SIAM J. APPL. MATH.
Vol. 72, No. 6, pp. 1742–1771

CAPTURING THE INTERELECTRON CUSP USING A GEMINAL LAYER ON AN UNCONSTRAINED SUM OF SLATER DETERMINANTS*

MARTIN J. MOHLENKAMP†

Abstract. Representations of the wavefunction of the multiparticle Schrödinger equation using an unconstrained sum of Slater determinants have the potential to be very efficient, except that they cannot capture the interelectron cusp. To remedy this shortcoming, we extend these representations to include interelectron geminals and analyze the algorithmic implications of doing so.

Key words. multiparticle Schrödinger equation, geminal, interelectron cusp, R12 methods, F12 methods

AMS subject classifications. 65Z05, 65D15, 81-08

DOI. 10.1137/110823900

1. Introduction.

We consider the time-independent, nonrelativistic, N-electron multiparticle Schrödinger equation with the Born–Oppenheimer approximation that the nuclei are point charges. This equation is an eigenproblem $\mathcal{H}\psi = \lambda\psi$. The eigenvalues λ correspond to energies and the smallest energies are of greatest interest. The wavefunction ψ is a function of N variables, each of which has a three-dimensional spatial part $\mathbf{r} = (x, y, z)$ and a spin variable $\sigma \in \{-1/2, 1/2\}$, which we combine as $\gamma = (\mathbf{r}, \sigma)$. The wavefunction ψ is also required to be antisymmetric under the exchange of any two γ_i and γ_j for $i \neq j$, so that, for example, $\psi(\gamma_1, \gamma_2, \dots) = -\psi(\gamma_2, \gamma_1, \dots)$. The Hamiltonian $\mathcal{H} = \mathcal{T} + \mathcal{V} + \mathcal{W}$ consists of kinetic, nuclear potential and interelectron potential operators defined respectively by

$$\mathcal{T} = -\frac{1}{2}\sum_{i=1}^{N} \nabla_i^2 \,, \tag{1.1}$$

$$\mathcal{V} = \sum_{i=1}^{N} V(\mathbf{r}_i) \,, \quad \text{and} \tag{1.2}$$

$$\mathcal{W} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N} \frac{1}{\|\mathbf{r}_i - \mathbf{r}_j\|} \,. \tag{1.3}$$

The differential operator ∇^2 is a three-dimensional Laplacian, and the potential $V(\mathbf{r})$ is a sum of terms of the form $Z_a/\|\mathbf{r} - \mathbf{r}_a\|$ from a nucleus of charge Z_a at position $\mathbf{r}_a$.

Explicit formulas for solutions of the multiparticle Schrödinger equation are essentially limited to $N = 1$. Nearly a century of work on numerical approximations has yielded great successes, but the problem remains challenging. A tremendous amount of computer power is still spent applying current methods, and a tremendous amount of brain power is still spent improving these methods. In very general terms, our

*Received by the editors February 9, 2011; accepted for publication (in revised form) September 17, 2012; published electronically November 15, 2012. This material is based upon work supported by the National Science Foundation under grant DMS-0545895.
http://www.siam.org/journals/siap/72-6/82390.html
†Department of Mathematics, Morton Hall 321, 1 Ohio University, Athens, OH 45701 (mohlenka@ohio.edu).

analysis of the situation is as follows. Current methods impose extra constraints on the wavefunction representation in order to make the algorithms and analysis easier. Such an approach is quite effective for relatively low-quality solutions, which were all that could be considered on the computers available when these methods originated. However, when one requests higher-quality solutions, these constraints cause the wavefunction representation to be extremely inefficient and the computational cost to explode. Many recent algorithmic improvements consist of finding the least damaging way to truncate the representation or the best way to extrapolate and capture the effect of neglected terms. These improvements do not address the central impediment of inefficient representation.

Consequently, our approach is to remove constraints and seek very efficient representations for the wavefunction. Such unconstrained representations require more complicated algorithms to compute. In [4] we presented the necessary algorithms in the basic case, and in [35] we extended them to scale properly for large systems. In this paper we extend the method so it can also capture the interelectron cusp by including geminal functions that depend on the interelectron distances $\|\mathbf{r}_i - \mathbf{r}_j\|$. While we think our method has great potential, it is not mature, so at this point we cannot make any specific claims about its efficacy. In the following narrative section we explain the rationale behind our computational method and the motivation for the current paper.

1.1. Motivational narrative. At minimum, our numerical approximation for the wavefunction ψ must be an antisymmetric function of N variables. The simplest way to construct such a function is to antisymmetrize a sum of products of functions of one electronic variable, as in

$$\psi(\gamma_1, \gamma_2, \ldots, \gamma_N) = \mathcal{A} \sum_{l=1}^{r} \prod_{i=1}^{N} \phi_i^l(\gamma_i)\,, \tag{1.4}$$

which could also be written as a sum of Slater determinants. Having chosen the representation (1.4), the next question is how to compute approximate wavefunctions of this form. The configuration interaction (CI) family of methods (see, e.g., [19]) uses notions of reference states and excitations to preselect the functions $\{\phi_i^l\}$ and impose structure on the representation. It then introduces coefficients s_l, inserts (1.4) into a variational formulation of the equation, and uses linear algebra to determine the s_l to give the best approximate wavefunction. The advantage of such a procedure is that the computations needed to determine s_l are relatively straightforward. The disadvantage of CI is that it imposes both structural constraints and pairwise constraints (such as orthogonality) on the $\{\phi_i^l\}$. As with any optimization problem, imposing constraints prevents one from achieving the (unconstrained) optimum. For example, suppose the true wavefunction is

$$\psi = \mathcal{A} \prod_{i=1}^{N} \theta_i(\gamma_i) + \mathcal{A} \prod_{i=1}^{N} (\theta_i(\gamma_i) + \theta_{i+N}(\gamma_i))\,, \tag{1.5}$$

where $\{\theta_j\}_{j=1}^{2N}$ form an orthogonal set. If there are no constraints, then (1.5) is already of the form (1.4) with $r = 2$. Suppose we impose the simplest orthogonality constraint, which holds in CI: for all i, j, l, and m, the functions ϕ_i^l and ϕ_j^m are either orthogonal or scalar multiples of one another. To put (1.5) in the form (1.4) while satisfying this constraint, we can multiply out the second term, obtaining $r = 2^N$ terms. Although

this is not the only way to satisfy the constraint, it captures the essential feature: all such representations have $r \geq 2^N$. As this simple example shows, constraints on $\{\phi_i^l\}$ could lead to representations that are "exponentially" inefficient in terms of r. (Note that sparse grid methods (see, e.g., [8]) applied to the multiparticle Schrödinger equation in many ways parallel CI; see [14]. The sparse grid structure also makes the computations relatively straightforward, the analysis tractable [52, 53, 54, 56], and the representations potentially exponentially inefficient.)

To obtain an optimal approximate wavefunction of the form (1.4), one must remove the constraints on $\{\phi_i^l\}$. Here optimal means either the minimum error for fixed r or the minimum r for fixed error. Since the unconstrained approach may be exponentially more efficient than the constrained approach, it is potentially very powerful and worth developing. The disadvantage of the unconstrained approach is that one has to compute all the $\{\phi_i^l\}$, and the method to do so is not at all straightforward, mainly due to the multilinear structure. General methods for computing on functions of many variables by using a multilinear structure were developed in [2, 3]; see also [17, 12, 15, 16, 10]. In [4] we introduced an algorithm to produce an approximation of the form (1.4) without any constraints on $\{\phi_i^l\}$. The algorithm is based on the iteration $\psi_n = -\mathcal{G}_\mu(\mathcal{V} + \mathcal{W})\psi_{n-1}$ using the Green function $\mathcal{G}_\mu = (\mathcal{T} - \mu\mathcal{I})^{-1}$ with μ the current energy estimate. Within the iteration, we alternately update the functions $\{\phi_i^l\}$ for a single electron index i while fixing the functions in the other electron indexes. To update these functions and the energy estimate we need to compute (antisymmetric) inner products of the forms

$$\left\langle \tilde{\psi}, \psi \right\rangle , \quad \left\langle \tilde{\psi}, (\mathcal{V} + \mathcal{W})\psi \right\rangle , \quad \text{and} \quad \left\langle \tilde{\psi}, \mathcal{G}_\mu(\mathcal{V} + \mathcal{W})\psi \right\rangle , \tag{1.6}$$

all with ψ and $\tilde{\psi}$ of the form (1.4). It is yet to be determined if the benefit of having (hopefully) optimal representations outweighs the computational expense of solving the unconstrained problem. However, even if the unconstrained approach in [4] using (1.4) were a complete success and rendered CI obsolete, we could not declare victory because of well-known, serious flaws in the representation (1.4).

