + \mathcal{V}_*)[f](\gamma) = \left(-\frac{1}{2}\Delta + v(\mathbf{r})\right)f(\gamma) = \left(-\frac{1}{2}\Delta + v(\mathbf{r})\right)f(\mathbf{r}, \sigma). \quad (22)$$

In what follows, we will reduce the action of the interelectron potential operator  $\mathcal{W}$  to convolutions with the Poisson kernel, so we define

$$\mathcal{W}_p[f](\mathbf{r}) = \int \frac{1}{\|\mathbf{r} - \mathbf{r}'\|} f(\gamma') d\gamma' = \sum_{\sigma' \in \{-1/2, 1/2\}} \int \frac{1}{\|\mathbf{r} - \mathbf{r}'\|} f(\mathbf{r}', \sigma') d\mathbf{r}'. \quad (23)$$

We allow these operators to be applied componentwise to vectors and matrices of functions.

Next, we define  $\tilde{\Phi} = \prod_{i=1}^N \phi_i(\gamma_i)$ , so, for example, we can write  $\langle \tilde{\Phi}, \Phi \rangle_{\mathcal{A}}$  instead of  $\langle \prod_{i=1}^N \tilde{\phi}_i(\gamma_i), \prod_{i=1}^N \phi_i(\gamma_i) \rangle_{\mathcal{A}}$ . We also associate with the product  $\Phi$  a vector of  $N$  functions of a single variable,

$$\Phi = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_N \end{bmatrix}. \quad (24)$$

We can then, for example, construct a new vector of functions  $\Theta$  by applying a matrix to an old one, as in  $\Theta = \mathbf{L}^{-1}\tilde{\Phi}$ . Although we do linear algebra operations on these vectors, we note that  $\Phi + \tilde{\Phi}$  does not correspond to  $\Phi + \tilde{\Phi}$ , so there is not a true vector-space structure. Our formulas contain fairly complicated expressions with such vectors, such as  $\int \Phi^* \mathcal{W}_p[\Theta \Phi^*] \Theta d\gamma$ . To parse this expression, we note that  $\Theta$  is a column vector of functions and  $\Phi^*$  is a row vector of functions, so  $\Theta \Phi^*$  is a matrix of functions. Then,  $\mathcal{W}_p[\Theta \Phi^*]$  is still a matrix of functions but applying  $\Phi^*$  on the left and  $\Theta$  on the right yields a single function, which is integrated in the implied variable  $\gamma$  to yield a number. When explicit specification of the variable involved is needed, the notation  $\Phi(\gamma)$  indicates that the single variable  $\gamma$  is used in all the functions.

#### B. Determinants of low-rank perturbations of matrices

Since the antisymmetric inner product involves determinants, we will use some linear algebra relations for them. Proposition 3 in this section is used heavily and is the key to avoiding rather unpleasant cofactor expansions.

*Proposition 2 (determinant via Schur complement):* Let  $\mathbf{A}$  be a nonsingular square matrix,  $\mathbf{D}$  a square matrix, and  $\mathbf{B}$  and  $\mathbf{C}$  matrices of appropriate size. Then,

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = |A| |D - CA^{-1}B|. \quad (25)$$

*Proof:* (See, e.g., Ref. 52) It is easy to verify directly that

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I & 0 \\ CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & D - CA^{-1}B \end{bmatrix} \begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix}. \quad (26)$$

Since the determinants of the first and third matrices are equal to 1, the determinant of the middle matrix gives the desired result.  $\square$

*Proposition 3 (determinant of a perturbation of the identity):* Let  $\{\mathbf{u}_q\}_{q=1}^Q$  and  $\{\mathbf{v}_q\}_{q=1}^Q$  be two sets of vectors of the same length, and  $\mathbf{u}_q \mathbf{v}_q^*$  denote the outer product of  $\mathbf{u}_q$  and  $\mathbf{v}_q$ . Then,

$$\left| I + \sum_{q=1}^Q \mathbf{u}_q \mathbf{v}_q^* \right| = \begin{vmatrix} 1 + \mathbf{v}_1^* \mathbf{u}_1 & \mathbf{v}_1^* \mathbf{u}_2 & \cdots & \mathbf{v}_1^* \mathbf{u}_Q \\ \mathbf{v}_2^* \mathbf{u}_1 & 1 + \mathbf{v}_2^* \mathbf{u}_2 & \cdots & \mathbf{v}_2^* \mathbf{u}_Q \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{v}_Q^* \mathbf{u}_1 & \mathbf{v}_Q^* \mathbf{u}_2 & \cdots & 1 + \mathbf{v}_Q^* \mathbf{u}_Q \end{vmatrix}. \quad (27)$$

*Proof:* Let  $U$  be the matrix with the vectors  $\{\mathbf{u}_q\}$  as its columns and  $V$  the matrix with the vectors  $\{\mathbf{v}_q\}$  as its columns. Note that  $U$  and  $V$  are of the same size. By Proposition 2, we have

$$\begin{vmatrix} I & U \\ -V^* & I \end{vmatrix} = |I + V^*U|, \quad (28)$$

which evaluates to the right side of (27). Exchanging the roles of  $A$  and  $D$  in Proposition 2, we have

$$\begin{vmatrix} I & U \\ -V^* & I \end{vmatrix} = |I + UV^*|, \quad (29)$$

which evaluates to the left side of (27).  $\square$

The  $Q=1$  case is well known (see, e.g., Ref. 52) but we have not found the general case in the literature.

### C. The modified pseudoinverse

The singular value decomposition (SVD) (e.g., Ref. 23) of a  $N \times N$  matrix is

$$A = \sum_{i=1}^N s_i \mathbf{u}_i \mathbf{v}_i^* = USV^*, \quad (30)$$

where the matrices  $U$  and  $V$  are unitary and the singular values  $\{s_i\}$  are non-negative and in descending order. The left singular vectors  $\{\mathbf{u}_i\}$  form an orthonormal set, as do the right singular vectors  $\{\mathbf{v}_i\}$ . The pseudoinverse is defined as

$$A^\dagger = \sum_{i=1}^{N-Q} s_i^{-1} \mathbf{v}_i \mathbf{u}_i^*, \quad (31)$$

where  $Q$  is the dimension of the (numerical) null space. We also define a projection matrix onto the null space.

*Definition 4:*

$$\mathbb{A}^\perp = \sum_{i=N-Q+1}^N \mathbf{v}_i \mathbf{u}_i^*. \quad (32)$$

For a modified pseudoinverse, we have the following.

*Definition 5 (modified pseudoinverse):*

$$\mathbb{A}^\ddagger = \mathbb{A}^\dagger + \mathbb{A}^\perp. \quad (33)$$

Note that  $\mathbb{A}^\perp$ , and thus  $\mathbb{A}^\ddagger$ , is not uniquely defined since the choice of basis for the null space is not unique. For our purposes, any consistent choice works. The modified pseudoinverse behaves much like the pseudoinverse but always has a nonzero determinant,

$$|\mathbb{A}^\ddagger| = \left( |\mathbb{U}| |\mathbb{V}^*| \prod_{s_i \neq 0} s_i \right)^{-1} \neq 0. \quad (34)$$

#### D. The antisymmetrizer and Löwdin's rule

Given a separable function, its antisymmetric projection can be found by applying the *antisymmetrizer*  $\mathcal{A}$  (see, e.g., Ref. 49), also called the *skew symmetrization* or *alternation* (see, e.g., Refs. 46 and 52), resulting in a Slater determinant. In the vector notation (24), we have

$$\mathcal{A}\Phi = \frac{1}{N!} |[\Phi(\gamma_1) \cdots \Phi(\gamma_N)]| = \frac{1}{N!} \begin{vmatrix} \phi_1(\gamma_1) & \phi_1(\gamma_2) & \cdots & \phi_1(\gamma_N) \\ \phi_2(\gamma_1) & \phi_2(\gamma_2) & \cdots & \phi_2(\gamma_N) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_N(\gamma_1) & \phi_N(\gamma_2) & \cdots & \phi_N(\gamma_N) \end{vmatrix}. \quad (35)$$

One cannot explicitly form a Slater determinant  $\mathcal{A}\Phi$  for large  $N$  since it would have  $N!$  terms. However, one can compute the antisymmetric pseudo-inner-product

$$\langle \tilde{\Phi}, \Phi \rangle_{\mathcal{A}} \stackrel{\text{def}}{=} \langle \mathcal{A}\tilde{\Phi}, \mathcal{A}\Phi \rangle = \langle \tilde{\Phi}, \mathcal{A}\Phi \rangle = \langle \mathcal{A}\tilde{\Phi}, \Phi \rangle, \quad (36)$$

where the first equality is a definition and the others follow since  $\mathcal{A}$  is an orthogonal projector. It is not a true inner product because it has a null space. To compute (36), first construct the matrix  $\mathbb{L}$  with entries

$$L(i, j) = \langle \tilde{\phi}_i, \phi_j \rangle \quad (37)$$

at cost  $\mathcal{O}(N^2M)$ . Then use  $\langle \tilde{\Phi}, \Phi \rangle_{\mathcal{A}} = \langle \mathcal{A}\tilde{\Phi}, \Phi \rangle$  and move the integrals inside the determinant to obtain

$$\langle \tilde{\Phi}, \Phi \rangle_{\mathcal{A}} = \frac{1}{N!} |\mathbb{L}|, \quad (38)$$

which is the so-called Löwdin's rule (e.g., Refs. 41 and 49). Since  $\mathbb{L}$  is an ordinary matrix, its determinant can be computed with cost  $\mathcal{O}(N^3)$  (or less). The denominator  $N!$  need never be computed, since it will occur in every term in our equations and so cancels.

Our method for enforcing the antisymmetry constraint, as described in Ref. 6, is to use the pseudonorm based on the antisymmetric inner product  $\langle \cdot, \cdot \rangle_{\mathcal{A}}$  for the least-squares fitting (12).

#### E. Maximum coincidence

Consider two products,  $\Phi = \prod_{i=1}^N \phi_i(\gamma_i)$  and  $\tilde{\Phi} = \prod_{i=1}^N \tilde{\phi}_i(\gamma_i)$ , stored in the vector notation of (24) as  $\Phi$  and  $\tilde{\Phi}$ . To specify which functions were used to compute  $\mathbb{L}$  in (37), we use the notation  $\mathbb{L}(\tilde{\Phi}, \Phi)$ . The matrix of inner products  $\mathbb{L} = \mathbb{L}(\tilde{\Phi}, \Phi)$  is in general full. Defining

$$\Theta = L^{-1}\tilde{\Phi}, \quad (39)$$

we have

$$\mathcal{A}\Theta = \frac{1}{N!} [|(L^{-1}\tilde{\Phi})(\gamma_1) \cdots (L^{-1}\tilde{\Phi})(\gamma_N)|] = |L^{-1}| \frac{1}{N!} [|\tilde{\Phi}(\gamma_1) \cdots \tilde{\Phi}(\gamma_N)|] = |L^{-1}| \mathcal{A}\tilde{\Phi}. \quad (40)$$

Thus, the antisymmetrizations of  $\tilde{\Phi}$  and  $\Theta$  are the same up to a constant, and we can use  $\Theta$  instead of  $\tilde{\Phi}$  in calculations. The advantage of using  $\Theta$  is that the resulting matrix of inner products  $\hat{L} = L(\Theta, \Phi) = I$ ; in other words, we have the biorthogonality property  $\langle \theta_i, \phi_j \rangle = \delta_{ij}$ . To show this, write the matrix  $\hat{L}$  as  $\int \Theta \Phi^* d\gamma$ , where the integration is elementwise. Substituting for  $\Theta$ , we have  $\int (L^{-1}\tilde{\Phi}) \Phi^* d\gamma$ . Since the integration is elementwise, it commutes with  $L^{-1}$  and we have  $L^{-1} \int \tilde{\Phi} \Phi^* d\gamma = L^{-1} L = I$ . The computational cost to construct  $\Theta$  is  $\mathcal{O}(N^2(N+M))$ .

When the matrix  $L$  in (37) is singular, we define  $\Theta = L^{\ddagger} \tilde{\Phi}$  using the modified pseudoinverse of Definition 5. By the same argument as before, we have  $|L^{\ddagger}|^{-1} \mathcal{A}\Theta = \mathcal{A}\tilde{\Phi}$ . The matrix  $\int \Theta \Phi^* d\gamma$  evaluates to  $L^{\ddagger} L = I - \sum_{i=N-Q+1}^N \mathbf{v}_i \mathbf{v}_i^*$ . For notational convenience in later sections, we will reindex our singular values and vectors so that the first  $Q$  generates the null space rather than the last  $Q$ .