The first flaw is that (1.4) is untenable for large systems. Consider a simple thought experiment with K noninteracting, identical subsystems. The overall wavefunction is then the product of the wavefunctions for the subsystems, and the overall eigenvalue is the sum of the eigenvalues of the subsystems. The wavefunction representation should correctly obtain the sum of the eigenvalues of the subsystems, a requirement called "size-extensivity" in [19]. (The terms "size-extensive" and "size-consistent" are used extensively and inconsistently in the literature; see, e.g., [1].) The CI method with a fixed level of excitation fails in this requirement (see, e.g., [19]). It is not clear how to apply the argument from CI in the unconstrained case, but another simple argument shows that it also is untenable for large systems: If the wavefunction for each system is of the form (1.4), then representing the overall wavefunction also in the form (1.4) would require us to multiply out and obtain r^K terms, which grows exponentially in K. The standard approach to achieving size-extensivity is coupled cluster (CC) (e.g., [19, 9]), which replaces (1.4) by a representation obtained by applying the exponential of a sum of excitation operators to a reference state. Although not as straightforward as CI, CC is well developed, which is an advantage. The disadvantage of CC is it, like CI, imposes constraints on the component $\{\phi_i^l\}$ and thus may be exponentially inefficient. We have so far been unable to construct an unconstrained version of CC.

Instead, we consider wavefunctions of the recursive form

$$\psi = \mathcal{A} \sum_{l=1}^{r} \prod_{k=1}^{K} \left(\sum_{l_k=1}^{r_k} \prod_{i=1}^{N_k} \phi_i^{l,k}(\gamma_{i,k}) \right) . \tag{1.7}$$

Applied to a set of noninteracting subsystems, we can take $r = 1$, and the representation (1.7) is then size-extensive and has computational cost growing only linearly with K. Allowing larger r handles interacting subsystems. The representation does assume there are identifiable subsystems and so would not be appropriate for an electron gas, for example. Computing (1.6) with ψ and $\tilde{\psi}$ of the form (1.7) is a conceptual and algorithmic challenge, but in [35] we presented the principles and algorithms to do so. To our knowledge, the form (1.7) has not been used for the Schrödinger equation, but related forms have been considered in multilinear algebra [38, 13].

The second flaw in (1.4) is that the interelectron cusp will cause the approximation error to decay slowly with r and thus r to grow rapidly for increasing accuracy. Since ψ must be continuous [23], so must $\mathcal{T}\psi + \mathcal{V}\psi + \mathcal{W}\psi$, even though $\mathcal{V}$ and $\mathcal{W}$ have singularities. The singularities in $\mathcal{V}$ are at the nuclei $\mathbf{r}_a$ and the singularities in $\mathcal{W}$ are whenever two electron coordinates coincide as $\mathbf{r}_i = \mathbf{r}_j$, so these two singularities cannot cancel. When ψ is not zero for a value of $\mathbf{r}_i = \mathbf{r}_j$, the singularity in $\mathcal{W}\psi$ must be canceled by a singularity in $\mathcal{T}\psi$, which means ψ must not be twice differentiable at that $\mathbf{r}_i = \mathbf{r}_j$. Kato [23] characterized the interelectron cusp for $\mathbf{r}_2 \to \mathbf{r}_1$ by

$$\frac{\partial}{\partial t} \left(\frac{1}{4\pi t^2} \int_{\|\mathbf{r}_1 - \mathbf{r}_2\| = t} \psi(\gamma_1, \gamma_2, \gamma_3, \dots) d\omega_2 \right) \Bigg|_{t=0} = \frac{1}{2} \psi(\gamma_1, (\mathbf{r}_1, \sigma_2), \gamma_3, \dots) , \tag{1.8}$$

where $d\omega_2$ indicates area measure on the sphere with $\mathbf{r}_2$ varying. (A characterization without using the spherical average appears in [39].) When $\sigma_1 = \sigma_2$ the antisymmetry condition forces the right-hand side of (1.8) to be zero, and there is no cusp. When $\sigma_1 = -\sigma_2$, (1.8) says that ψ can be written locally as a differentiable function times the linear cusp $1 + (1/2)\|\mathbf{r}_1 - \mathbf{r}_2\|$. Approximations of such a function using representations like (1.4) will converge slowly in r, so obtaining a highly accurate approximation would require many terms.

The issue of the interelectron cusp has long been recognized. For two electrons, the early (1928) work of Hylleraas [21] showed that including "R12" terms of the form $w(\|\mathbf{r}_1 - \mathbf{r}_2\|)\theta(\mathbf{r}_1)\theta(\mathbf{r}_2)$ can greatly improve the accuracy of short approximations. The connecting function w is called a *geminal* (or a Jastrow function after [22]). The condition (1.8) shows that w should satisfy $w'(0) = w(0)/2 \neq 0$ but one has some flexibility. For example, the "Slater-type" geminal $w(t) = 1 - \exp(-ct)/(2c + 1)$ satisfies this conditions for any $0 < c$. Such Slater-type geminals perform better numerically than simple linear geminals and are preferred in, e.g., [48, 51, 20, 26, 45, 28, 24, 50]. On further consideration, one notices that the cusp itself occurs on a set of measure (volume) zero, so it cannot directly affect the values of integrals, and so it is in some sense irrelevant. What the asymptotics (1.8) tell you is that there is a region (the "correlation hole") around $\mathbf{r}_1 = \mathbf{r}_2$ where the dependence of ψ on $\|\mathbf{r}_1 - \mathbf{r}_2\|$ is more important than its dependence individually on $\mathbf{r}_1$ and $\mathbf{r}_2$. This realization led to the use of geminals that do not satisfy the cusp condition but are easier to compute with, in particular Gaussian geminals [6, 46, 33, 43, 30, 29, 41, 42, 7, 25, 47].

For $2 < N$ one has to decide how the pairwise geminals are combined. The first candidate is the multiplicative form

$$\prod_{m=1}^{N} \prod_{n=m+1}^{N} w(\|\mathbf{r}_m - \mathbf{r}_n\|)\,. \tag{1.9}$$

Assuming $w'(0) = w(0)/2 \neq 0$, taking (1.9) times a differentiable function results in a function that satisfies the cusp condition (1.8) when $\mathbf{r}_m = \mathbf{r}_n$ for every $m \neq n$ and all noncoinciding values of the other $\mathbf{r}_k$. As a consequence, encouraging theoretical results are available [11, 55]. However, integrating (1.9) in any variable couples all the remaining variables, so it currently appears intractable to use within the class of methods we consider. One therefore seeks a tractable, though theoretically inferior, form. Suppose we chose to use a Slater-type geminal $w(t) = 1 - \exp(-ct)/(2c+1)$. Inserting in (1.9) and expanding, we obtain

(1.10)

$$\begin{aligned}
&\prod_{m=1}^{N} \prod_{n=m+1}^{N} \left(1 - \frac{\exp(-c\|\mathbf{r}_m - \mathbf{r}_n\|)}{2c+1}\right) \\
&\quad = 1 + \sum_{m=1}^{N} \sum_{n=m+1}^{N} \frac{-\exp(-c\|\mathbf{r}_m - \mathbf{r}_n\|)}{2c+1} \\
&\qquad + \sum_{m=1}^{N} \sum_{n=m+1}^{N} \sum_{j=1}^{N} \sum_{\substack{k=j+1 \\ (j,k)\neq(m,n)\neq(k,j)}}^{N} \frac{\exp(-\|\mathbf{r}_m - \mathbf{r}_n\|)\exp(-\|\mathbf{r}_j - \mathbf{r}_k\|)}{(2c+1)^2} + \cdots .
\end{aligned}$$

If all $\mathbf{r}_k$ are far apart, then (1.9) reduces to the first term 1 in (1.10). If any two $\mathbf{r}_k$ are close but all others are far apart, then (1.9) reduces to the first two terms. Thus we see that the additive form

$$1 + \frac{1}{2} \sum_{m \neq n} w(\|\mathbf{r}_m - \mathbf{r}_n\|) \tag{1.11}$$

with $w(0) = -1/(2c+1)$ and $w'(0) = c/(2c+1)$ results in a function that satisfies the cusp condition (1.8) when $\mathbf{r}_m = \mathbf{r}_n$ for every $m \neq n$, but only when the other $\mathbf{r}_k$ are all far apart. Note that this additive form is similar in structure to the inter-electron potential operator $\mathcal{W}$ in (1.3). The CI-R12 methods build upon CI (see the review [27]) are based on the additive form (1.11). The advantage of these methods is they are well-developed (although no longer straightforward) and the disadvantage is that the constraints may cause exponential inefficiencies. The CC-R12 methods (see the reviews [48, 51]) incorporate the geminals into the excitation operators in the exponential. An expansion of the exponential includes (1.11) and additional terms with multiple geminals such as $w(\|\mathbf{r}_m - \mathbf{r}_n\|)w(\|\mathbf{r}_j - \mathbf{r}_k\|)$ but does not include all of the third term in (1.10). CC-R12 inherits structure from CC and so has the advantage of being well-developed and the disadvantage of potential exponential inefficiency. See the text [19] for a general overview of R12 methods.