*Remark 6:* Within CI methods, the functions in  $\Phi$  and  $\tilde{\Phi}$  are taken from a master set of orthonormal functions, and  $\Theta$  is simply a signed permutation of  $\tilde{\Phi}$  so that  $\phi_j = \theta_j$  for as many  $j$  as possible. This is known as the “maximum coincidence” ordering. The construction we use generalizes this notion.

## F. Antisymmetric inner product with the electron-electron potential $\mathcal{W}$ present

In this section, we derive formulas for computing antisymmetric inner products that include the electron-electron interaction potential. Although the derivation is somewhat messy, the resulting formulas are rather clean, and we use them verbatim in the computations. The main ideas are given in this section and then reused in later sections for other cases.

*Proposition 7:* When  $L$  from (37) is nonsingular,

$$\langle \tilde{\Phi}, \mathcal{W}\Phi \rangle_{\mathcal{A}} \stackrel{\text{def}}{=} \left\langle \mathcal{A} \prod_{j=1}^N \tilde{\phi}_j(\gamma_j), \left( \frac{1}{2} \sum_{i \neq j} \frac{1}{\|\mathbf{r}_i - \mathbf{r}_j\|} \right) \prod_{j=1}^N \phi_j(\gamma_j) \right\rangle \quad (41)$$

is equal to

$$\frac{1}{2} \frac{|L|}{N!} \int \Phi^* \Theta \mathcal{W}_p[\Phi^* \Theta] - \Phi^* \mathcal{W}_p[\Theta \Phi^*] \Theta d\gamma, \quad (42)$$

where  $\Theta = L^{-1}\tilde{\Phi}$ .

*Proof:* Using the maximum coincidence procedure in Sec. III E, (41) is equal to  $|L| \langle \Theta, \mathcal{W}\Phi \rangle_{\mathcal{A}}$ . We reorganize and find that we must compute

$$\frac{1}{2} \frac{|L|}{N!} \int \left( \sum_{i \neq j} \frac{1}{\|\mathbf{r}_i - \mathbf{r}_j\|} \right) \prod_{j=1}^N \tilde{\phi}_j(\gamma_j) \begin{vmatrix} \theta_1(\gamma_1) & \theta_1(\gamma_2) & \cdots & \theta_1(\gamma_N) \\ \theta_2(\gamma_1) & \theta_2(\gamma_2) & \cdots & \theta_2(\gamma_N) \\ \vdots & \vdots & \ddots & \vdots \\ \theta_N(\gamma_1) & \theta_N(\gamma_2) & \cdots & \theta_N(\gamma_N) \end{vmatrix} d\gamma_1 \cdots d\gamma_N. \quad (43)$$

By moving the sum outside of the integral, we can integrate in all directions except  $\gamma_i$  and  $\gamma_j$ . Using  $\langle \theta_m, \phi_n \rangle = \delta_{mn}$ , we obtain

$$\frac{1}{2 N!} \sum_{i \neq j} \int \frac{1}{\|\mathbf{r}-\mathbf{r}'\|} \begin{vmatrix} 1 & \cdots & \bar{\phi}_i(\gamma) \theta_1(\gamma) & \cdots & \bar{\phi}_j(\gamma') \theta_1(\gamma') & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & \bar{\phi}_i(\gamma) \theta_i(\gamma) & \cdots & \bar{\phi}_j(\gamma') \theta_i(\gamma') & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & \bar{\phi}_i(\gamma) \theta_j(\gamma) & \cdots & \bar{\phi}_j(\gamma') \theta_j(\gamma') & \cdots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & \bar{\phi}_i(\gamma) \theta_N(\gamma) & \cdots & \bar{\phi}_j(\gamma') \theta_N(\gamma') & \cdots & 1 \end{vmatrix} d\gamma d\gamma'$$

$$= \frac{1}{2 N!} \sum_{i \neq j} \int \frac{1}{\|\mathbf{r}-\mathbf{r}'\|} \left[ \mathbb{I} + (\bar{\phi}_i(\gamma) \Theta(\gamma) - \mathbf{e}_i) \mathbf{e}_i^* + (\bar{\phi}_j(\gamma') \Theta(\gamma') - \mathbf{e}_j) \mathbf{e}_j^* \right] d\gamma d\gamma'. \quad (44)$$

Since the inner matrix is a low-rank perturbation of the identity, we reduce its determinant using Proposition 3 and obtain

$$\frac{1}{2 N!} \sum_{i \neq j} \int \frac{1}{\|\mathbf{r}-\mathbf{r}'\|} \bar{\phi}_i(\gamma) \bar{\phi}_j(\gamma') \begin{vmatrix} \theta_i(\gamma) & \theta_i(\gamma') \\ \theta_j(\gamma) & \theta_j(\gamma') \end{vmatrix} d\gamma d\gamma'. \quad (45)$$

The determinant is zero if  $j=i$ , so we do not need to explicitly prohibit it as we needed to in (44) and above. The antisymmetrization has caused a convenient cancellation of a fictitious self-interaction, and thus allowed us to decouple the two sums. Expanding out the determinant and rearranging the terms, we obtain

$$\begin{aligned} & \frac{1}{2 N!} \int \left( \sum_i \bar{\phi}_i(\gamma) \theta_i(\gamma) \right) \left[ \int \frac{1}{\|\mathbf{r}-\mathbf{r}'\|} \left( \sum_j \bar{\phi}_j(\gamma') \theta_j(\gamma') \right) d\gamma' \right] d\gamma \\ & - \frac{1}{2 N!} \int \sum_i \sum_j \bar{\phi}_i(\gamma) \theta_j(\gamma) \left[ \int \frac{1}{\|\mathbf{r}-\mathbf{r}'\|} \bar{\phi}_j(\gamma') \theta_i(\gamma') d\gamma' \right] d\gamma. \end{aligned} \quad (46)$$

In our compact notation, this yields (42). □

We now consider the computational cost of (42). In the first term in (42), computing  $\Phi^* \Theta$  costs  $\mathcal{O}(NM)$ , applying  $\mathcal{W}_p[\cdot]$  to it costs  $\mathcal{O}(M_p)$ , and the integral in  $\gamma$  costs  $\mathcal{O}(M)$ . In the second term,  $\Phi \Theta^*$  costs  $\mathcal{O}(N^2M)$ , applying  $\mathcal{W}_p[\cdot]$  to it costs  $\mathcal{O}(N^2M_p)$ , applying  $\Theta^*$  and then  $\Phi$  costs  $\mathcal{O}(N^2M)$ , and then the integral in  $\gamma$  costs  $\mathcal{O}(M)$ . Including the cost to construct  $\Theta$ , our total cost is  $\mathcal{O}(N^2(N+M_p))$ .

**1. The singular case**

In this section, we investigate the case when the matrix  $L$  from (37) is singular. Inserting the definition  $\Theta = L^{-1} \tilde{\Phi}$  into our main formula (42), we have

$$\frac{1}{2 N!} \int \Phi^* L^{-1} \tilde{\Phi} \mathcal{W}_p[\Phi^* L^{-1} \tilde{\Phi}] - \Phi^* \mathcal{W}_p[L^{-1} \tilde{\Phi} \Phi^*] L^{-1} \tilde{\Phi} d\gamma. \quad (47)$$

In terms of the SVD (30), we can express

$$L^{-1} = \sum_{j=1}^N s_j^{-1} \mathbf{v}_j \mathbf{u}_j^* \text{ and } |L| = |\mathbf{U}| |\mathbf{V}^*| \prod_i s_i. \quad (48)$$

Inserting these expressions into (47), we have

$$\begin{aligned}
& \frac{1}{2} \frac{|U||V^*| \prod_i s_i}{N!} \int \Phi^* \sum_{j=1}^N s_j^{-1} \mathbf{v}_j \mathbf{u}_j^* \tilde{\Phi} \mathcal{W}_P \left[ \Phi^* \sum_{k=1}^N s_k^{-1} \mathbf{v}_k \mathbf{u}_k^* \tilde{\Phi} \right] \\
& - \Phi^* \mathcal{W}_P \left[ \sum_{j=1}^N s_j^{-1} \mathbf{v}_j \mathbf{u}_j^* \tilde{\Phi} \Phi^* \right] \sum_{k=1}^N s_k^{-1} \mathbf{v}_k \mathbf{u}_k^* \tilde{\Phi} d\gamma \\
& = \frac{1}{2} \frac{|U||V^*|}{N!} \sum_{j=1}^N \sum_{k=1}^N \prod_{i \neq j,k} s_i \int \Phi^* \mathbf{v}_j \mathbf{u}_j^* \tilde{\Phi} \mathcal{W}_P [\Phi^* \mathbf{v}_k \mathbf{u}_k^* \tilde{\Phi}] - \Phi^* \mathbf{v}_j \mathcal{W}_P [\mathbf{u}_j^* \tilde{\Phi} \Phi^* \mathbf{v}_k] \mathbf{u}_k^* \tilde{\Phi} d\gamma.
\end{aligned} \tag{49}$$

If  $\mathbb{L}$  is singular, then at least one  $s_i$  is zero, and only terms that exclude those from the product in (49) are nonzero. Since we exclude two indices in the product, if more than two  $s_i$  are zero, then the entire inner product is zero. If exactly two are zero, then only one term in the sum survives. If exactly one is zero, then we can simplify from a double to a single sum using symmetry. Recalling the modified pseudoinverse from Definition 5 and sorting the zero  $s_i$  to the beginning for notational convenience, we obtain the following propositions.

*Proposition 8:* When the rank deficiency of  $\mathbb{L}$  is more than 2, the antisymmetric inner product (41) evaluates to zero.

*Proposition 9:* When the rank deficiency of  $\mathbb{L}$  is equal to 2, the antisymmetric inner product (41) is equal to

$$\frac{1}{|\mathbb{L}^\ddagger|N!} \int \Phi^* \mathbf{v}_1 \mathbf{u}_1^* \tilde{\Phi} \mathcal{W}_P [\Phi^* \mathbf{v}_2 \mathbf{u}_2^* \tilde{\Phi}] - \Phi^* \mathbf{v}_1 \mathcal{W}_P [\Phi^* \mathbf{v}_2 \mathbf{u}_1^* \tilde{\Phi}] \mathbf{u}_2^* \tilde{\Phi} d\gamma. \tag{50}$$

*Proposition 10:* When the rank deficiency of  $\mathbb{L}$  is equal to 1, defining  $\Theta = \mathbb{L}^\dagger \tilde{\Phi}$  or  $\Theta = \mathbb{L}^\ddagger \tilde{\Phi}$ , the antisymmetric inner product (41) is equal to

$$\frac{1}{|\mathbb{L}^\ddagger|N!} \int \Phi^* \mathbf{v}_1 \mathbf{u}_1^* \tilde{\Phi} \mathcal{W}_P [\Phi^* \Theta] - \Phi^* \mathbf{v}_1 \mathcal{W}_P [\mathbf{u}_1^* \tilde{\Phi} \Phi^*] \Theta d\gamma. \tag{51}$$

In computing (50), constructing  $\Phi^* \mathbf{v}_1$ ,  $\Phi^* \mathbf{v}_2$ ,  $\mathbf{u}_1^* \tilde{\Phi}$ , and  $\mathbf{u}_2^* \tilde{\Phi}$  costs  $\mathcal{O}(NM)$ , applying  $\mathcal{W}_P[\cdot]$  costs  $\mathcal{O}(M_P)$  and, finally, the integral in  $\gamma$  costs  $\mathcal{O}(M)$ . In computing (51), the first term costs  $\mathcal{O}(NM)$  to form  $\Phi^* \Theta$ ,  $\mathcal{O}(M_P)$  to apply  $\mathcal{W}_P[\cdot]$ , and  $\mathcal{O}(M)$  to integrate in  $\gamma$ . The second term costs  $\mathcal{O}(NM)$  to form  $\mathbf{u}_1^* \tilde{\Phi} \Phi^*$ ,  $\mathcal{O}(NM_P)$  to apply  $\mathcal{W}_P[\cdot]$ ,  $\mathcal{O}(NM)$  to apply  $\Theta$ , and  $\mathcal{O}(M)$  to integrate in  $\gamma$ . In total, the computational cost for the singular cases is less than the cost of the nonsingular case.

*Remark 11:* In the CI context, rank deficiency 2 corresponds to a double excitation. The vectors  $\mathbf{u}_i$  and  $\mathbf{v}_i$  would be zero except for a single entry, and so select the locations of the excited electrons out of  $\Phi$  and  $\tilde{\Phi}$ . Proposition 9 then reduces to the Slater–Condon rules.<sup>14</sup>

### G. Antisymmetric inner product with $\mathcal{T}$ and/or $\mathcal{V}$ present

Since  $\mathcal{T}$  and  $\mathcal{V}$  both have the structure of a sum of one-directional operators, we state the formulas for their sum, although of course they can be treated individually.