In this work we use a slight generalization of (1.11). We allow a set of different geminals $\{w_p\}_{p=0}^{P}$ to be used in order to capture different length scales; they could, for example, be $\exp(-ct)$ for different values of c. We allow the constant term to act independently of the geminals. To unify the notation we include it by setting $w_0 \equiv 1$; in computations this will result in a special, but easier, case, which we

will usually not describe. To allow dependence on the difference in spin, we denote $|\gamma - \gamma'| = (\|\mathbf{r} - \mathbf{r}'\|, |\sigma - \sigma'|)$ and consider geminals of the form $w(|\gamma - \gamma'|)$. The resulting wavefunction representation is

$$\psi = \mathcal{A} \sum_{p=0}^{P} \left(\frac{1}{2} \sum_{m \neq n} w_p(|\gamma_m - \gamma_n|) \right) \sum_{l=1}^{r_p} \prod_{i=1}^{N} \phi_i^{p,l}(\gamma_i) \,. \tag{1.12}$$

The purpose of the current paper is to develop methods based on the unconstrained form of (1.12), thus overcoming the second flaw in (1.4). Specifically, we present algorithms for computing (1.6) with ψ and $\tilde{\psi}$ of the form (1.12). As discussed above, the advantage of this approach is possibly much more efficient representations, and the disadvantage is the difficulty in constructing the algorithms. We hope through our work here to alleviate this disadvantage so that one can obtain the benefits of the advantage.

To overcome both the first and second flaws at the same time will require merging (1.7) and (1.12). While we see no obstruction to doing so, it will be algorithmically challenging, and we do not attempt to do so here. Left unresolved are the possibility of doing an unconstrained version of CC and/or including geminal layers of multiplicative form. Based on experience with the Schrödinger equation, there are likely further flaws to be overcome.

It is worth noting at this point some other differences in the approach we are developing versus traditional approaches. First, a great deal of the conceptual structure of traditional approaches does not exist within our framework. Specifically, there is no Fock operator, reference state, excitation, active space, virtual space, quantization, level of theory, level of perturbation, hierarchy of models, projection operator, etc. Second, the approximations within our methods occur at specific places. The first approximation is a decomposition of the Green function $\mathcal{G}_\mu$; as shown in [4], very accurate and efficient approximations are available. The second approximation (of a sort) is in the choice of grouping of variables in (1.7) and choice of geminals in (1.12). We expect these choices to be made either empirically or adaptively. The third and main approximation is in the choice of r in (1.4), r_k in (1.7), and r_p in (1.12). Our essential assumption is that good approximations may be obtained with small values of these r. At the level presented here and in [4, 35], the remainder of the method is exact. There are of course approximations in the manipulations of $\{\phi_i^l\}$ and computations of integrals, but these are decoupled from the multivariate problem and deferred to a numerical method on one or two electronic variables.

1.2. Main algorithmic analysis. In section 2 we generalize the method in [4] to compute the inner products (1.6) using the wavefunction representation (1.12). We consider the inner product $\langle \tilde{\psi}, (\mathcal{V} + \mathcal{W})\psi \rangle$ in detail and then simplify to $\langle \tilde{\psi}, \psi \rangle$ and extend to $\langle \tilde{\psi}, \mathcal{G}_\mu(\mathcal{V} + \mathcal{W})\psi \rangle$. The basic method is to use linear algebra to reduce from formulas with $N \times N$ Slater determinants to formulas with Slater determinants of size given by the number of variables involved in the geminals and potentials. With a geminal in $\tilde{\psi}$, another in ψ, and the geminal-like structure in $\mathcal{W}$, at most six variables are involved. These variables may coincide in various ways, so we have a set of formulas with two to six variables. We represent the central structures in these formulas graphically, using ● to represent a variable such as γ_1 and ●—● to represent a geminal such as $w(|\gamma_1 - \gamma_2|)$ connecting two variables. The geminals in the inner product $\langle \tilde{\psi}, \mathcal{W}\psi \rangle$ result in formulas corresponding to the eight overlap structures ●—●—●—●—●—●, ●—●—●—●—●, ⬭—●—●, ●—●—●—●, ●—●—●, ⬭—●,

TABLE 1.1
Heuristically minimal total operations sorted by complexity. The entry gives the number of operations of the complexity given by the product of the row and column headings.

	1	M	M_*	M^2	MM_*
1		104	28	6	3
N	411	241		3	
N^2	181	184	44	2	1
N^3	124	8			
N^4	122	65	3		
N^5	6				
N^6	6				

 , and  . For example,  corresponds to formulas with geminals configured like $w_1(|\gamma_1 - \gamma_2|)w_2(|\gamma_1 - \gamma_2|)w_3(|\gamma_2 - \gamma_3|)$. Similarly, $\langle\tilde{\psi}, \mathcal{V}\psi\rangle$ results in seven distinct structures. To compute these formulas, one needs to expand out the determinants. After combining equivalent terms, we obtain 584 terms to compute. Since generating so many terms by hand would be tedious and error prone, we present an algorithm to generate them.

The cost to compute each of these terms depends on three parameters, the first of which is the number of electrons N. The second parameter, which we denote M, is the cost to represent, add, multiply, or integrate a function of γ. If a nonadaptive method is used for the functions of γ, then M would be the number of basis functions or grid points. The third parameter, which we denote M_*, is the cost to perform a convolution such as $g(\gamma_0) = \int w_1(|\gamma_0 - \gamma_1|)f(\gamma_1)d\gamma_1$. Since convolution costs at least one integration total and at most one integration per output value, we have $M < M_* < M^2$. If M represents the number of points on an equispaced grid, then $M_* = \mathcal{O}(M \log M)$ using the fast Fourier transform. For our analysis we will assume $N \ll M$, so that we prefer even a high power of N to a factor of M.

In section 3 we determine how to actually compute these 584 terms and the cost to do so. The formulas each include two to six integrals over space and the same number of sums over the electrons. Brute-force summation and integration would thus lead to cost $\mathcal{O}(N^6M^6)$. Instead, we choose the order of summation and integration to involve few indexes and variables at a time and thus to minimize the cost. To handle all these terms we develop an algorithm for reducing cases to simpler cases and then handle a few residual cases individually. This procedure indicates that we need 5516 summations and integrations of various complexities. Many of these operations are duplicates, however, so we can cache and reuse them. Heuristically optimizing such reuse, we obtain the operations counts in Table 1.1. The scaling with N of N^6 is misleading because following our assumption $N \ll M$ we chose to include N^6 and N^5 computations rather than additional N^4M_* computations. The overall dominant computation is N^2MM_*. The computations with complexity including M^2 or MM_* all come from the cyclic overlap structure  , which does not allow one to integrate in any variable without involving both other variables. Thus although some terms cannot occur for $N < 6$, the dominant terms are already present by $N = 3$.

1.3. Is it worthwhile? The goal in replacing the representation (1.4) by (1.12) is to allow a more efficient computation of the wavefunction. Since computing with (1.12) is more expensive, we should consider whether this replacement is worthwhile. For each inner product, both (1.4) and (1.12) share an initialization cost of $2N^2M + \mathcal{O}(N^3)$. The main cost for (1.4) is $N^2M_* + M_* + N^2M + 2NM + 2M + N$ and that

for (1.12) is given in Table 1.1. The number of inner products is quadratic in the number of summands in the representation, which is r for (1.4) and $R = \sum_p r_p$ for (1.12). Using the dominant cost for the two methods, (1.12) is worthwhile only if

$$R^2N^2MM_* < r^2N^2M_* \quad \leftrightarrow \quad R < r/\sqrt{M}\,. \tag{1.13}$$

A construction from [2, 3] allows us to convert a representation (1.12) to (1.4) and thus relate r to R. Suppose we have approximations

$$w_p(|\gamma - \gamma'|) \approx \sum_{q=1}^{Q} u_p^q(\gamma)u_p^q(\gamma') \tag{1.14}$$

that are sufficiently accurate on the region of space where the wavefunction is significant. Noting that

$$\frac{d^2}{dt^2}\left(\prod_{i=1}^{N}(1 + tu_p^q(\gamma_i))\right)\bigg|_{t=0} = 2\sum_{m\neq n} u_p^q(\gamma_m)u_p^q(\gamma_n)\,, \tag{1.15}$$

we can use a finite difference with h sufficiently small to obtain the approximation

$$\frac{1}{4h^2}\sum_{q=1}^{Q}\left(\prod_{i=1}^{N}(1 + hu_p^q(\gamma_i)) - 2 + \prod_{i=1}^{N}(1 - hu_p^q(\gamma_i))\right) \approx \frac{1}{2}\sum_{m\neq n} w_p(|\gamma_m - \gamma_n|)\,. \tag{1.16}$$

(We neglect $\log N$ factors due to loss-of-precision effects; see [2, 3].) Inserting into (1.12) and multiplying out yields a (potentially suboptimal) representation of the form (1.4) with $r = (2Q + 1)R$. Inserting into (1.13), this means the representation (1.12) is worthwhile only if $\sqrt{M} < 2Q + 1$.