*Proposition 12:* If  $\mathbb{L}$  from (37) is nonsingular,

$$\langle \tilde{\Phi}, (\mathcal{T} + \mathcal{V}) \Phi \rangle_{\mathcal{A}} \stackrel{\text{def}}{=} \left\langle \mathcal{A} \prod_{j=1}^N \tilde{\phi}_j(\gamma_j), \left( \sum_i -\frac{1}{2} \Delta_i + v(\mathbf{r}_i) \right) \prod_{j=1}^N \phi_j(\gamma_j) \right\rangle \tag{52}$$

is equal to

$$\frac{|\mathbb{L}|}{N!} \int (\mathcal{T}_* + \mathcal{V}_*) [\Phi]^* \Theta d\gamma. \tag{53}$$

*Proof:* We follow the same procedure as we used for the electron-electron operator  $\mathcal{W}$  in Sec. III F. Instead of (44), we have the simpler expression

$$\frac{|\mathbb{L}|}{N!} \sum_i \int |1 + ((\mathcal{T}_* + \mathcal{V}_*)[\bar{\phi}_i](\gamma)\Theta(\gamma) - \mathbf{e}_i)\mathbf{e}_i^*| d\gamma. \tag{54}$$

Applying Proposition 3, we obtain (53). □

To analyze the computational cost to compute (53), we note that it costs  $\mathcal{O}(NM)$  to apply  $(\mathcal{T}_* + \mathcal{V}_*)[\cdot]$ . Including the cost for the maximum coincidence transformation, our total cost is thus  $\mathcal{O}(N^2(N+M))$ .

**1. The Singular Case**

We now state the formula when  $\mathbb{L}$  is singular. The analysis is similar to that for  $\mathcal{W}$  in Sec. III F 1.

*Proposition 13:* *If the rank deficiency of  $\mathbb{L}$  is greater than 1, (52) evaluates to zero. If it is equal to 1, we have*

$$\frac{1}{|\mathbb{L}^\ddagger|N!} \int (\mathcal{T}_* + \mathcal{V}_*)[\Phi^*\mathbf{v}_1]\mathbf{u}_1^*\tilde{\Phi}d\gamma. \tag{55}$$

To compute (55), it costs  $\mathcal{O}(NM)$  to form  $\Phi^*\mathbf{v}_1$  and  $\mathbf{u}_1^*\tilde{\Phi}$ , and  $\mathcal{O}(M)$  to apply  $(\mathcal{T}_* + \mathcal{V}_*)[\cdot]$ .

**IV. INCORPORATING DELTA FUNCTIONS INTO THE ANTISYMMETRIC INNER PRODUCTS**

In this section, we show how to compute antisymmetric inner products when one of the component functions is replaced by a delta function. For concreteness, we will replace  $\tilde{\phi}_1(\gamma_1)$  by  $\delta(\gamma - \gamma_1)$ .

**A. Löwdin’s rule with  $\delta(\gamma - \gamma_1)$  present**

The matrix  $\mathbb{L}$  from (37) is defined by  $L(i, j) = \langle \tilde{\phi}_i, \phi_j \rangle$ . If we replace  $\tilde{\phi}_1(\gamma_1)$  by  $\delta(\gamma - \gamma_1)$ , then the first row depends on  $\gamma$  and is given by  $L(1, j) = \langle \delta(\gamma - \cdot), \phi_j \rangle = \phi_j(\gamma)$ . We thus have a matrix that depends on  $\gamma$ ,

$$\mathbb{L}(\gamma) = \begin{bmatrix} \phi_1(\gamma) & \phi_2(\gamma) & \cdots & \phi_N(\gamma) \\ \langle \tilde{\phi}_2, \phi_1 \rangle & \langle \tilde{\phi}_2, \phi_2 \rangle & \cdots & \langle \tilde{\phi}_2, \phi_N \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \tilde{\phi}_N, \phi_1 \rangle & \langle \tilde{\phi}_N, \phi_2 \rangle & \cdots & \langle \tilde{\phi}_N, \phi_N \rangle \end{bmatrix}. \tag{56}$$

To compute with  $\mathbb{L}(\gamma)$  without resorting to cofactor expansions, we express  $\mathbb{L}(\gamma)$  as a rank-1 perturbation of a matrix of numbers. Define

$$\mathbb{E} = \begin{bmatrix} \bar{d}(1) & \bar{d}(2) & \cdots & \bar{d}(N) \\ \langle \tilde{\phi}_2, \phi_1 \rangle & \langle \tilde{\phi}_2, \phi_2 \rangle & \cdots & \langle \tilde{\phi}_2, \phi_N \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \tilde{\phi}_N, \phi_1 \rangle & \langle \tilde{\phi}_N, \phi_2 \rangle & \cdots & \langle \tilde{\phi}_N, \phi_N \rangle \end{bmatrix}, \tag{57}$$

where the vector  $\mathbf{d}^*$  is chosen to be a unit vector orthogonal to the remaining rows of  $\mathbb{E}$ . This choice assures that the rank deficiency of  $\mathbb{E}$  will be smaller than or equal to the rank deficiency of the matrix with any other first row. It also gives us some convenient properties, namely,  $\mathbb{E}\mathbf{d} = \mathbf{e}_1$ ,  $\mathbf{d}^*\mathbb{E}^\ddagger = \mathbf{e}_1^*$ ,  $\mathbb{E}^\ddagger\mathbf{e}_1 = \mathbf{d}$ , and  $\mathbf{e}_1^*\mathbb{E} = \mathbf{d}^*$ , where  $\mathbb{E}^\ddagger$  is the modified pseudoinverse of Definition 5. It costs  $\mathcal{O}(N^2M)$  to construct  $\mathbb{E}$  and  $\mathcal{O}(N^3)$  to compute  $\mathbb{E}^\ddagger$  and  $|\mathbb{E}|$ .

We then have

$$\mathbb{L}(\gamma) = \mathbb{E} + \mathbf{e}_1(\Phi(\gamma) - \mathbf{d})^* \quad (58)$$

and, with the help of Proposition 3, compute

$$|\mathbb{L}(\gamma)| = |\mathbb{E}| |\mathbb{I} + \mathbf{d}(\Phi(\gamma) - \mathbf{d})^*| = |\mathbb{E}| (1 + (\Phi(\gamma) - \mathbf{d})^* \mathbf{d}) = |\mathbb{E}| \Phi(\gamma)^* \mathbf{d}, \quad (59)$$

which yields the following

*Proposition 14:*

$$\left\langle \delta(\gamma - \gamma_1) \prod_{i=2}^N \tilde{\phi}_i(\gamma_i), \prod_{i=1}^N \phi_i(\gamma_i) \right\rangle_{\mathcal{A}} = |\mathbb{E}| \Phi(\gamma)^* \mathbf{d}, \quad (60)$$

where  $\mathbb{E}$  and  $\mathbf{d}$  are defined as above.

*Remark 15:* If  $i > 1$ , then,

$$\langle |\mathbb{E}| \Phi^* \mathbf{d}, \tilde{\phi}_i \rangle = |\mathbb{E}| \langle \Phi, \tilde{\phi}_i \rangle^* \mathbf{d} = |\mathbb{E}| E(i, \cdot)^* \mathbf{d} = 0, \quad (61)$$

since  $\mathbf{d}$  is orthogonal to  $E(i, \cdot)$ , which is row number  $i$  of  $\mathbb{E}$ . Thus, the function (60) is orthogonal to  $\tilde{\phi}_i$  for  $i > 1$ . The same property will hold when the operators  $\mathcal{T}$ ,  $\mathcal{V}$ , and  $\mathcal{W}$  are present in the antisymmetric inner product, as described in the following sections.

### B. Antisymmetric inner product with $\delta(\gamma - \gamma_1)$ and ( $\mathcal{T}$ and/or $\mathcal{V}$ ) present

To compute antisymmetric inner products involving operators, we will modify formulas from Sec. III. The first (trivial) modification is to denote the variable of integration in those formulas by  $\gamma'$ , so as not to confuse it with the variable  $\gamma$  in  $\delta(\gamma - \gamma_1)$ . Next, we replace  $|\mathbb{L}|$  with  $|\mathbb{L}(\gamma)|$  given by (59). Using (58), we can express

$$\mathbb{L}(\gamma)^{-1} = (\mathbb{E} + \mathbf{e}_1(\Phi(\gamma) - \mathbf{d})^*)^{-1} = (\mathbb{E}(\mathbb{I} + \mathbf{d}(\Phi(\gamma) - \mathbf{d})^*))^{-1} = (\mathbb{I} + \mathbf{d}(\Phi(\gamma) - \mathbf{d})^*)^{-1} \mathbb{E}^{-1}. \quad (62)$$

Using the Sherman–Morrisson formula [see, e.g., Ref. 23 and (B5) in Appendix B], we then have

$$\mathbb{L}(\gamma)^{-1} = \left( \mathbb{I} - \frac{\mathbf{d}(\Phi(\gamma) - \mathbf{d})^*}{1 + (\Phi(\gamma) - \mathbf{d})^* \mathbf{d}} \right) \mathbb{E}^{-1} = \left( \mathbb{I} + \mathbf{d} \frac{(\mathbf{d} - \Phi(\gamma))^*}{\Phi(\gamma)^* \mathbf{d}} \right) \mathbb{E}^{-1}. \quad (63)$$

The vector of functions  $\Theta$ , which was defined by  $\mathbb{L}^{-1} \tilde{\Phi}$ , now depends on the variable  $\gamma$  in  $\delta(\gamma - \gamma_1)$  as well as its own internal variable  $\gamma'$ . Replacing  $\mathbb{L}^{-1}$  with (63) and  $\tilde{\Phi}$  with  $\tilde{\Phi}(\gamma') + \mathbf{e}_1(\delta(\gamma - \gamma') - \tilde{\phi}_1(\gamma'))$ , we obtain

$$\Theta(\gamma, \gamma') = \left( \mathbb{I} + \mathbf{d} \frac{(\mathbf{d} - \Phi(\gamma))^*}{\Phi(\gamma)^* \mathbf{d}} \right) \mathbb{E}^{-1} (\tilde{\Phi}(\gamma') + \mathbf{e}_1(\delta(\gamma - \gamma') - \tilde{\phi}_1(\gamma'))). \quad (64)$$

To compute it, we first compute the base case  $\tilde{\Theta}(\gamma') = \mathbb{E}^{-1} \tilde{\Phi}(\gamma')$ . Multiplying out (64) and noting  $\mathbf{d}^* \tilde{\Theta} = \mathbf{d}^* \mathbb{E} \tilde{\Phi} = \tilde{\phi}_1$ , we obtain

$$\begin{aligned} \Theta(\gamma, \gamma') &= \tilde{\Theta}(\gamma') + \mathbf{d} \frac{\mathbf{d}^* \tilde{\Theta}(\gamma') - \Phi(\gamma)^* \tilde{\Theta}(\gamma') + \delta(\gamma - \gamma') - \tilde{\phi}_1(\gamma')}{\Phi(\gamma)^* \mathbf{d}} \\ &= \tilde{\Theta}(\gamma') - \mathbf{d} \frac{\Phi(\gamma)^* \tilde{\Theta}(\gamma') - \delta(\gamma - \gamma')}{\Phi(\gamma)^* \mathbf{d}}. \end{aligned} \quad (65)$$

We are now ready to state our main formulas.

*Proposition 16:* When  $\mathbb{E}$  is nonsingular,



$$\left\langle \delta(\gamma - \gamma_1) \prod_{i=2}^N \tilde{\phi}_i(\gamma_i), (\mathcal{T} + \mathcal{V}) \prod_{i=1}^N \phi_i(\gamma_i) \right\rangle_{\mathcal{A}} \tag{66}$$

is equal to

$$\frac{|\mathbb{E}|}{N!} \left[ \Phi(\gamma)^* \left( \mathbf{d} \int (\mathcal{T}_* + \mathcal{V}_*) [\Phi]^* \tilde{\Theta} d\gamma' - \int (\mathcal{T}_* + \mathcal{V}_*) [\Phi^* \mathbf{d}] \tilde{\Theta} d\gamma' \right) + (\mathcal{T}_* + \mathcal{V}_*) [\Phi^* \mathbf{d}] (\gamma) \right], \tag{67}$$

which can be computed with total cost  $\mathcal{O}(N^3 + N^2M)$ .

*Proof:* To compute (66), we start with  $(|\mathbb{L}|/N!) \int (\mathcal{T}_* + \mathcal{V}_*) [\Phi]^* \Theta d\gamma'$  from (53) and substitute in (59) and (65) to obtain

$$\frac{|\mathbb{E}| \Phi(\gamma)^* \mathbf{d}}{N!} \int (\mathcal{T}_* + \mathcal{V}_*) [\Phi](\gamma')^* \left( \tilde{\Theta}(\gamma') - \mathbf{d} \frac{\Phi(\gamma)^* \tilde{\Theta}(\gamma') - \delta(\gamma - \gamma')}{\Phi(\gamma)^* \mathbf{d}} \right) d\gamma'. \tag{68}$$

Distributing out and rearranging, we have

$$\begin{aligned} & \frac{|\mathbb{E}|}{N!} \int \Phi(\gamma)^* \mathbf{d} (\mathcal{T}_* + \mathcal{V}_*) [\Phi]^*(\gamma') \tilde{\Theta}(\gamma') - (\mathcal{T}_* + \mathcal{V}_*) [\Phi](\gamma')^* \mathbf{d} \Phi(\gamma)^* \tilde{\Theta}(\gamma') \\ & + (\mathcal{T}_* + \mathcal{V}_*) [\Phi](\gamma')^* \mathbf{d} \delta(\gamma - \gamma') d\gamma', \end{aligned} \tag{69}$$

which yields (67). Although in (63) and (65) we divide by  $\Phi^* \mathbf{d}$ , which could be zero, this denominator cancels in the final expression, so we can argue by continuity that the final expression is still valid. One can also prove this directly by determining the null space of  $\mathbb{L}$  and then using (55).  $\square$

*Remark 17:* It is the term with pointwise multiplication,  $(\mathcal{T}_* + \mathcal{V}_*) [\Phi^* \mathbf{d}]$  in (67), that allows adaptive refinement around the nuclei in the numerical algorithm.

To obtain the formulas when  $\mathbb{E}$  is singular, we follow the same logic as in Sec. III F 1. Denote the singular vectors in the null space of  $\mathbb{E}$  by  $\{(\tilde{\mathbf{u}}_i, \tilde{\mathbf{v}}_i)\}$ .

*Proposition 18:* When  $\mathbb{E}$  has rank deficiency greater than 1, (66) is zero. When  $\mathbb{E}$  has rank deficiency 1, (66) is equal to

$$\frac{1}{|\mathbb{E}^\ddagger| N!} \Phi(\gamma)^* \left( \mathbf{d} \int (\mathcal{T}_* + \mathcal{V}_*) [\Phi^* \tilde{\mathbf{v}}_1] \tilde{\mathbf{u}}_1^* \tilde{\Phi} d\gamma' - \tilde{\mathbf{v}}_1 \int (\mathcal{T}_* + \mathcal{V}_*) [\Phi^* \mathbf{d}] \tilde{\mathbf{u}}_1^* \tilde{\Phi} d\gamma' \right), \tag{70}$$

which can be computed with total cost  $\mathcal{O}(N^3 + N^2M)$ .

### C. Antisymmetric inner product with $\delta(\gamma - \gamma_1)$ and $\mathcal{W}$ present

Conceptually, the derivation if  $\mathcal{W}$  is present in the inner product is the same and we obtain the following propositions.

*Proposition 19:* When  $\mathbb{E}$  is nonsingular,

$$\left\langle \delta(\gamma - \gamma_1) \prod_{i=2}^N \tilde{\phi}_i(\gamma_i), \mathcal{W} \prod_{i=1}^N \phi_i(\gamma_i) \right\rangle_{\mathcal{A}} \tag{71}$$

is equal to

$$\frac{1}{2N!} \mathbb{E} \left[ 2(\Phi(\gamma)^* \mathbf{d} \mathcal{W}_p[\Phi^* \tilde{\Theta}] (\gamma) - \Phi(\gamma)^* \mathcal{W}_p[\tilde{\Theta} \Phi^* \mathbf{d}] (\gamma)) + \Phi(\gamma)^* \left( \mathbf{d} \int \Phi^* \tilde{\Theta} \mathcal{W}_p[\Phi^* \tilde{\Theta}] - \Phi^* \mathcal{W}_p[\tilde{\Theta} \Phi^*] \tilde{\Theta} d\gamma' - 2 \int \tilde{\Theta} \mathcal{W}_p[\Phi^* \tilde{\Theta}] \Phi^* \mathbf{d} - \tilde{\Theta} \Phi^* \mathcal{W}_p[\tilde{\Theta} \Phi^* \mathbf{d}] d\gamma' \right) \right], \quad (72)$$

which can be computed with total cost  $\mathcal{O}(N^3 + N^2 M_p)$ .

*Proposition 20:* When  $\mathbb{E}$  has rank deficiency of 1, (71) is equal to

$$\frac{1}{|\mathbb{E}^\ddagger|N!} \left[ (\Phi(\gamma)^* \mathbf{d} \mathcal{W}_p[\Phi^* \tilde{\mathbf{v}}_1 \tilde{\mathbf{u}}_1^* \tilde{\Phi}] (\gamma) - \Phi(\gamma)^* \tilde{\mathbf{v}}_1 \mathcal{W}_p[\tilde{\mathbf{u}}_1^* \tilde{\Phi} \Phi^* \mathbf{d}] (\gamma)) + \Phi(\gamma)^* \left( \mathbf{d} \int \Phi^* \tilde{\mathbf{v}}_1 (\tilde{\mathbf{u}}_1^* \tilde{\Phi} \mathcal{W}_p[\Phi^* \tilde{\Theta}] - \mathcal{W}_p[\tilde{\mathbf{u}}_1^* \tilde{\Phi} \Phi^*] \tilde{\Theta}) d\gamma' + \int \tilde{\Theta} (\Phi^* \tilde{\mathbf{v}}_1 \mathcal{W}_p[\tilde{\mathbf{u}}_1^* \tilde{\Phi} \Phi^* \mathbf{d}] - \mathcal{W}_p[\Phi^* \tilde{\mathbf{v}}_1 \tilde{\mathbf{u}}_1^* \tilde{\Phi}] \Phi^* \mathbf{d}) d\gamma' - \tilde{\mathbf{v}}_1 \int \Phi^* \mathbf{d} (\tilde{\mathbf{u}}_1^* \tilde{\Phi} \mathcal{W}_p[\Phi^* \tilde{\Theta}] - \mathcal{W}_p[\tilde{\mathbf{u}}_1^* \tilde{\Phi} \Phi^*] \tilde{\Theta}) d\gamma' \right) \right], \quad (73)$$

which can be computed with total cost  $\mathcal{O}(N^3 + N^2 M + NM_p)$ .

*Proposition 21:* When  $\mathbb{E}$  has rank deficiency of 2, (71) is equal to

$$\frac{1}{|\mathbb{E}^\ddagger|N!} \Phi(\gamma)^* \left[ \mathbf{d} \int \Phi^* \tilde{\mathbf{v}}_1 \tilde{\mathbf{u}}_1^* \tilde{\Phi} \mathcal{W}_p[\Phi^* \tilde{\mathbf{v}}_2 \tilde{\mathbf{u}}_2^* \tilde{\Phi}] - \Phi^* \tilde{\mathbf{v}}_2 \mathcal{W}_p[\tilde{\mathbf{u}}_2^* \tilde{\Phi} \Phi^* \tilde{\mathbf{v}}_1] \tilde{\mathbf{u}}_1^* \tilde{\Phi} d\gamma - \tilde{\mathbf{v}}_1 \int \Phi^* \tilde{\mathbf{v}}_2 \tilde{\mathbf{u}}_2^* \tilde{\Phi} \mathcal{W}_p[\Phi^* \mathbf{d} \tilde{\mathbf{u}}_1^* \tilde{\Phi}] - \Phi^* \tilde{\mathbf{v}}_2 \mathcal{W}_p[\tilde{\mathbf{u}}_2^* \tilde{\Phi} \Phi^* \mathbf{d}] \tilde{\mathbf{u}}_1^* \tilde{\Phi} d\gamma - \tilde{\mathbf{v}}_2 \int \Phi^* \tilde{\mathbf{v}}_1 \tilde{\mathbf{u}}_1^* \tilde{\Phi} \mathcal{W}_p[\Phi^* \mathbf{d} \tilde{\mathbf{u}}_2^* \tilde{\Phi}] - \Phi^* \tilde{\mathbf{v}}_1 \mathcal{W}_p[\tilde{\mathbf{u}}_1^* \tilde{\Phi} \Phi^* \mathbf{d}] \tilde{\mathbf{u}}_2^* \tilde{\Phi} d\gamma \right], \quad (74)$$

which can be computed with total cost  $\mathcal{O}(N^3 + NM + M_p)$ .

## V. DETAILS OF THE GREEN'S FUNCTION ITERATION

In this section, we fill in the missing pieces in the Green's function iteration algorithm outlined in Sec. II B. First, we give a representation for the Green's function itself. Then we use the methods in the previous sections to construct the vector  $\mathbf{b}$  in (18) and the matrix  $\mathbb{A}$  in (17) to form the normal equations (15). Next, we give the algorithm from Sec. II B in outline form as a pseudocode. Finally, we gather the computational cost of the whole method and present some linear algebra techniques to reduce it.

### A. Representing the Green's function

In this section, we construct a separated representation for the Green's function  $\mathcal{G}_\mu$  in (7) following the ideas in Refs. 7 and 6 (see also Refs. 25 and 26). We will use this representation in Sec. V B when constructing the right-hand side of the normal equations.

We begin by constructing an approximation of  $1/t$  with exponentials such that

$$\left| \frac{1}{t} - \sum_{p=1}^L w_p \exp(-\tau_p t) \right| < \epsilon \quad (75)$$

on the interval  $t \in [1, \infty)$ , with  $w_p$  and  $\tau_p$  positive. Expansions of  $1/t$  into exponentials have been used in several applications and constructed by diverse techniques, see, Refs. 9, 32, 60, 8, 10, and 24 and the references therein. The interval  $[1, \infty)$  is addressed specifically in Ref. 10, where it is shown that the error rate  $\epsilon = \mathcal{O}(\exp(-c\sqrt{L}))$  can be achieved, which means we can achieve  $L = \mathcal{O}((\ln \epsilon)^2)$ .

Substituting  $t=s/(-\mu)$  for  $\mu < 0$  into (75) and dividing by  $-\mu$ , one has

$$\left| \frac{1}{s} - \sum_{p=1}^L \frac{w_p}{-\mu} \exp\left(-\frac{\tau_p}{-\mu} s\right) \right| < \frac{\epsilon}{-\mu}, \tag{76}$$

valid on the interval  $s \in [-\mu, \infty)$ . In Fourier coordinates, we can express

$$\mathcal{G}_\mu = \frac{1}{2\pi^2 \sum \xi_i^2 - \mu}, \tag{77}$$

from which we see that  $\|\mathcal{G}_\mu\| = 1/(-\mu)$ . Since the denominator is at least  $-\mu > 0$ , we can substitute into (76) and obtain

$$\left| \mathcal{G}_\mu - \sum_{p=1}^L \frac{w_p}{-\mu} e^{-\tau_p \otimes} \exp\left(-\frac{2\pi^2 \tau_p}{-\mu} \xi_i^2\right) \right| < \frac{\epsilon}{-\mu} = \epsilon \|\mathcal{G}_\mu\|. \tag{78}$$

Thus, we obtain an approximation of  $\mathcal{G}_\mu$  with relative error  $\epsilon$  in norm using  $L$  terms, with  $L$  independent of  $N$  and  $\mu$ . To construct  $\mathcal{G}_\mu$  as an integral operator in spatial coordinates, we apply the inverse Fourier transform to obtain

$$\mathcal{G}_\mu \approx \sum_{p=1}^L \otimes_{i=1}^N \mathcal{F}_{\mathbf{r}_i}^p, \tag{79}$$

where the convolution operator  $\mathcal{F}_{\mathbf{r}_i}^p$ , which depends implicitly on  $\mu$ , is defined by

$$\begin{aligned} \mathcal{F}_{\mathbf{r}_i}^p f(\gamma_1, \dots, \gamma_N) &= \left(\frac{w_p}{-\mu e^{\tau_p}}\right)^{1/N} \left(\frac{-\mu}{2\pi\tau_p}\right)^{3/2} \\ &\times \int \exp\left(-\frac{-\mu}{2\tau_p} \|\mathbf{r}_i - \mathbf{r}'\|^2\right) f(\gamma_1, \dots, \gamma_{i-1}, (\mathbf{r}', \sigma_i), \gamma_{i+1}, \dots, \gamma_N) d\mathbf{r}'. \end{aligned} \tag{80}$$

This construction has a theoretical value since it has proved the following theorem.