We have not answered the question, but we have learned that the answer does not explicitly depend on N. We have also learned that the answer depends on how well the geminal separates in (1.14). For highly localized geminals, $Q \to M$ and (1.12) is worthwhile. However, delocalized geminals have $Q \to 1$ and are not worth using.

In section 4.1 we consider further implications of assuming that the geminals (excluding $w_0 \equiv 1$) are highly localized. We find that if the geminals are localized to a region of volume K, then the dominant cost is reduced from N^2MM_* to K^2N^2M. If $K^2M < N^2M_*$ then the cyclic structure is no longer dominant in cost. Using the now-dominant cost $3N^4M_*$ from Table 1.1, using (1.12) becomes worthwhile if

$$3R^2N^4M_* < r^2N^2M_* \quad \leftrightarrow \quad R < r/(\sqrt{3}N) \quad \leftrightarrow \quad \sqrt{3}N < 2Q + 1\,. \tag{1.17}$$

We thus conclude that localized geminals have the dual effects of making (1.4) less efficient and making (1.12) faster to compute. Whether such localized geminals are needed to represent ψ will depend on the specific problem.

1.4. Comparison with the literature. We have found that in our method there is a profusion of terms to compute. This finding is entirely in agreement with methods using geminals in the literature. As a sampling, [18, 32, 37, 36, 42, 25, 48, 51] describe many cases that need to be computed and lament that there are so many. We have also found that the cyclic case is the most difficult. Again this is consistent with the literature. This case is specifically mentioned as difficult in [31, 32, 42, 27].

To deal with all these difficult terms, many different approximation schemes have been developed. One set consists of simply neglecting terms that are too difficult.

Although this method makes numerical analysts cringe, one should remember that the goal of the practitioner is to obtain the best results for the available computational resources, not to obtain a "correct" result for a given model. An analysis of the sizes of the terms neglected in various methods is given in [28]. Arguments that the cyclic case often produces small results appear in [31, 27].

A second, very common method of approximation is the resolution of the identity (RI). In our survey of papers from 2010, RI was used in [48, 51, 20, 49, 44, 5, 45]. We consider RI in section 4.2 applied to our dominant cyclic case. This approximation reduces the dominant cost from N^2MM_* to QN^2M_*, where Q is the length of a separation like (1.14) for the identity operator (delta function). Although this approximation is popular in the literature, we see no reason to expect Q to be much smaller than M, and so we judge this approximation to not be useful for us. The failure of RI for difficult cases was noted in [27].

Another technique is to restrict to Gaussian geminals and use analytic formulas involving Gaussians whenever possible. This method is used in [6, 46, 33, 43, 30, 29, 41, 42, 7, 25, 47]. (Additionally, approximations of other geminals as a sum of Gaussians are sometimes used.) We consider the implications of using Gaussian geminals in section 4.3 applied to our dominant cyclic case. We find that Gaussians allow us to reduce the costs, but only by replacing MM_* factors by M^2 factors. The dominant cost is thus reduced from N^2MM_* to N^2M^2, which is still dominant.

2. Antisymmetric inner products including geminals. In this section we derive formulas for the antisymmetric inner products (1.6) where the wavefunction approximations are of the form (1.12). Since these inner products are linear, we can pull out the summations over p and l and consider each term separately. We thus consider (1.6) for

$$\begin{aligned} \psi &= \mathcal{A}\left(\frac{1}{2}\sum_{a\neq b} w_1(|\gamma_a-\gamma_b|)\right)\prod_{i=1}^{N}\phi_i(\gamma_i) \quad \text{and} \\ \tilde{\psi} &= \mathcal{A}\left(\frac{1}{2}\sum_{m\neq n} w_2(|\gamma_m-\gamma_n|)\right)\prod_{i=1}^{N}\tilde{\phi}_i(\gamma_i)\,. \end{aligned} \tag{2.1}$$

The operator $\mathcal{W}$ in (1.3) has the same structure as the geminal layer, so for notational convenience we let $w_3(|\gamma_u-\gamma_v|) = 1/(\|\mathbf{r}_u-\mathbf{r}_v\|)$ and use u and v as the summation indexes for $\mathcal{W}$.

In section 2.1 we set some notation and review standard results on antisymmetric inner products. In section 2.2 we generalize the methods of [4] that allow nonseparable elements in the inner products. In section 2.3 we derive the formulas for $\langle\tilde{\psi},\mathcal{W}\psi\rangle_{\mathcal{A}}$ and in section 2.4 we derive those for $\langle\tilde{\psi},\mathcal{V}\psi\rangle_{\mathcal{A}}$. In section 2.5 we present an algorithm for generating the results from sections 2.3 and 2.4 automatically and an algorithm for expanding into the terms we will actually compute. In section 2.6 we show how to modify the formulas in this section to include $\mathcal{G}_\mu$. In section 2.7 we state the results for the simpler case $\langle\tilde{\psi},\psi\rangle_{\mathcal{A}}$.

In section 3 we will analyze the formulas for $\langle\tilde{\psi},\mathcal{W}\psi\rangle_{\mathcal{A}}$ and $\langle\tilde{\psi},\mathcal{V}\psi\rangle_{\mathcal{A}}$ to determine their computational cost.

2.1. Preliminaries. Let $\bar{(\cdot)}$ denote complex conjugate and $(\cdot)^*$ denote conjugate transpose. An inner product is denoted $\langle\cdot,\cdot\rangle$ and includes conjugation of its second argument.

2.1.1. The antisymmetrizer and Löwdin's rule. The *antisymmetrizer* $\mathcal{A}$ (see, e.g., [40]) projects a function on its antisymmetric part, which for a separable function results in a Slater determinant, given by

$$\mathcal{A}\prod_{i=1}^{N}\phi_i(\gamma_i) = \frac{1}{N!}\begin{vmatrix} \phi_1(\gamma_1) & \phi_1(\gamma_2) & \cdots & \phi_1(\gamma_N) \\ \phi_2(\gamma_1) & \phi_2(\gamma_2) & \cdots & \phi_2(\gamma_N) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_N(\gamma_1) & \phi_N(\gamma_2) & \cdots & \phi_N(\gamma_N) \end{vmatrix}. \tag{2.2}$$

We define the antisymmetric inner product by

$$\langle\tilde{\psi},\psi\rangle_{\mathcal{A}} \stackrel{\text{def}}{=} \langle\mathcal{A}\tilde{\psi},\mathcal{A}\psi\rangle = \langle\tilde{\psi},\mathcal{A}\psi\rangle = \langle\mathcal{A}\tilde{\psi},\psi\rangle\,. \tag{2.3}$$

For separable functions, we can compute it by

$$\begin{aligned} &\left\langle \mathcal{A}\prod_{i=1}^{N}\tilde{\phi}_i(\gamma_i), \prod_{i=1}^{N}\phi_i(\gamma_i)\right\rangle \\ &\quad= \int\cdots\int \prod_{i=1}^{N}\bar{\phi}_i(\gamma_i)\frac{1}{N!}\begin{vmatrix} \tilde{\phi}_1(\gamma_1) & \tilde{\phi}_1(\gamma_2) & \cdots & \tilde{\phi}_1(\gamma_N) \\ \tilde{\phi}_2(\gamma_1) & \tilde{\phi}_2(\gamma_2) & \cdots & \tilde{\phi}_2(\gamma_N) \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{\phi}_N(\gamma_1) & \tilde{\phi}_N(\gamma_2) & \cdots & \tilde{\phi}_N(\gamma_N) \end{vmatrix} d\gamma_1\ldots d\gamma_N \\ &\quad= \frac{1}{N!}\begin{vmatrix} \langle\tilde{\phi}_1,\phi_1\rangle & \langle\tilde{\phi}_1,\phi_2\rangle & \cdots & \langle\tilde{\phi}_1,\phi_N\rangle \\ \langle\tilde{\phi}_2,\phi_1\rangle & \langle\tilde{\phi}_2,\phi_2\rangle & \cdots & \langle\tilde{\phi}_2,\phi_N\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle\tilde{\phi}_N,\phi_1\rangle & \langle\tilde{\phi}_N,\phi_2\rangle & \cdots & \langle\tilde{\phi}_N,\phi_N\rangle \end{vmatrix} = \frac{1}{N!}|\mathbb{L}| \end{aligned} \tag{2.4}$$

using the matrix $\mathbb{L}$ with entries

$$L(i,j) = \langle\tilde{\phi}_i,\phi_j\rangle = \int\tilde{\phi}_i(\gamma)\bar{\phi}_j(\gamma)d\gamma = \sum_{\sigma\in\{-1/2,1/2\}}\int\tilde{\phi}_i(\mathbf{r},\sigma)\bar{\phi}_j(\mathbf{r},\sigma)d\mathbf{r}\,. \tag{2.5}$$

The formula (2.4) is known as Löwdin's rule (e.g., [34, 40]).

2.2. Antisymmetric inner products with interference. To obtain formulas for the antisymmetric inner product involving nonseparable functions, we generalize the method in [4] for the potential operators. The main idea is to use Löwdin's rule (2.4) in all the variables with which the nonseparable function does not interfere, thereby reducing to a smaller problem. Suppose g is a (nonseparable) function of Q variables, which for simplicity we assume to be the first Q variables.

Let $\mathbb{L}$ be the matrix from (2.5), define the column vector $\tilde{\mathbf{\Phi}}$ whose entries are the functions $\tilde{\phi}_i$, and let $\mathbf{\Theta} = \mathbb{L}^{-1}\tilde{\mathbf{\Phi}}$ define the functions θ_i. (If $\mathbb{L}$ is singular a pseudoinverse can be used; see [4].) Determinants satisfy the general property $|\mathbb{A}| = |\mathbb{L}\mathbb{L}^{-1}\mathbb{A}| = |\mathbb{L}||\mathbb{L}^{-1}\mathbb{A}|$, so by inserting the Slater determinant $\mathcal{A}\prod_{i=1}^{N}\tilde{\phi}_i(\gamma_i)$ in for $|\mathbb{A}|$, we have $|\mathbb{L}|\mathcal{A}\prod_{i=1}^{N}\theta_i(\gamma_i) = \mathcal{A}\prod_{i=1}^{N}\tilde{\phi}_i(\gamma_i)$. Thus by simply including the scalar $|\mathbb{L}|$ we can replace $\tilde{\phi}_i$ by θ_i in our formulas. The advantage of $\{\theta_i\}$ is that it is biorthogonal to $\{\phi_i\}$, meaning that $\langle\theta_j,\phi_i\rangle = \delta_{ji}$. This property can be verified by writing the matrix of inner products as $\int\mathbf{\Theta}\mathbf{\Phi}^*d\gamma = \int\mathbb{L}^{-1}\tilde{\mathbf{\Phi}}\mathbf{\Phi}^*d\gamma = \mathbb{L}^{-1}\int\tilde{\mathbf{\Phi}}\mathbf{\Phi}^*d\gamma = \mathbb{L}^{-1}\mathbb{L} = \mathbb{I}$.

Using the transformation of $\tilde{\phi}_i$ to θ_i, to compute $\langle \prod_{i=1}^N \tilde{\phi}_i(\gamma_i), g \prod_{i=1}^N \phi_i(\gamma_i) \rangle_{\mathcal{A}}$ we must evaluate

(2.6)
$$\frac{|\mathbb{L}|}{N!} \int \cdots \int \bar{g}(\gamma_1, \ldots, \gamma_Q) \prod_{i=1}^N \bar{\phi}_i(\gamma_i) \begin{vmatrix} \theta_1(\gamma_1) & \theta_1(\gamma_2) & \cdots & \theta_1(\gamma_N) \\ \theta_2(\gamma_1) & \theta_2(\gamma_2) & \cdots & \theta_2(\gamma_N) \\ \vdots & \vdots & \ddots & \vdots \\ \theta_N(\gamma_1) & \theta_N(\gamma_2) & \cdots & \theta_N(\gamma_N) \end{vmatrix} d\gamma_1 \ldots d\gamma_N .$$

For $i > Q$ we can move $\bar{\phi}_i(\gamma_i)$ onto column i in the determinant and integrate. Since $\langle \theta_j, \phi_i \rangle = \delta_{ji}$ the only nonzero entry in column i is a one in row i, so the determinant collapses and we obtain

$$\frac{|\mathbb{L}|}{N!} \int \cdots \int \bar{g}(\gamma_1, \ldots, \gamma_Q) \prod_{i=1}^Q \bar{\phi}_i(\gamma_i) \begin{vmatrix} \theta_1(\gamma_1) & \cdots & \theta_1(\gamma_Q) \\ \vdots & \ddots & \vdots \\ \theta_Q(\gamma_1) & \cdots & \theta_Q(\gamma_Q) \end{vmatrix} d\gamma_1 \ldots d\gamma_Q . \tag{2.7}$$

From this point, the method will depend on g. As an illustration, let us consider $g = \frac{1}{2} \sum_{a \neq b} w_1(|\gamma_a - \gamma_b|)$. Although g couples all variables, it does so only two at a time, so we can use these methods to obtain

$$\begin{aligned} &\left\langle \prod_{i=1}^N \tilde{\phi}_i(\gamma_i), \frac{1}{2} \sum_{a \neq b} w_1(|\gamma_a - \gamma_b|) \prod_{i=1}^N \phi_i(\gamma_i) \right\rangle_{\mathcal{A}} \\ &= \frac{1}{2} \sum_{a \neq b} \left\langle \prod_{i=1}^N \tilde{\phi}_i(\gamma_i), w_1(|\gamma_a - \gamma_b|) \prod_{i=1}^N \phi_i(\gamma_i) \right\rangle_{\mathcal{A}} \\ &= \frac{1}{2} \frac{|\mathbb{L}|}{N!} \sum_{a \neq b} \int \int w_1(|\gamma_a - \gamma_b|) \bar{\phi}_a(\gamma_a) \bar{\phi}_b(\gamma_b) \begin{vmatrix} \theta_a(\gamma_a) & \theta_a(\gamma_b) \\ \theta_b(\gamma_a) & \theta_b(\gamma_b) \end{vmatrix} d\gamma_a d\gamma_b . \end{aligned} \tag{2.8}$$

When $a = b$ the determinant is zero, so we do not need to exclude it from the sum. The variables γ_a and γ_b are just variables of integration, so they do not actually depend on the indexes a and b and so we can rename them γ_0 and γ_1. We can then rename a and b as j_0 and j_1 and obtain

$$\frac{1}{2} \frac{|\mathbb{L}|}{N!} \sum_{j_0, j_1} \int \int w_1(|\gamma_0 - \gamma_1|) \bar{\phi}_a(\gamma_0) \bar{\phi}_b(\gamma_1) \begin{vmatrix} \theta_{j_0}(\gamma_0) & \theta_{j_0}(\gamma_1) \\ \theta_{j_1}(\gamma_0) & \theta_{j_1}(\gamma_1) \end{vmatrix} d\gamma_0 d\gamma_1 . \tag{2.9}$$

To actually compute this formula we would expand the determinant and for each term choose some order in which to integrate and sum.

2.3. Antisymmetric inner product including $\mathcal{W}$. To compute $\langle \tilde{\psi}, \mathcal{W}\psi \rangle_{\mathcal{A}}$ with ψ and $\tilde{\psi}$ in (2.1), we first follow section 2.2 to convert $\tilde{\phi}_i$ to θ_i. Then, as in the example in section 2.2, we move the sums and scalars outside the integrals and integrate in all the variables with which the nonseparable function $g = w_1(|\gamma_a - \gamma_b|)\bar{w}_2(|\gamma_m - \gamma_n|)w_3(|\gamma_u - \gamma_v|)$ does not interfere, i.e., all variables except $\{\gamma_a, \gamma_b, \gamma_m, \gamma_n, \gamma_u, \gamma_v\}$. If a, b, m, n, u, v are all distinct, then so are the six variables $\{\gamma_a, \gamma_b, \gamma_m, \gamma_n, \gamma_u, \gamma_v\}$ and we can rename them $\gamma_0, \ldots, \gamma_5$. Renaming the indexes as $j_0, \ldots, j_5$ we obtain the

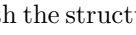
formula

$$\frac{1}{8}\frac{|\mathbb{L}|}{N!}\sum_{j_0,j_1,j_2,j_3,j_4,j_5}\int\int\int\int\int\int \bar{w}_1(|\gamma_0-\gamma_1|)w_2(|\gamma_2-\gamma_3|)\bar{w}_3(|\gamma_4-\gamma_5|)\bar{\phi}_{j_0}(\gamma_0)\bar{\phi}_{j_1}(\gamma_1)\bar{\phi}_{j_2}(\gamma_2)\bar{\phi}_{j_3}(\gamma_3)\bar{\phi}_{j_4}(\gamma_4)\bar{\phi}_{j_5}(\gamma_5)$$
$$\begin{vmatrix} \theta_{j_0}(\gamma_0) & \theta_{j_0}(\gamma_1) & \theta_{j_0}(\gamma_2) & \theta_{j_0}(\gamma_3) & \theta_{j_0}(\gamma_4) & \theta_{j_0}(\gamma_5) \\ \theta_{j_1}(\gamma_0) & \theta_{j_1}(\gamma_1) & \theta_{j_1}(\gamma_2) & \theta_{j_1}(\gamma_3) & \theta_{j_1}(\gamma_4) & \theta_{j_1}(\gamma_5) \\ \theta_{j_2}(\gamma_0) & \theta_{j_2}(\gamma_1) & \theta_{j_2}(\gamma_2) & \theta_{j_2}(\gamma_3) & \theta_{j_2}(\gamma_4) & \theta_{j_2}(\gamma_5) \\ \theta_{j_3}(\gamma_0) & \theta_{j_3}(\gamma_1) & \theta_{j_3}(\gamma_2) & \theta_{j_3}(\gamma_3) & \theta_{j_3}(\gamma_4) & \theta_{j_3}(\gamma_5) \\ \theta_{j_4}(\gamma_0) & \theta_{j_4}(\gamma_1) & \theta_{j_4}(\gamma_2) & \theta_{j_4}(\gamma_3) & \theta_{j_4}(\gamma_4) & \theta_{j_4}(\gamma_5) \\ \theta_{j_5}(\gamma_0) & \theta_{j_5}(\gamma_1) & \theta_{j_5}(\gamma_2) & \theta_{j_5}(\gamma_3) & \theta_{j_5}(\gamma_4) & \theta_{j_5}(\gamma_5) \end{vmatrix} d\gamma_0 d\gamma_1 d\gamma_2 d\gamma_3 d\gamma_4 d\gamma_5 .$$