*Theorem 22: For any  $\epsilon > 0$ ,  $\mu < 0$ , and  $N$ , the  $N$ -particle Green's function  $\mathcal{G}_\mu$  has a separated representation with relative error in operator norm bounded by  $\epsilon$  using  $L = \mathcal{O}((\ln \epsilon)^2)$  terms, with  $L$  independent of  $\mu$  and  $N$ .*

**B. Constructing the right-hand side vector  $\mathbf{b}$  in (18)**

In order to do a step in the iteration, we need to construct the right-hand side  $\mathbf{b}$  in the normal equations (15) in Sec. II B 2. Since  $\mathcal{A}$  is an orthogonal projection,  $\mathcal{A}$  and  $\mathcal{G}_\mu$  commute, and  $\mathcal{G}_\mu$  is self-adjoint, the entry (18) is equal to

$$b(l)(\gamma) = -\tilde{s}_l \sum_m^r s_m \left\langle \mathcal{A} \mathcal{G}_\mu \delta(\gamma - \gamma_1) \prod_{i=2}^N \tilde{\phi}_i^l(\gamma_i), [\mathcal{V} + \mathcal{W}] \prod_{i=1}^N \phi_i^m(\gamma_i) \right\rangle. \tag{81}$$

Substituting (79) in for  $\mathcal{G}_\mu$  and rearranging, we have

$$b(l)(\gamma) = -\tilde{s}_l \sum_m^r s_m \sum_{p=1}^L \left\langle \mathcal{A} \mathcal{F}_{\mathbf{r}_1}^p \delta(\gamma - \gamma_1) \prod_{i=2}^N \mathcal{F}_{\mathbf{r}_i}^p \tilde{\phi}_i^l(\gamma_i), [\mathcal{V} + \mathcal{W}] \prod_{i=1}^N \phi_i^m(\gamma_i) \right\rangle. \tag{82}$$

The computation is of the same form for each value of the indices  $l, m$ , and  $p$ , so we can consider a single term and suppress the indices.

To evaluate a single term  $\langle \mathcal{A} \mathcal{F}_{\mathbf{r}_1}^p \delta(\gamma - \gamma_1) \prod_{i=2}^N \mathcal{F}_{\mathbf{r}_i}^p \tilde{\phi}_i^l(\gamma_i), [\mathcal{V} + \mathcal{W}] \prod_{i=1}^N \phi_i^m(\gamma_i) \rangle$ , we use the formulas in Propositions 16–21 in Secs. IV B and IV C, with two modifications. The first modification is that  $\tilde{\Phi}$  is replaced with  $\mathcal{F}\tilde{\Phi}$  throughout. This replacement causes no structural change to the

formulas; it just changes the inputs. The second modification is caused by the replacement of  $\delta(\gamma - \gamma_1)$  by  $\mathcal{F}_r \delta(\gamma - \gamma_1)$ . The first row of  $\mathbb{L}(\gamma)$  in (56) becomes  $\mathcal{F}\Phi(\gamma)^*$ , which makes  $|\mathbb{L}(\gamma)| = |\mathbb{E}|\mathcal{F}\Phi(\gamma)^*\mathbf{d}$ . Similarly, (65) becomes

$$\Theta(\gamma, \gamma') = \tilde{\Theta}(\gamma') - \mathbf{d} \frac{\mathcal{F}\Phi(\gamma)^* \tilde{\Theta}(\gamma') - \mathcal{F}\delta(\gamma - \gamma')}{\mathcal{F}\Phi(\gamma)^* \mathbf{d}}. \quad (83)$$

Tracking  $\mathcal{F}$  through the formulas, we find that all we need to do is to modify the formulas in Secs. IV B and IV C by applying  $\mathcal{F}$  to the final result.

### C. Constructing the matrix $\mathbb{A}$ in (17)

In this section, we construct the kernels in (17) for the normal equations (15) using the same ideas in Sec. IV. We fix  $l$  and  $l'$  and define

$$K(\gamma, \gamma') = \frac{\mathcal{A}(l, l')(\gamma, \gamma')}{\tilde{s}_l \tilde{s}_{l'}}, \quad (84)$$

$$\mathbf{w}(\gamma') = [\tilde{\phi}_2^l(\gamma') \cdots \tilde{\phi}_N^l(\gamma')]^*, \quad (85)$$

$$\mathbf{y}(\gamma) = [\tilde{\phi}_2^{l'}(\gamma) \cdots \tilde{\phi}_N^{l'}(\gamma)]^*, \quad (86)$$

$$\mathbb{D} = \begin{bmatrix} \langle \tilde{\phi}_2^l, \tilde{\phi}_2^{l'} \rangle & \cdots & \langle \tilde{\phi}_2^l, \tilde{\phi}_N^{l'} \rangle \\ \vdots & \ddots & \vdots \\ \langle \tilde{\phi}_N^l, \tilde{\phi}_2^{l'} \rangle & \cdots & \langle \tilde{\phi}_N^l, \tilde{\phi}_N^{l'} \rangle \end{bmatrix}. \quad (87)$$

Using Löwdin's rules (38), we have

$$K(\gamma, \gamma') = \frac{|\mathbb{L}|}{N!} = \frac{1}{N!} \begin{vmatrix} \delta(\gamma - \gamma') & \mathbf{y}^*(\gamma) \\ \mathbf{w}(\gamma') & \mathbb{D} \end{vmatrix}. \quad (88)$$

Expressing  $\mathbb{L}$  as a low-rank perturbation of

$$\begin{bmatrix} 1 & 0 \\ 0 & \mathbb{D} \end{bmatrix},$$

we have

$$\begin{aligned} K(\gamma, \gamma') &= \frac{1}{N!} \left| \begin{bmatrix} 1 & 0 \\ 0 & \mathbb{D} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} [0 \ \mathbf{y}^*(\gamma)] + \begin{bmatrix} \delta(\gamma - \gamma') - 1 \\ \mathbf{w}(\gamma') \end{bmatrix} [1 \ 0] \right| \\ &= \frac{1}{N!} \left| \begin{bmatrix} 1 & 0 \\ 0 & \mathbb{D} \end{bmatrix} \left[ \mathbb{I} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} [0 \ \mathbf{y}^*(\gamma)] + \begin{bmatrix} \delta(\gamma - \gamma') - 1 \\ \mathbb{D}^{-1} \mathbf{w}(\gamma') \end{bmatrix} [1 \ 0] \right] \right| = \frac{|\mathbb{D}|}{N!} \left| \begin{bmatrix} 1 & \mathbf{y}^*(\gamma) \mathbb{D}^{-1} \mathbf{w}(\gamma') \\ 1 & \delta(\gamma - \gamma') \end{bmatrix} \right| \\ &= \frac{|\mathbb{D}|}{N!} (\delta(\gamma - \gamma') - \mathbf{y}^*(\gamma) \mathbb{D}^{-1} \mathbf{w}(\gamma')). \end{aligned} \quad (89)$$

If  $\mathbb{D}$  is singular, then we apply the same logic as in Sec. III F 1. If  $\mathbb{D}$  has rank deficiency greater than 1, then  $K(\gamma, \gamma') = 0$ . If it has rank deficiency of 1, then we have

$$\begin{aligned}
K(\gamma, \gamma') &= \frac{1}{|\mathbb{D}^\ddagger|N!} \left| \mathbb{I} + \begin{bmatrix} 0 \\ -\mathbf{v} \end{bmatrix} [0 \ \mathbf{v}^*] + \begin{bmatrix} 1 \\ 0 \end{bmatrix} [0 \ \mathbf{y}^*(\gamma)] + \begin{bmatrix} \delta(\gamma - \gamma') - 1 \\ \mathbb{D}^\ddagger \mathbf{w}(\gamma') \end{bmatrix} [1 \ 0] \right| \\
&= \frac{1}{|\mathbb{D}^\ddagger|N!} \left| \begin{array}{ccc} 0 & 0 & \mathbf{v}^* \mathbb{D}^\ddagger \mathbf{w}(\gamma') \\ -\mathbf{y}^*(\gamma) \mathbf{v} & 1 & \mathbf{y}^*(\gamma) \mathbb{D}^\ddagger \mathbf{w}(\gamma') \\ 0 & 1 & \delta(\gamma - \gamma') \end{array} \right| = \frac{-(\mathbf{y}^*(\gamma) \mathbf{v})(\mathbf{v}^* \mathbb{D}^\ddagger \mathbf{w}(\gamma'))}{|\mathbb{D}^\ddagger|N!} \\
&= \frac{-(\mathbf{y}^*(\gamma) \mathbf{v})(\mathbf{u}^* \mathbf{w}(\gamma'))}{|\mathbb{D}^\ddagger|N!}, \tag{90}
\end{aligned}$$

where  $\mathbb{D}^\ddagger$  is the modified pseudoinverse of Definition 5.

In the nonsingular case, we can construct  $\mathbb{D}$  at cost  $\mathcal{O}(N^2M)$  and compute  $\mathbb{D}^{-1}$  at cost  $\mathcal{O}(N^3)$ . Applying this kernel costs  $\mathcal{O}(NM)$  to integrate against a function in  $\gamma'$ ,  $\mathcal{O}(N^2)$  to apply  $\mathbb{D}^{-1}$ , and then  $\mathcal{O}(NM)$  to apply  $\mathbf{y}^*$  to the result. In the singular case, we can compute  $\mathbb{D}^\ddagger$  at cost  $\mathcal{O}(N^3)$  and construct  $\mathbf{y}^* \mathbf{v}$  and  $\mathbf{u}^* \mathbf{w}$  at cost  $\mathcal{O}(NM)$ . Since the variables separate, applying this kernel costs  $\mathcal{O}(M)$ .

*Remark 23:* In the case  $r=1$ , which corresponds to the HF formulation,  $\mathbb{D}=\mathbb{I}$  and  $K(\gamma, \gamma')$  is just the projector orthogonal to  $\{\tilde{\phi}_{ij=2}^N\}$ .

#### D. Pseudocode

In this section, we give the algorithm in outline form as a pseudocode. We do not indicate when objects can be recalled or updated from previous computations.

Loop through  $l$  Green's function iterations (9), (10), and (13). For each of these:

Construct  $\mathcal{G}_\mu$  as in Sec. V A, obtaining the operators  $\mathcal{F}^p$  in (80).

Loop through the  $N$  directions (electrons). For each of these:

Compute  $A(l, l')$  via (89) for all  $(l, l')$ .

Compute  $b(l)(\gamma)$  in (82) by:

Loop in the  $r$  values of  $l$  and for each:

Sum over the  $L$  values of  $p$  and for each:

Compute  $\mathcal{F}^p \phi_i^l$  for all  $i$ .

Sum over the  $r$  values of  $m$  and for each:

Using  $\mathcal{F}^p \tilde{\Phi}$  in place of  $\tilde{\Phi}$ , construct  $\mathbb{E}$  in (57).

Compute  $|\mathbb{E}|$  and  $\mathbb{E}^{-1}$ .

Construct  $\tilde{\Theta} = \mathbb{E}^{-1} \mathcal{F}^p \tilde{\Phi}$ .

Construct  $\Phi^* \tilde{\Theta}$ ,  $\Phi^* \mathbf{d}$ , and  $\tilde{\Theta} \Phi^*$ .

Compute  $\mathcal{W}_p[\Phi^* \tilde{\Theta}]$  and  $\mathcal{W}_p[\tilde{\Theta} \Phi^*]$ .

Compute (67) and (72) using these ingredients.

Apply  $\mathcal{F}^p$  to [(67) + (72)].

Apply conjugate gradient to solve the normal equations (15).

Renormalize as in (10).

Update  $\mu$  via (13).

*Remark 24:* We have presented the algorithm in serial form for clarity. The loop in  $l$ , sum in  $p$ , and sum in  $m$  can be trivially parallelized. Parallelizing the loop through the  $N$  electrons would represent a change in the algorithm, which we will develop elsewhere.

#### E. Overall computational cost

The computational cost is dominated by the repeated construction and solution of the normal equations (15). For a fixed direction, the construction cost is dominated by (82), which has  $r^2L$  inner products. The most costly portion of the inner products is (72), which requires  $\mathcal{O}(N^3 + N^2M_p)$  operations, giving us the net construction cost

$$\mathcal{O}(r^2LN^2(N + M_p)). \quad (91)$$

The operation count to solve the normal equations (15) by applying the matrix of integral operators  $A$   $S$  times is

$$\mathcal{O}(r^2SN(N + M)). \quad (92)$$

As we loop through the directions, we may reuse several quantities, so the total cost of the construction is less than  $N$  times the cost for one direction. In fact, the construction cost for the entire loop through  $N$  directions is of the same order as the cost for one direction. The application cost is simply multiplied by  $N$ . In the sections below, we show how to update the construction for direction  $k=2$  using what we already have for direction  $k=1$  and then determine the cost for one loop through the directions. We defer the development of the technical linear algebra rules on low-rank updates to Appendix B and here only show how to apply them to our problem. Our final conclusion is the computational cost

$$\mathcal{O}(Ir^2N^2[L(N + M_p) + S(N + M)]), \quad (93)$$

where  $I$  is the number of Green's function iterations.