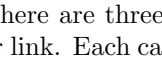
We represent (2.10) by the graph  with each ● representing a variable and  representing a geminal connection like $w_1(|\gamma_0-\gamma_1|)$. When a, b, m, n, u, v are not all distinct, then the determinant in (2.10) evaluates to zero. That is correct when $a = b$, $m = n$, or $u = v$ but is incorrect when other indexes coincide. For these cases we follow section 2.2 again, but fewer than six variables remain. We now work out the different ways this can occur and the resulting formulas. To obtain $\langle\tilde{\psi}, \mathcal{W}\psi\rangle_{\mathcal{A}}$ we add (2.10) and all these additional terms.

The only way to have five variables remaining is with the structure  . This structure has three cases, counted by which of w_1, w_2, or w_3 is the isolated portion  . Each case occurs four times, which when combined gives a scalar factor 4. When w_3 is isolated, this multiplicity comes from the ways in which a or b equals m or n. This case, including multiplicity, contributes

$$\frac{1}{2}\frac{|\mathbb{L}|}{N!}\sum_{j_0,j_1,j_2,j_3,j_4}\int\int\int\int\int \bar{w}_1(|\gamma_0-\gamma_1|)w_2(|\gamma_1-\gamma_2|)\bar{w}_3(|\gamma_3-\gamma_4|)$$
$$\bar{\phi}_{j_0}(\gamma_0)\bar{\phi}_{j_1}(\gamma_1)\bar{\phi}_{j_2}(\gamma_2)\bar{\phi}j_3(\gamma_3)\bar{\phi}_{j_4}(\gamma_4)$$
$$\begin{vmatrix} \theta_{j_0}(\gamma_0) & \theta_{j_0}(\gamma_1) & \theta_{j_0}(\gamma_2) & \theta_{j_0}(\gamma_3) & \theta_{j_0}(\gamma_4) \\ \theta_{j_1}(\gamma_0) & \theta_{j_1}(\gamma_1) & \theta_{j_1}(\gamma_2) & \theta_{j_1}(\gamma_3) & \theta_{j_1}(\gamma_4) \\ \theta_{j_2}(\gamma_0) & \theta_{j_2}(\gamma_1) & \theta_{j_2}(\gamma_2) & \theta_{j_2}(\gamma_3) & \theta_{j_2}(\gamma_4) \\ \theta j_3(\gamma_0) & \theta j_3(\gamma_1) & \theta j_3(\gamma_2) & \theta j_3(\gamma_3) & \theta j_3(\gamma_4) \\ \theta_{j_4}(\gamma_0) & \theta_{j_4}(\gamma_1) & \theta_{j_4}(\gamma_2) & \theta_{j_4}(\gamma_3) & \theta_{j_4}(\gamma_4) \end{vmatrix} d\gamma_0 d\gamma_1 d\gamma_2 d\gamma_3 d\gamma_4 .$$

Four variables can occur with the structures ,  , and , where  indicates that two geminals connect those two variables. There are three cases for  , counted by which of w_1, w_2, or w_3 is the isolated portion  . Each case has multiplicity two; when w_3 is isolated, the multiplicity comes from the choice of whether m or n equals a. This case, including multiplicity, contributes

$$\frac{1}{4}\frac{|\mathbb{L}|}{N!}\sum_{j_0,j_1,j_2,j_3}\int\int\int\int \bar{w}_1(|\gamma_0-\gamma_1|)w_2(|\gamma_0-\gamma_1|)\bar{w}_3(|\gamma_2-\gamma_3|)\bar{\phi}_{j_0}(\gamma_0)\bar{\phi}_{j_1}(\gamma_1)$$
$$\bar{\phi}_{j_2}(\gamma_2)\bar{\phi}_{j_3}(\gamma_3)\begin{vmatrix} \theta_{j_0}(\gamma_0) & \theta_{j_0}(\gamma_1) & \theta_{j_0}(\gamma_2) & \theta_{j_0}(\gamma_3) \\ \theta_{j_1}(\gamma_0) & \theta_{j_1}(\gamma_1) & \theta_{j_1}(\gamma_2) & \theta_{j_1}(\gamma_3) \\ \theta_{j_2}(\gamma_0) & \theta_{j_2}(\gamma_1) & \theta_{j_2}(\gamma_2) & \theta_{j_2}(\gamma_3) \\ \theta_{j_3}(\gamma_0) & \theta_{j_3}(\gamma_1) & \theta_{j_3}(\gamma_2) & \theta_{j_3}(\gamma_3) \end{vmatrix} d\gamma_0 d\gamma_1 d\gamma_2 d\gamma_3 . \tag{2.12}$$

There are three cases for  , counted by which of w_1, w_2, or w_3 is the center link. Each case has multiplicity eight; when w_2 is the center link, a multiplicity four comes from the ways in which a or b equals m or n, and another multiplicity two comes from whether the other of m or n equals u or v. This case, including multiplicity, contributes

$$\frac{|\mathbb{L}|}{N!} \sum_{j_0,j_1,j_2,j_3} \int\int\int\int \bar{w}_1(|\gamma_0-\gamma_1|) w_2(|\gamma_1-\gamma_2|) \bar{w}_3(|\gamma_2-\gamma_3|) \bar{\phi}_{j_0}(\gamma_0) \bar{\phi}_{j_1}(\gamma_1) \bar{\phi}_{j_2}(\gamma_2) \bar{\phi}_{j_3}(\gamma_3) \begin{vmatrix} \theta_{j_0}(\gamma_0) & \theta_{j_0}(\gamma_1) & \theta_{j_0}(\gamma_2) & \theta_{j_0}(\gamma_3) \\ \theta_{j_1}(\gamma_0) & \theta_{j_1}(\gamma_1) & \theta_{j_1}(\gamma_2) & \theta_{j_1}(\gamma_3) \\ \theta_{j_2}(\gamma_0) & \theta_{j_2}(\gamma_1) & \theta_{j_2}(\gamma_2) & \theta_{j_2}(\gamma_3) \\ \theta_{j_3}(\gamma_0) & \theta_{j_3}(\gamma_1) & \theta_{j_3}(\gamma_2) & \theta_{j_3}(\gamma_3) \end{vmatrix} d\gamma_0 d\gamma_1 d\gamma_2 d\gamma_3 \,. \tag{2.13}$$

There is only one case for  . It has multiplicity eight, counted by the ways in which a or b, m or n, and u or v are the center node. This case contributes

$$\frac{|\mathbb{L}|}{N!} \sum_{j_0,j_1,j_2,j_3} \int\int\int\int \bar{w}_1(|\gamma_0-\gamma_1|) w_2(|\gamma_1-\gamma_2|) \bar{w}_3(|\gamma_1-\gamma_3|) \bar{\phi}_{j_0}(\gamma_0) \bar{\phi}_{j_1}(\gamma_1) \bar{\phi}_{j_2}(\gamma_2) \bar{\phi}_{j_3}(\gamma_3) \begin{vmatrix} \theta_{j_0}(\gamma_0) & \theta_{j_0}(\gamma_1) & \theta_{j_0}(\gamma_2) & \theta_{j_0}(\gamma_3) \\ \theta_{j_1}(\gamma_0) & \theta_{j_1}(\gamma_1) & \theta_{j_1}(\gamma_2) & \theta_{j_1}(\gamma_3) \\ \theta_{j_2}(\gamma_0) & \theta_{j_2}(\gamma_1) & \theta_{j_2}(\gamma_2) & \theta_{j_2}(\gamma_3) \\ \theta_{j_3}(\gamma_0) & \theta_{j_3}(\gamma_1) & \theta_{j_3}(\gamma_2) & \theta_{j_3}(\gamma_3) \end{vmatrix} d\gamma_0 d\gamma_1 d\gamma_2 d\gamma_3 \,. \tag{2.14}$$

Three variables can occur with the structures and  . There are three cases for  , counted by which of w_1, w_2, or w_3 sticks out by itself. Each case has multiplicity eight; when w_3 sticks out, a multiplicity two comes from whether m or n equals a, and another multiplicity four comes from the ways in which u or v equals a or b. This case, including multiplicity, contributes

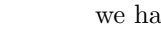
$$\frac{|\mathbb{L}|}{N!} \sum_{j_0,j_1,j_2} \int\int\int \bar{w}_1(|\gamma_0-\gamma_1|) w_2(|\gamma_0-\gamma_1|) \bar{w}_3(|\gamma_1-\gamma_2|) \bar{\phi}_{j_0}(\gamma_0) \bar{\phi}_{j_1}(\gamma_1) \bar{\phi}_{j_2}(\gamma_2) \begin{vmatrix} \theta_{j_0}(\gamma_0) & \theta_{j_0}(\gamma_1) & \theta_{j_0}(\gamma_2) \\ \theta_{j_1}(\gamma_0) & \theta_{j_1}(\gamma_1) & \theta_{j_1}(\gamma_2) \\ \theta_{j_2}(\gamma_0) & \theta_{j_2}(\gamma_1) & \theta_{j_2}(\gamma_2) \end{vmatrix} d\gamma_0 d\gamma_1 d\gamma_2 \,. \tag{2.15}$$

There is only one case for  . It has multiplicity eight; a multiplicity four comes from the ways in which a or b equals m or n, and another multiplicity two comes from whether the other of m or n equals u or v. This case contributes

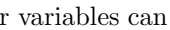
$$\frac{|\mathbb{L}|}{N!} \sum_{j_0,j_1,j_2} \int\int\int \bar{w}_1(|\gamma_0-\gamma_1|) w_2(|\gamma_1-\gamma_2|) \bar{w}_3(|\gamma_0-\gamma_1|) \bar{\phi}_{j_0}(\gamma_0) \bar{\phi}_{j_1}(\gamma_1) \bar{\phi}_{j_2}(\gamma_2) \begin{vmatrix} \theta_{j_0}(\gamma_0) & \theta_{j_0}(\gamma_1) & \theta_{j_0}(\gamma_2) \\ \theta_{j_1}(\gamma_0) & \theta_{j_1}(\gamma_1) & \theta_{j_1}(\gamma_2) \\ \theta_{j_2}(\gamma_0) & \theta_{j_2}(\gamma_1) & \theta_{j_2}(\gamma_2) \end{vmatrix} d\gamma_0 d\gamma_1 d\gamma_2 \,. \tag{2.16}$$

The only way two variables can occur is with the structure  , which has one case with multiplicity four, coming from whether m or n equals a and whether u or v equals a. This case, with multiplicity, contributes

$$\frac{|\mathbb{L}|}{N!}\sum_{j_0,j_1}\int\int \bar{w}_1(|\gamma_0-\gamma_1|)w_2(|\gamma_0-\gamma_1|)\bar{w}_3(|\gamma_0-\gamma_1|) \bar{\phi}_{j_0}(\gamma_0)\bar{\phi}_{j_1}(\gamma_1)\begin{vmatrix}\theta_{j_0}(\gamma_0) & \theta_{j_0}(\gamma_1)\\ \theta_{j_1}(\gamma_0) & \theta_{j_1}(\gamma_1)\end{vmatrix} d\gamma_0 d\gamma_1\,. \tag{2.17}$$

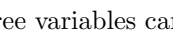
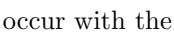
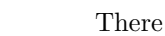
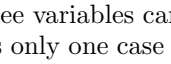
2.4. Antisymmetric inner product including $\mathcal{V}$. To compute $\langle\tilde{\psi},\mathcal{V}\psi\rangle_{\mathcal{A}}$ with ψ and $\tilde{\psi}$ in (2.1), we follow the same procedure as in section 2.3. To indicate the presence of V on a variable  we decorate it as  . When all indexes are distinct, we have the structure  and formula

$$\frac{1}{4}\frac{|\mathbb{L}|}{N!}\sum_{j_0,j_1,j_2,j_3,j_4}\int\int\int\int\int \bar{w}_1(|\gamma_0-\gamma_1|)w_2(|\gamma_2-\gamma_3|)V(\mathbf{r}_4)\bar{\phi}_{j_0}(\gamma_0)\bar{\phi}_{j_1}(\gamma_1)\bar{\phi}_{j_2}(\gamma_2) \bar{\phi}_{j_3}(\gamma_3)\bar{\phi}_{j_4}(\gamma_4)\begin{vmatrix}\theta_{j_0}(\gamma_0) & \theta_{j_0}(\gamma_1) & \theta_{j_0}(\gamma_2) & \theta_{j_0}(\gamma_3) & \theta_{j_0}(\gamma_4)\\ \theta_{j_1}(\gamma_0) & \theta_{j_1}(\gamma_1) & \theta_{j_1}(\gamma_2) & \theta_{j_1}(\gamma_3) & \theta_{j_1}(\gamma_4)\\ \theta_{j_2}(\gamma_0) & \theta_{j_2}(\gamma_1) & \theta_{j_2}(\gamma_2) & \theta_{j_2}(\gamma_3) & \theta_{j_2}(\gamma_4)\\ \theta_{j_3}(\gamma_0) & \theta_{j_3}(\gamma_1) & \theta_{j_3}(\gamma_2) & \theta_{j_3}(\gamma_3) & \theta_{j_3}(\gamma_4)\\ \theta_{j_4}(\gamma_0) & \theta_{j_4}(\gamma_1) & \theta_{j_4}(\gamma_2) & \theta_{j_4}(\gamma_3) & \theta_{j_4}(\gamma_4)\end{vmatrix} d\gamma_0 d\gamma_1 d\gamma_2 d\gamma_3 d\gamma_4\,. \tag{2.18}$$

Four variables can occur with the structures  and  . There are two cases for  , counted by whether V shares a variable with w_1 or w_2. Each case has multiplicity two; when V shares a variable with w_2, the multiplicity is from whether u equals m or n. The contribution for this case is obtained from (2.12) but replacing $(1/4)w_2(|\gamma_0-\gamma_1|)\bar{w}_3(|\gamma_2-\gamma_3|)$ by $(1/2)w_2(|\gamma_2-\gamma_3|)V(\mathbf{r}_3)$.

There is only one case for  . It has multiplicity four, from the ways in which a or b equals m or n. This case contributes

$$\frac{|\mathbb{L}|}{N!}\sum_{j_0,j_1,j_2,j_3}\int\int\int\int \bar{w}_1(|\gamma_0-\gamma_1|)w_2(|\gamma_1-\gamma_2|)V(\mathbf{r}_3)\bar{\phi}_{j_0}(\gamma_0)\bar{\phi}_{j_1}(\gamma_1)\bar{\phi}_{j_2}(\gamma_2)\bar{\phi}j_3(\gamma_3) \begin{vmatrix}\theta_{j_0}(\gamma_0) & \theta_{j_0}(\gamma_1) & \theta_{j_0}(\gamma_2) & \theta_{j_0}(\gamma_3)\\ \theta_{j_1}(\gamma_0) & \theta_{j_1}(\gamma_1) & \theta_{j_1}(\gamma_2) & \theta_{j_1}(\gamma_3)\\ \theta_{j_2}(\gamma_0) & \theta_{j_2}(\gamma_1) & \theta_{j_2}(\gamma_2) & \theta_{j_2}(\gamma_3)\\ \theta j_3(\gamma_0) & \theta j_3(\gamma_1) & \theta j_3(\gamma_2) & \theta j_3(\gamma_3)\end{vmatrix} d\gamma_0 d\gamma_1 d\gamma_2 d\gamma_3\,. \tag{2.19}$$

Three variables can occur with the structures  ,  , and  . There is only one case for  but two for  since V can share a variable with w_1 only or with w_2 only. Each of these three cases has multiplicity four from the ways a or b equals m or n. The contribution is the same as (2.15) but replacing $w_2(|\gamma_0-\gamma_1|)\bar{w}_3(|\gamma_1-\gamma_2|)$ by $w_2(|\gamma_1-\gamma_2|)$ and including $V(\mathbf{r}_0)$, $V(\mathbf{r}_1)$, or $V(\mathbf{r}_2)$.

There is only one case for  . It has multiplicity two from whether a or b equals m and contributes

$$\frac{1}{2}\frac{|\mathbb{L}|}{N!}\sum_{j_0,j_1,j_2}\int\int\int \bar{w}_1(|\gamma_0-\gamma_1|)w_2(|\gamma_0-\gamma_1|)V(\mathbf{r}_2)\bar{\phi}_{j_0}(\gamma_0)\bar{\phi}_{j_1}(\gamma_1)\bar{\phi}_{j_2}(\gamma_2) \left|\begin{array}{ccc}\theta_{j_0}(\gamma_0) & \theta_{j_0}(\gamma_1) & \theta_{j_0}(\gamma_2)\\ \theta_{j_1}(\gamma_0) & \theta_{j_1}(\gamma_1) & \theta_{j_1}(\gamma_2)\\ \theta_{j_2}(\gamma_0) & \theta_{j_2}(\gamma_1) & \theta_{j_2}(\gamma_2)\end{array}\right| d\gamma_0 d\gamma_1 d\gamma_2\,. \tag{2.20}$$

The only way two variables can occur is with the structure , which has one case. It has multiplicity four, with two coming from whether m or n equals a and two from whether u equals a or b. The contribution is the same as (2.17) but replacing $w_2(|\gamma_0-\gamma_1|)\bar{w}_3(|\gamma_0-\gamma_1|)$ by $w_2(|\gamma_0-\gamma_1|)V(\mathbf{r}_1)$.