### 1. Reuse in computing $A$

Let  $D_1$  denote  $D$  in (87) for directions 1 and  $D_2$  the version for direction 2. We let  $\hat{\phi}'_1$  denote the updated version of  $\tilde{\phi}'_1$ . To construct  $D_2$  requires only the first column and row of  $D_1$  to be updated, specifically

$$D_2 = D_1 + \mathbf{e}_1 [0 \langle \hat{\phi}'_1, \tilde{\phi}'_3 \rangle - \langle \tilde{\phi}'_2, \tilde{\phi}'_3 \rangle \cdots] + \begin{bmatrix} \langle \hat{\phi}'_1, \hat{\phi}'_1 \rangle - \langle \tilde{\phi}'_2, \tilde{\phi}'_2 \rangle \\ \vdots \end{bmatrix} \mathbf{e}_1^*. \quad (94)$$

Computing those inner products involving  $\hat{\phi}'_1$  and  $\tilde{\phi}'_1$  costs  $\mathcal{O}(NM)$ . Using Proposition 25 twice, we compute  $D_2^\ddagger$ ,  $|D_2^\ddagger|$ , and if appropriate  $\mathbf{v}$ , all at cost  $\mathcal{O}(N^2)$ . The formulas (88) and following are modified by inserting the extra column and row in the second place instead of the first, but otherwise, the procedure is unchanged. The cost for one loop through the  $N$  directions is thus  $\mathcal{O}(N^3 + N^2M)$ .

### 2. Reuse in computing antisymmetric inner products with $\delta(\gamma - \gamma_1)$ and operators

We again let  $\hat{\phi}'_1$  denote the updated version of  $\tilde{\phi}'_1$  computed during the solving of  $k=1$ . The inner products needed to construct  $E_2$  require only the one row involving  $\hat{\phi}'_1$  to be updated, at cost  $\mathcal{O}(NM)$ . The vector  $\mathbf{d}_1$  can be constructed by doing the SVD of  $E_1$  with the first row set to zero and then selecting one of the right singular vectors  $\mathbf{v}_i$  with zero singular value. Using Proposition 25, we obtain the SVD of  $E_2$  with first row set to zero and second row containing the new inner products, and thus can find  $\mathbf{d}_2$ . Putting the first and second rows back in proper position, we then have

$$E_2 = E_1 + \mathbf{e}_1 ([\langle \mathcal{F}\hat{\phi}'_1, \phi_1 \rangle \cdots \langle \mathcal{F}\hat{\phi}'_1, \phi_N \rangle] - \mathbf{d}_1^*) + \mathbf{e}_2 (\mathbf{d}_2^* - [\langle \mathcal{F}\tilde{\phi}'_2, \phi_1 \rangle \cdots \langle \mathcal{F}\tilde{\phi}'_2, \phi_N \rangle]), \quad (95)$$

and we can compute  $|E_2^\ddagger|$  and  $E_2^\ddagger$  using Proposition 25 twice, at cost  $\mathcal{O}(N^2)$ .

Proposition 25 produces a rank-2 update and we must apply it twice. For notational ease, we will show how to use a rank-1 update applied once; the method easily extends. Assuming  $E_2^\ddagger = E_1^\ddagger + \mathbf{f}\mathbf{g}^*$ , we next update

$$\tilde{\Theta}_2 = \mathbb{E}_2^\dagger \mathcal{F} \tilde{\Theta}_2 = (\mathbb{E}_1^\dagger + \mathbf{fg}^*)(\mathcal{F} \tilde{\Theta}_1 + \mathbf{e}_1(\hat{\phi}_1 - \tilde{\phi}_1)) = \tilde{\Theta}_1 + \mathbf{d}_1(\hat{\phi}_1 - \tilde{\phi}_1) + \mathbf{fg}^* \mathcal{F} \tilde{\Theta}_1 + \mathbf{fg}^* \mathbf{e}_1(\hat{\phi}_1 - \tilde{\phi}_1) \quad (96)$$

at cost  $\mathcal{O}(NM)$ . It is insufficient to just update  $\tilde{\Theta}_2$  in this way, since it would still cost  $\mathcal{O}(N^2 M_p)$  to compute  $\mathcal{W}_p[\tilde{\Theta}_2 \Phi^*]$  in (72). Instead, we update the combined quantity

$$\begin{aligned} \Phi^* \mathcal{W}_p[\tilde{\Theta}_2 \Phi^*] &= \Phi^* \mathcal{W}_p[\tilde{\Theta}_1 \Phi^*] + \Phi^* \mathbf{d}_1 \mathcal{W}_p[(\hat{\phi}_1 - \tilde{\phi}_1) \Phi^*] + \Phi^* \mathbf{f} \mathcal{W}_p[\mathbf{g}^* \mathcal{F} \tilde{\Theta}_1 \Phi^*] \\ &\quad + \Phi^* \mathbf{fg}^* \mathbf{e}_1 \mathcal{W}_p[(\hat{\phi}_1 - \tilde{\phi}_1) \Phi] \end{aligned} \quad (97)$$

at cost  $\mathcal{O}(NM_p)$ . With this quantity and  $\tilde{\Theta}_2$ , we can compute (72) at cost  $\mathcal{O}(NM_p)$ . The singular cases work similarly. The cost for one loop through the  $N$  directions is thus  $\mathcal{O}(N^2 M_p)$ .

## ACKNOWLEDGMENTS

We would like to thank Dr. Robert Harrison (University of Tennessee and Oak Ridge National Laboratory) and Dr. Lucas Monzón (University of Colorado) for many useful conversations. Portions of this research were conducted while G.B. and M.J.M. were in residence at the Institute for Pure and Applied Mathematics during the fall of 2004.

This material is based upon work supported by the National Science Foundation under Grant Nos. DMS-0219326 (G.B. and M.J.M.), DMS-0612358 (G.B. and F.P.), and DMS-0545895 (M.J.M.), the DARPA/ARO under Grant Nos. W911NF-04-1-0281 and W911NF-06-1-0254 (G.B. and M.J.M.), and the Department of Energy under Grant No. DE-FG02-03ER25583 and DOE/ORNL Grant No. 4000038129 (G.B. and F.P.).

## APPENDIX A: ALGORITHMS BASED ON GRADIENT DESCENT

We prefer the integral iteration in Sec. II B 1 due to the generally superior numerical properties of integral formulations. One could, however, try to minimize (4) directly with a method based on gradients. Since the machinery that we have constructed applies to these methods as well, we sketch how it can be used.

To minimize (4), we could use a gradient descent starting at some initial guess for  $\psi$ . Inserting our current approximation  $\psi$  and formally taking the gradient, we have

$$2 \frac{\langle \mathcal{H} \psi, \nabla \psi \rangle_A \langle \psi, \psi \rangle_A - \langle \mathcal{H} \psi, \psi \rangle_A \langle \psi, \nabla \psi \rangle_A}{\langle \psi, \psi \rangle_A^2}. \quad (A1)$$

Defining  $\mu$  to be our current value of (4), the gradient reduces to

$$\frac{2}{\langle \psi, \psi \rangle_A} (\langle \mathcal{H} \psi, \nabla \psi \rangle_A - \mu \langle \psi, \nabla \psi \rangle_A). \quad (A2)$$

The gradient is with respect to the parameters that are used to minimize (4). In our case, that is the values of the functions  $\phi_j^l$ . Taking the gradient with respect to the point values of  $\phi_j^l$  results in a vector  $\mathbf{g}$  of functions, defined by

$$g_j^l(\gamma) = \frac{2}{\langle \psi, \psi \rangle_A} s_l \sum_{m=1}^r s_m \left\langle \delta(\gamma - \gamma_j) \prod_{i \neq j}^N \phi_i^l(\gamma_i), (\mathcal{H} - \mu \mathcal{I}) \prod_{i=1}^N \phi_i^m(\gamma_i) \right\rangle_A, \quad (A3)$$

where  $\delta(\gamma - \gamma_j)$  is the delta function. The methods in Sec. IV can be used to construct  $\mathbf{g}$ .

Moving  $t$  in the direction opposite the gradient replaces  $\psi$  with

$$\sum_{l=1}^r s_l \prod_{i=1}^N (\phi_i^l - t g_i^l). \quad (\text{A4})$$

Some search procedure can then be used to find an appropriate  $t$ . Then  $\psi$  is updated and the procedure repeated.

Alternatively, we can use an alternating direction approach and take the gradient with respect to the functions  $\phi_i^l$  for one direction  $i$  while fixing the functions in the other directions and then loop through the directions. This loop through the directions is then repeated  $I$  times until we obtain the desired accuracy. We describe the  $i=1$  case. Moving  $t$  in the direction opposite the gradient replaces  $\psi$  with

$$\sum_{l=1}^r s_l (\phi_1^l - t g_1^l) \prod_{i=2}^N \phi_i^l = \psi - t \sum_{l=1}^r s_l g_1^l \prod_{i=2}^N \phi_i^l = \psi - t \tilde{\psi}. \quad (\text{A5})$$

Inserting (A5) into (4) results in

$$\frac{\langle \mathcal{H}(\psi - t \tilde{\psi}), \psi - t \tilde{\psi} \rangle_{\mathcal{A}}}{\langle \psi - t \tilde{\psi}, \psi - t \tilde{\psi} \rangle_{\mathcal{A}}} = \frac{\langle \mathcal{H}\psi, \psi \rangle_{\mathcal{A}} - 2t \langle \mathcal{H}\psi, \tilde{\psi} \rangle_{\mathcal{A}} + t^2 \langle \mathcal{H}\tilde{\psi}, \tilde{\psi} \rangle_{\mathcal{A}}}{\langle \psi, \psi \rangle_{\mathcal{A}} - 2t \langle \psi, \tilde{\psi} \rangle_{\mathcal{A}} + t^2 \langle \tilde{\psi}, \tilde{\psi} \rangle_{\mathcal{A}}}. \quad (\text{A6})$$

Once the inner products have been computed, we can find the minimizer for (A6) by solving a quadratic equation and then update  $\psi$  via (A5). The cost to construct  $\mathbf{g}$  for one direction is  $r^2$  times the cost for one inner product. The dominant cost for the inner product comes from (72), which costs  $\mathcal{O}(N^3 + N^2 M_p)$ , giving us the net construction cost

$$\mathcal{O}(r^2 N^2 (N + M_p)). \quad (\text{A7})$$

As described in Sec. V E 2, many of the computations can be reused, so the cost for a single loop through the  $N$  directions is of the same order. Thus, for  $I$  loops through the directions, the overall computational cost is

$$\mathcal{O}(I r^2 N^2 (N + M_p)). \quad (\text{A8})$$

## APPENDIX B: LOW-RANK UPDATES

In this section, we develop formulas for low-rank updates to  $\mathbb{A}^\dagger$ ,  $\mathbb{A}^\perp$ , and  $|\mathbb{A}^\ddagger|$  based on Refs. 43 and 4.

*Proposition 25:* Given  $\mathbb{A}$ ,  $\mathbb{A}^\dagger$ ,  $\mathbb{A}^\perp$ ,  $|\mathbb{A}^\ddagger|$ ,  $\mathbf{b}$ , and  $\mathbf{c}$ , let  $\mathbb{A}_1 = \mathbb{A} + \mathbf{b}\mathbf{c}^*$  and compute

$$\mathbf{d} = \mathbb{A}^\dagger \mathbf{b}, \quad \mathbf{e} = (\mathbb{A}^\dagger)^* \mathbf{c}, \quad \mathbf{f} = (\mathbb{I} - \mathbb{A}\mathbb{A}^\dagger) \mathbf{b}, \quad \mathbf{g} = (\mathbb{I} - \mathbb{A}^\dagger \mathbb{A}) \mathbf{c},$$

$$d = \mathbf{d}^* \mathbf{d}, \quad e = \mathbf{e}^* \mathbf{e}, \quad f = \mathbf{f}^* \mathbf{f}, \quad g = \mathbf{g}^* \mathbf{g},$$

$$\lambda = 1 + \mathbf{c}^* \mathbb{A}^\dagger \mathbf{b}, \quad \mu = |\lambda|^2 + dg, \quad \nu = |\lambda|^2 + ef,$$

$$\mathbf{p} = \bar{\lambda} \mathbf{d} + d \mathbf{g}, \quad \mathbf{q} = \lambda \mathbf{e} + e \mathbf{f}. \quad (\text{B1})$$

(1) If  $\lambda=0$ ,  $f=0$ , and  $g=0$ , then  $\text{rank}(\mathbb{A}_1) = \text{rank}(\mathbb{A}) - 1$  and

$$\mathbb{A}_1^\dagger = \mathbb{A}^\dagger - d^{-1} \mathbf{d} \mathbf{d}^* \mathbb{A}^\dagger + e^{-1} (-\mathbb{A}^\dagger \mathbf{e} + d^{-1} (\mathbf{d}^* \mathbb{A}^\dagger \mathbf{e}) \mathbf{d}) \mathbf{e}^*, \quad (\text{B2})$$

$$\mathbb{A}_1^\perp = \mathbb{A}^\perp + (1/\sqrt{de}) \mathbf{d} \mathbf{e}^*, \quad (\text{B3})$$



$$|A_1^\ddagger| = -(1/\sqrt{de})|A^\ddagger|. \quad (\text{B4})$$

(2) If  $\lambda \neq 0$ ,  $f=0$ , and  $g=0$ , then  $\text{rank}(A_1)=\text{rank}(A)$  and

$$A_1^\dagger = A^\dagger - \lambda^{-1} \mathbf{d} \mathbf{e}^*, \quad (\text{B5})$$

$$A_1^\perp = A^\perp, \quad (\text{B6})$$

$$|A_1^\ddagger| = |A^\ddagger| \lambda^{-1}. \quad (\text{B7})$$

(3) If  $f=0$  and  $g \neq 0$ , then  $\text{rank}(A_1)=\text{rank}(A)$  and

$$A_1^\dagger = A^\dagger - \mu^{-1} \mathbf{d} (\mathbf{g} \mathbf{d}^* A^\dagger + \bar{\lambda} \mathbf{e}^*) + \mu^{-1} \mathbf{g} (-\mathbf{d} \mathbf{e}^* + \lambda \mathbf{d}^* A^\dagger), \quad (\text{B8})$$

$$A_1^\perp = A^\perp - \frac{|\lambda|(\sqrt{\mu} - |\lambda|) \mathbf{g} + \lambda \mathbf{g} \mathbf{d}}{g|\lambda|\sqrt{\mu}} \mathbf{g}^* A^\perp, \quad (\text{B9})$$

$$|A_1^\ddagger| = |A^\ddagger| \frac{(\bar{\lambda} - \lambda)|\lambda|^2 + \lambda \mu}{\mu|\lambda|\sqrt{\mu}}. \quad (\text{B10})$$

(4) If  $f \neq 0$  and  $g=0$ , then  $\text{rank}(A_1)=\text{rank}(A)$  and

$$A_1^\dagger = A^\dagger - \nu^{-1} (f A^\dagger \mathbf{e} + \bar{\lambda} \mathbf{d}) \mathbf{e}^* + \nu^{-1} (-\mathbf{e} \mathbf{d} + \lambda A^\dagger \mathbf{e}) \mathbf{f}^*, \quad (\text{B11})$$

$$A_1^\perp = A^\perp - A^\perp \mathbf{f} \frac{(|\lambda|(\sqrt{\nu} - |\lambda|) \mathbf{f} + \bar{\lambda} \mathbf{f} \mathbf{e})^*}{f|\lambda|\sqrt{\nu}}, \quad (\text{B12})$$

$$|A_1^\ddagger| = |A^\ddagger| \frac{(\lambda - \bar{\lambda})|\lambda|^2 + \bar{\lambda} \nu}{\nu|\lambda|\sqrt{\nu}}. \quad (\text{B13})$$

(5) If  $f \neq 0$  and  $g \neq 0$ , then  $\text{rank}(A_1)=\text{rank}(A)+1$  and

$$A_1^\dagger = A^\dagger - f^{-1} \mathbf{d} \mathbf{f}^* + g^{-1} \mathbf{g} (-\mathbf{e}^* + \lambda f^{-1} \mathbf{f}^*), \quad (\text{B14})$$

$$A_1^\perp = A^\perp - (1/\sqrt{gf}) \mathbf{g} \mathbf{f}^*, \quad (\text{B15})$$

$$|A_1^\ddagger| = |A^\ddagger| [1 + (g^{-1} f^{-1} - (1/\sqrt{gf})) \mathbf{g}^* A^\perp \mathbf{f}], \quad (\text{B16})$$

The cost to compute  $A_1^\dagger$ ,  $A_1^\perp$ , and  $|A_1^\ddagger|$  is  $\mathcal{O}(N^2)$ .

*Proof:* The overall method, update rules for  $\text{rank}(A_1)$ , and update rules for  $A_1^\dagger$  are taken from Ref. 4, which also lists the useful properties

$$\mathbf{c}^* \mathbf{d} = \mathbf{e}^* \mathbf{b} = \lambda - 1, \quad \mathbf{b}^* \mathbf{f} = f, \quad \mathbf{c}^* \mathbf{g} = g, \quad \mathbf{d}^* \mathbf{g} = 0, \quad \mathbf{e}^* \mathbf{f} = 0,$$

$$A^\dagger A \mathbf{d} = \mathbf{d}, \quad A A^\dagger \mathbf{e} = \mathbf{e}, \quad A^* \mathbf{f} = A^\dagger \mathbf{f} = 0, \quad A \mathbf{g} = (A^\dagger)^* \mathbf{g} = 0. \quad (\text{B17})$$

They give update rules for the row and column spans of  $A_1$ , which we translate into update rules for  $A^\perp$ . The cases (B3), (B6), and (B15) follow directly. Corresponding to (B9), their update rule is that the row span of  $A^\perp$  should be extended (orthogonally) by  $\mathbf{d}$  and then reduced by projecting orthogonal to  $\mathbf{p}$ . We translate this into a (Householder) reflection of the vector  $\mathbf{g}$  into a vector in

the span of  $\mathbf{d}$  and  $\mathbf{g}$  perpendicular to  $\mathbf{p}$ . Adjusting these vectors to have equal norm and real inner product yields the reflection of the vector  $\bar{\lambda}\sqrt{\mu}\mathbf{g}$  to  $-|\lambda|(g\mathbf{d}-\bar{\lambda}\mathbf{g})$ , resulting in

$$\left( \mathbb{I} - \frac{2(\bar{\lambda}\sqrt{\mu}\mathbf{g} + |\lambda|(g\mathbf{d}-\bar{\lambda}\mathbf{g}))(\bar{\lambda}\sqrt{\mu}\mathbf{g} + |\lambda|(g\mathbf{d}-\bar{\lambda}\mathbf{g}))^*}{\|(\bar{\lambda}\sqrt{\mu}\mathbf{g} + |\lambda|(g\mathbf{d}-\bar{\lambda}\mathbf{g}))\|^2} \right)_{A^\perp}, \quad (\text{B18})$$

which simplifies to (B9). To obtain (B12), we use the same process, extending the column span by  $\mathbf{e}$  and then projecting orthogonal to  $\mathbf{q}$  by a reflection of  $\lambda\sqrt{f}\mathbf{f}$  to  $-|\lambda|(f\mathbf{e}-\lambda\mathbf{f})$ .

To derive the update rules for  $|A_1^\ddagger|$ , first add the update rules for  $A_1^\ddagger$  and  $A_1^\perp$  and then take the determinant. On the right-hand side, factor out a copy of  $A^\ddagger$  leaving a low-rank perturbation of the identity, to which we can apply Proposition 3. To simplify the results, we use (B1) and (B17), and the further observations

$$\begin{aligned} (A^\ddagger)^{-1}\mathbf{d} &= \mathbf{b} - \mathbf{f}, & (A^\ddagger)^{-1}A^\dagger\mathbf{e} &= \mathbf{c} - \mathbf{g}, & (A^\ddagger)^{-1}\mathbf{g} &= (A^\perp)^*\mathbf{c}, \\ \mathbf{e}^*(A^\ddagger)^{-1} &= \mathbf{c}^* - \mathbf{g}^*, & \mathbf{f}^*(A^\ddagger)^{-1} &= \mathbf{b}^*(A^\perp)^*. \end{aligned} \quad (\text{B19})$$

To obtain (B4), we compute

$$\begin{aligned} |A_1^\ddagger| &= |A^\ddagger| \left| \mathbb{I} - d^{-1}\mathbf{b}\mathbf{d}^*A^\dagger + ((1/\sqrt{de})\mathbf{b} - e^{-1}\mathbf{e} + d^{-1}e^{-1}(\mathbf{d}^*A^\dagger\mathbf{e})\mathbf{b})\mathbf{e}^* \right| \\ &= |A^\ddagger| \begin{vmatrix} 1 - d^{-1}\mathbf{d}^*A^\dagger\mathbf{b} & \mathbf{d}^*A^\dagger((1/\sqrt{de})\mathbf{b} - e^{-1}\mathbf{e} + d^{-1}e^{-1}(\mathbf{d}^*A^\dagger\mathbf{e})\mathbf{b}) \\ -d^{-1}\mathbf{e}^*\mathbf{b} & 1 + \mathbf{e}^*((1/\sqrt{de})\mathbf{b} - e^{-1}\mathbf{e} + d^{-1}e^{-1}(\mathbf{d}^*A^\dagger\mathbf{e})\mathbf{b}) \end{vmatrix} \\ &= |A^\ddagger| \begin{vmatrix} 0 & \mathbf{d}^*A^\dagger(1/\sqrt{de})\mathbf{b} \\ d^{-1} & \mathbf{e}^*((1/\sqrt{de})\mathbf{b} + d^{-1}e^{-1}(\mathbf{d}^*A^\dagger\mathbf{e})\mathbf{b}) \end{vmatrix} = |A^\ddagger|(-1/\sqrt{de}). \end{aligned} \quad (\text{B20})$$

For (B7), we have  $|A_1^\ddagger| = |A^\ddagger| |\mathbb{I} - \lambda^{-1}\mathbf{b}\mathbf{e}^*| = |A^\ddagger| (1 - \lambda^{-1}\mathbf{e}^*\mathbf{b}) = |A^\ddagger| \lambda^{-1}$ . To obtain (B10), we compute

$$\begin{aligned} |A^\ddagger| &\left| \mathbb{I} + (A^\ddagger)^{-1} \left( \mathbf{d} \left( -\mu^{-1}(g\mathbf{d}^*A^\dagger + \bar{\lambda}\mathbf{e}^*) - \frac{\lambda\mathbf{g}^*A^\perp}{|\lambda|\sqrt{\mu}} \right) \right. \right. \\ &\quad \left. \left. + \mathbf{g} \left( \mu^{-1}(-d\mathbf{e}^* + \lambda\mathbf{d}^*A^\dagger) - \frac{(\sqrt{\mu}-|\lambda|)\mathbf{g}^*A^\perp}{g\sqrt{\mu}} \right) \right) \right| \\ &= |A^\ddagger| \begin{vmatrix} 1 + (-\mu^{-1}(g\mathbf{d}^*A^\dagger + \bar{\lambda}\mathbf{e}^*))\mathbf{b} & (\mu^{-1}(-d\mathbf{e}^* + \lambda\mathbf{d}^*A^\dagger))^*\mathbf{b} \\ (-\lambda\mathbf{g}^*A^\perp/|\lambda|\sqrt{\mu})(A^\perp)^*\mathbf{c} & 1 - ((\sqrt{\mu}-|\lambda|)\mathbf{g}^*A^\perp/g\sqrt{\mu})^*(A^\perp)^*\mathbf{c} \end{vmatrix} \\ &= |A^\ddagger| \begin{vmatrix} \bar{\lambda}/\mu & d/\mu \\ -\lambda g/|\lambda|\sqrt{\mu} & |\lambda|\sqrt{\mu} \end{vmatrix} = |A^\ddagger| \frac{(\bar{\lambda}-\lambda)|\lambda|^2 + \lambda\mu}{\mu|\lambda|\sqrt{\mu}}. \end{aligned} \quad (\text{B21})$$

A similar calculation yields (B13). To obtain (B16), we compute

$$\begin{aligned} |A^\ddagger| &\left| \mathbb{I} + (A^\ddagger)^{-1}(-f^{-1}\mathbf{d}\mathbf{f}^* + \mathbf{g}(g^{-1}(-\mathbf{e}^* + \lambda f^{-1}\mathbf{f}^*) - (1/\sqrt{gf})\mathbf{f}^*)) \right| \\ &= |A^\ddagger| \begin{vmatrix} 1 & \mathbf{f}^*(A^\perp)^*\mathbf{c} \\ g^{-1}f^{-1}(\lambda-1) & 1 + (g^{-1}\lambda f^{-1} - (1/\sqrt{gf}))\mathbf{f}^*(A^\perp)^*\mathbf{c} \end{vmatrix} \\ &= |A^\ddagger| (1 + ((-1/\sqrt{gf}) + g^{-1}f^{-1})\mathbf{f}^*(A^\perp)^*\mathbf{c}) = |A^\ddagger| [1 + (g^{-1}f^{-1} - (1/\sqrt{gf}))\mathbf{g}^*A^\perp\mathbf{f}]. \end{aligned} \quad (\text{B22})$$

□

When  $\mathbb{A}$  and  $\mathbb{A}_1$  are nonsingular, (B5) is the Sherman–Morrison formula (see, e.g., Ref. 23). For our application, we need the singular vectors in  $\mathbb{A}^\perp$  rather than  $\mathbb{A}^\perp$  itself, but then only when  $\text{rank}(\mathbb{A}^\perp) \leq 3$ . These singular vectors can be extracted by a simple modification of the power method with deflation.

- <sup>1</sup> Agmon, S., *Schrödinger operators (Como, 1984)*, Lecture Notes in Mathematics, Vol. 1159 (Springer, Berlin, 1985), pp. 1–38.
- <sup>2</sup> Agren, H., Flores-Riveros, A., and Jensen, H. J. Aa, “Evaluation of first- and second-order nonadiabatic coupling elements from large multiconfigurational self-consistent-field wave functions,” *Phys. Rev. A* **34**, 4606 (1986).
- <sup>3</sup> Ayala, P. Y. and Schlegel, H. B., “A nonorthogonal CI treatment of symmetry breaking in sigma formylxyl radical,” *J. Chem. Phys.* **108**, 7560 (1998).
- <sup>4</sup> Baksalary, J. K., Baksalary, O. M., and Trenkler, G., “A revisit of formulae for the Moore-Penrose inverse of modified matrices,” *Linear Algebr. Appl.* **372**, 207 (2003).
- <sup>5</sup> Beylkin, G., Cheruvu, V., and Pérez, F., “Fast adaptive algorithms in the non-standard form for multidimensional problems,” *Appl. Comput. Harmon. Anal.* (in press).
- <sup>6</sup> Beylkin, G. and Mohlenkamp, M. J., “Algorithms for numerical analysis in high dimensions,” *SIAM J. Sci. Comput. (USA)* **26**, 2133 (2005).
- <sup>7</sup> Beylkin, G. and Mohlenkamp, M. J., “Numerical operator calculus in higher dimensions,” *Proc. Natl. Acad. Sci. U.S.A.* **99**, 10246 (2002).
- <sup>8</sup> Beylkin, G. and Monzón, L., “On approximation of functions by exponential sums,” *Appl. Comput. Harmon. Anal.* **19**, 17 (2005).
- <sup>9</sup> Braess, D., “Asymptotics for the approximation of wave functions by exponential sums,” *J. Approx. Theory* **83**, 93 (1995).
- <sup>10</sup> Braess, D. and Hackbusch, W., “Approximation of  $1/x$  by exponential sums in  $[1, \infty)$ ,” *IMA J. Numer. Anal.* **25**, 685 (2005).
- <sup>11</sup> Bro, R., Chemometrics Intelligence Laboratory System, Special Issue of Second Internet Conference in Chemometrics (INCINC’96) (unpublished), Vol. 38, pp. 149–171.
- <sup>12</sup> Cancès, E., Defranceschi, M., Kutzelnigg, W., Le Bris, C., and Maday, Y., *Handbook of Numerical Analysis* (North-Holland, Amsterdam, 2003), Vol. X, pp. 3–270.
- <sup>13</sup> Chen, B. and Anderson, J. R., “A simplified released-node quantum Monte Carlo calculation of the ground state of LiH,” *J. Chem. Phys.* **102**, 4491 (1995).
- <sup>14</sup> Condon, E. U. and Shortley, G. H., *The Theory of Atomic Spectra* (Cambridge University Press, Cambridge, 1967).
- <sup>15</sup> De Lathauwer, L., De Moor, B., and Vandewalle, J., “On the best rank-1 and rank- $(R_1, R_2, \dots, R_N)$  approximation of higher-order tensors,” *SIAM J. Matrix Anal. Appl.* **21**, 1324 (2000).
- <sup>16</sup> de Silva, V. and Lim, L.-H., “Tensor rank and the ill-posedness of the best low-rank approximation problem,” *Tensor Decompositions and Applications*, special Issue of SIAM J. Matrix Anal. Appl. (to be published).
- <sup>17</sup> Dijkstra, F. and Van Lenthe, J. H., “Gradients in valence bond theory,” *J. Chem. Phys.* **113**, 2100 (2000).
- <sup>18</sup> Dijkstra, F. and Van Lenthe, J. H., “On the rapid evaluation of cofactors in the calculation of nonorthogonal matrix elements,” *Int. J. Quantum Chem.* **67**, 77 (1998).
- <sup>19</sup> Ethridge, F. and Greengard, L., “A new fast-multipole accelerated Poisson solver in two dimensions,” *SIAM J. Sci. Comput. (USA)* **23**, 741 (2001).
- <sup>20</sup> Fink, R. and Staemmler, V., “A multi-configuration reference CEPA method based on pair natural orbitals,” *Theor. Chim. Acta* **87**, 129 (1993).
- <sup>21</sup> Gilbert, P., “The reconstruction of a three-dimensional structure from projections and its applications to electron microscopy II. Direct methods,” *Proc. R. Soc. London, Ser. B* **182**, 89 (1972).
- <sup>22</sup> Gilbert, T. L., “Multiconfiguration self-consistent-field theory for localized orbitals. I. The orbital equations,” *Phys. Rev. A* **6**, 580 (1972).
- <sup>23</sup> Golub, G. and Van Loan, C., *Matrix Computations* (Johns Hopkins University Press, Baltimore, 1996).
- <sup>24</sup> Hackbusch, W., “Entwicklungen nach exponentialsummen,” Max-Planck-Institut für Mathematik in den Naturwissenschaften, 2005 Technical Report No. 4 (see also [http://www.mis.mpg.de/scicomp/EXP\\_SUM/](http://www.mis.mpg.de/scicomp/EXP_SUM/)).
- <sup>25</sup> Hackbusch, W. and Khoromskij, B. N., “Low-rank Kronecker-product approximation to multi-dimensional nonlocal operators. I. Separable approximation of multi-variate functions,” *Computing* **76**, 177 (2006).
- <sup>26</sup> Hackbusch, W. and Khoromskij, B. N., “Low-rank Kronecker-product approximation to multi-dimensional nonlocal operators. II. HKT representation of certain operators,” *Computing* **76**, 203 (2006).
- <sup>27</sup> *Handbook of Numerical Analysis. Vol. X*, edited by Le Bris, C. (North-Holland, Amsterdam, 2003) (special volume on computational chemistry).
- <sup>28</sup> Harrison, R. J., Fann, G. I., Yanai, T., and Beylkin, G., *Lect. Notes Comput. Sci.* **2660**, 103 (2003).
- <sup>29</sup> Harrison, R. J., Fann, G. I., Yanai, T., Gan, Z., and Beylkin, G., “Multiresolution quantum chemistry: Basic theory and initial applications,” *J. Chem. Phys.* **121**, 11587 (2004).
- <sup>30</sup> Harshman, R. A., UCLA Working Papers in Phonetics No. 16, 1970 (<http://publish.uwo.ca/~harshman/wpppfac0.pdf>).
- <sup>31</sup> Helgaker, T. and Taylor, P. R., *Modern Electronic Structure Theory* (World Scientific, Singapore, 1995).
- <sup>32</sup> Hrycak, T. and Rokhlin, V., “An improved fast multipole algorithm for potential fields,” *SIAM J. Sci. Comput. (USA)* **19**, 1804 (1998).
- <sup>33</sup> Kalos, M. H., “Monte Carlo calculations of the ground state of three- and four-body nuclei,” *Phys. Rev.* **128**, 1791 (1962).
- <sup>34</sup> Kalos, M. H., “Monte Carlo integration of the Schrödinger equation,” *Trans. N. Y. Acad. Sci.* **26**, 497 (1963).

- <sup>35</sup> Kato, T., "Fundamental properties of Hamiltonian operators of Schrödinger type," *Trans. Am. Math. Soc.* **70**, 195 (1951).
- <sup>36</sup> Klopper, W., *Modern Methods and Algorithms of Quantum Chemistry*, NIC Series Vol. 1, edited by Grotendorst, J. (John von Neumann Institute for Computing, Jülich, 2000), pp. 153–201.
- <sup>37</sup> Klopper, W. and Samson, C. C. M., "Explicitly correlated second-order Møller Plesset methods with auxiliary basis sets," *J. Chem. Phys.* **116**, 6397 (2002).
- <sup>38</sup> Kroonenberg, P. M. and de Leeuw, J., "Principal component analysis of three-mode data by means of alternating least squares algorithms," *Psychometrika* **45**, 69 (1980).
- <sup>39</sup> Leurgans, S. E., Moyeed, R. A., and Silverman, B. W., "Canonical correlation analysis when the data are curves," *J. R. Stat. Soc. Ser. B (Methodol.)* **55**, 725 (1993).
- <sup>40</sup> Lindh, R., Olsen, J., and Roos, B. O., "Low-rank configuration interaction with orbital optimization—the LR SCF approach," *Chem. Phys. Lett.* **148**, 276 (1988).
- <sup>41</sup> Löwdin, P.-O., "Quantum theory of many-particle systems. I. Physical interpretations by means of density matrices, natural spin-orbitals, and convergence problems in the method of configuration interaction," *Phys. Rev.* **97**, 1474 (1955).
- <sup>42</sup> Lüchow, A. and Fink, R., "On the systematic improvement of fixed-node diffusion quantum Monte Carlo energies using natural orbital CI guide functions," *J. Chem. Phys.* **113**, 8457 (2000).
- <sup>43</sup> Meyer, Jr., C. D., "Generalized inversion of modified matrices," *SIAM J. Appl. Math.* **24**, 315 (1973).
- <sup>44</sup> Mohlenkamp, M. J. and Monzón, L., "Trigonometric identities and sums of separable functions," *Math. Intell.* **27**, 65 (2005).
- <sup>45</sup> Mohlenkamp, M. J. and Young, T., in *Recent Advances in Computational Science: Selected Papers from the International Workshop on Computational Sciences and Its Education*, edited by Jorgensen, P., Shen, X., Shu, C.-W., and Yan, N. (World Scientific, Singapore, 2007).
- <sup>46</sup> Muir, T., *A Treatise on the Theory of Determinants* (privately published, Albany, 1930), revised and enlarged by Metzler, W. H..
- <sup>47</sup> Noga, J., Kutzelnigg, W., and Klopper, W., "CC-R12, a correlation cusp corrected coupled-cluster method with a pilot application to the Be<sub>2</sub> potential curve," *Chem. Phys. Lett.* **199**, 497 (1992).
- <sup>48</sup> Olsen, J., Malmquist, P.-A., Roos, B. O., Lindh, R., and Widmark, P.-O., "A non-linear approach to configuration interaction: The low-rank CI method (LR CI)," *Chem. Phys. Lett.* **133**, 91 (1987).
- <sup>49</sup> Pauncz, R., *The Symmetric Group in Quantum Chemistry* (CRC, Boca Raton, FL, 1995).
- <sup>50</sup> Persson, B. J. and Taylor, P. T., "Accurate quantum-chemical calculations: The use of Gaussian-type geminal functions in the treatment of electron correlation," *J. Chem. Phys.* **105**, 5915 (1996).
- <sup>51</sup> Persson, B. J. and Taylor, P. R., "Molecular integrals over Gaussian-type geminal basis functions," *Theor. Chem. Acc.* **97**, 240 (1997).
- <sup>52</sup> Prasolov, V. V., *Problems and Theorems in Linear Algebra*, Translations of Mathematical Monographs Vol. 134 (American Mathematical Society, Providence, RI, 1994).
- <sup>53</sup> Reed, M., and Simon, B., *Methods of Modern Mathematical Physics. II. Fourier Analysis, Self-adjointness* (Academic, New York, 1975).
- <sup>54</sup> Reed, M. and Simon, B., *Methods of Modern Mathematical Physics. IV. Analysis of Operators* (Academic, New York, 1978).
- <sup>55</sup> Rellich, F., "Störungstheorie der Spektralzerlegung. V.," *Math. Ann.* **118**, 462 (1942).
- <sup>56</sup> Rudin, S. P., "Configuration interaction with non-orthogonal Slater determinants applied to the hubbard model, atoms, and small molecules," Ph.D. thesis, The Ohio State University, 1997.
- <sup>57</sup> Rychlewski, J., Cencek, W., and Komasa, J., "The equivalence of explicitly correlated slater and gaussian functions in variational quantum chemistry computations: The ground state of H<sub>2</sub>," *Chem. Phys. Lett.* **229**, 657 (1994).
- <sup>58</sup> Sherrill, C. D. and Schaefer, III, H. F., "The configuration interaction method: Advances in highly correlated approaches," *Adv. Quantum Chem.* **127**, 143 (1999).
- <sup>59</sup> Smilde, A., Bro, R., and Geladi, P., *Multi-Way Analysis. Applications in the Chemical Sciences* (Wiley, New York, 2004).
- <sup>60</sup> Yarvin, N. and Rokhlin, V., "Generalized Gaussian quadratures and singular value decompositions of integral operators," *SIAM J. Sci. Comput. (USA)* **20**, 699 (1999).
- <sup>61</sup> Zanghellini, J., "Multi-electron dynamics in the ionization of molecules by strong laser pulses," Ph.D. thesis, Vienna University of Technology, 2004.