2.5. Automatic generation and expansion of the cases. The Slater determinants in the formulas in sections 2.3 and 2.4 are nonseparable functions of two to six (three-dimensional) variables. Integrating with them directly costs up to M^6, which is prohibitively expensive. Our only choice is to expand out these determinants so that each term is separable so a proper choice for the order of summations and integrations allows us to use only low-dimensional functions. This expansion leads to many more terms than one can manage by hand, so we instead produce an algorithm to generate them.

In section 2.5.1 we present an algorithm to generate the 15 distinct structures and 25 total cases that we derived by hand in sections 2.3 and 2.4. In section 2.5.2 we present an algorithm to expand the determinants and combine duplicate terms to generate all the terms that we need to compute. In section 3 we study how to actually compute them and the cost to do so.

2.5.1. Generation of basic cases. The information needed to encode the basic cases for $\langle\tilde{\psi},\mathcal{W}\psi\rangle_{\mathcal{A}}$ is the variables that each geminal couples. Thus we can capture it by a list of three tuples as $[(i,j),(m,n),(u,v)]$. The geminal $\bar{w}_1$ always occurs as $\bar{w}_1(|\gamma_0-\gamma_1|)$ so we encode its presence with the tuple $(0,1)$ in the first position. The next index m can be a previously used index $\{0,1\}$ or the next available index 2. Its partner n can be a previously used index $\{0,1,m\}$ or the next available index, but it cannot be m. Thus the possibilities for (m,n) are $(0,1)$, $(0,2)$, $(1,0)$, $(1,2)$, $(2,0)$, $(2,1)$, or $(2,3)$. Similarly, u can be any of $\{0,1,m,n\}$ or the next available and v can be any of $\{0,1,m,n,u\}$ or the next available, but it cannot be u. The basic cases for $\langle\tilde{\psi},\mathcal{V}\psi\rangle_{\mathcal{A}}$ are encoded by two couples and a single as $[(i,j),(m,n),(u,)]$; they are generated the same way, simply neglecting v.

Some of the cases generated in this way are equivalent and should be combined and their multiplicity noted. To detect equivalent cases we put each case in a canonical form and then combine duplicates. The allowed operations are to sort a tuple (since $(3,2)$ and $(2,3)$ are equivalent) and to exchange any two numbers (since $[(0,1),(0,2)]$ and $[(1,0),(1,2)]$ are equivalent). The specification for the canonical form is as follows:

- Each tuple is sorted small to large.
- If $0\in\{m,n\}$ and $1\notin\{m,n\}$, then exchange 0 and 1.
- If $(m,n)\in\{(0,1),(2,3)\}$, $0\in\{u,v\}$, and $1\notin\{u,v\}$, then exchange 0 and 1.
- If $(m,n)=(2,3)$, $2\in\{u,v\}$, and $3\notin\{u,v\}$, then exchange 2 and 3.

This algorithm reproduces the results of sections 2.3 and 2.4, which we summarize in the first three columns of Tables 2.1 and 2.2.

TABLE 2.1
Summary of the structures and cases from $\langle\tilde{\psi}, \mathcal{W}\psi\rangle_{\mathcal{A}}$.

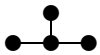
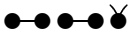
Structure	Base cases #	Mult	Expanded Symmetry group	#
	1	1	{(012345), (012354), (013245), (013254), (102345), (102354), (103245), (103254)}	120
	3	4	{(01234), (01243)}	198
	3	2	{(0123), (0132), (1023), (1032)}	30
	3	8	{(0123)}	72
	1	8	{(0123)}	24
	3	8	{(012)}	18
	1	8	{(012)}	6
	1	4	{(01), (10)}	2
Total: 8	16			470

TABLE 2.2
Summary of the structures and cases from $\langle\tilde{\psi}, \mathcal{V}\psi\rangle_{\mathcal{A}}$.

Structure	Base cases #	Mult	Expanded Symmetry group	#
	1	1	{(01234), (01324), (10234), (10324)}	38
	2	2	{(0123), (1023)}	28
	1	4	{(0123)}	24
	1	4	{(012)}	6
	2	4	{(012)}	12
	1	2	{(012), (102)}	4
	1	4	{(01)}	2
Total: 7	9			114

2.5.2. Expansion into all terms. The entries in our Slater determinants consist of the function θ with index given by its row and variable given by its column (both counting from 0). Thus the row 0 column 2 entry is $\theta_{j_0}(\gamma_2)$. To generate terms in the expansion, we fill a matrix of the desired size with tuples (a, b) giving that entry's position. We then symbolically expand the determinant by column expansion to obtain a sum (stored as a list) of products (stored as lists) with signs. For example, the 2×2 determinant yields $[(0,0),(1,1)]$ with sign 1 and $[(0,1),(1,0)]$ with sign -1.

Some of these summands are equivalent and should be combined. Two summands are equivalent if the symmetry group of the case contains a permutation that maps one summand to the other. In the fourth column of Tables 2.1 and 2.2 we show the symmetry group of each structure. In the fifth column we show the number of terms remaining after combining equivalent terms. In total there are 584 terms to compute.

2.6. Antisymmetric inner products including $\mathcal{G}_\mu$. For the main iteration in [4], we need to compute $\langle\tilde{\psi}, \mathcal{G}_\mu(\mathcal{V}+\mathcal{W})\psi\rangle_{\mathcal{A}}$ with ψ and $\tilde{\psi}$ in (2.1). Since $\mathcal{G}_\mu$ is self-adjoint, this is equal to $\langle\mathcal{G}_\mu\tilde{\psi}, (\mathcal{V}+\mathcal{W})\psi\rangle_{\mathcal{A}}$, which we will compute instead. In [4], we gave a construction that, for fixed μ and accuracy, approximates the Green function as a sum of separable operators as

$$\mathcal{G}_\mu = (\mathcal{T} - \mu\mathcal{I})^{-1} \approx \sum_{l=1}^{L} \bigotimes_{j=1}^{N} \mathcal{F}_j^l \,. \tag{2.21}$$

The index l determines the amplitude and exponent of a Gaussian, and $\mathcal{F}_j^l$ is the operator that convolves with this Gaussian in the variable $\mathbf{r}_j$. Since the inner product is linear, we can insert this approximation, pull out the sum, and consider the computation for a single index l, which we can then suppress. Thus we must compute

$$\left\langle \mathcal{A} \bigotimes_{j=1}^{N} \mathcal{F}_j \tilde{\psi}, (\mathcal{V} + \mathcal{W})\psi \right\rangle . \tag{2.22}$$

We can formally write

$$\mathcal{A} \bigotimes_{j=1}^{N} \mathcal{F}_j \tilde{\psi} = \mathcal{A} \left(\bigotimes_{j=1}^{d} \mathcal{F}_j \right) \left(\frac{1}{2} \sum_{m \neq n} w_2(|\gamma_m - \gamma_n|) \right) \left(\prod_{i=1}^{N} \tilde{\phi}_i(\gamma_i) \right) \tag{2.23}$$

$$= \left(\frac{1}{2} \sum_{m \neq n} \mathcal{F}_m \mathcal{F}_n w_2(|\gamma_m - \gamma_n|) \mathcal{F}_m^{-1} \mathcal{F}_n^{-1} \right) \mathcal{A} \prod_{i=1}^{N} \left(\mathcal{F} \tilde{\phi}_i(\gamma_i) \right) . \tag{2.24}$$

Structurally, we have an "operator geminal" and a modified separable function with $\hat{\phi}_i = \mathcal{F}\tilde{\phi}_i$.

We can then use formulas that we have already developed, with a few modifications. The matrix $\mathbb{L}$ in (2.5) is computed using $\hat{\phi}_i$ instead of $\tilde{\phi}_i$, and we set $\mathbf{\Theta} = \mathbb{L}^{-1}\hat{\mathbf{\Phi}}$. The portions of the operator geminal with $\mathcal{F}^{-1}$ are only applied to the functions θ; since $\mathcal{F}^{-1}\mathbf{\Theta} = \mathcal{F}^{-1}\mathbb{L}^{-1}\mathcal{F}\tilde{\mathbf{\Phi}} = \mathbb{L}^{-1}\tilde{\mathbf{\Phi}}$, $\mathcal{F}^{-1}$ can always be formally canceled and not computed. Some care must be taken that the operators are applied to the correct functions. For example, for (2.13) ●–●–●–● we have

$$\frac{|\mathbb{L}|}{N!} \sum_{j_0,j_1,j_2,j_3} \int\int\int\int \bar{w}_1(|\gamma_0 - \gamma_1|) \bar{w}_3(|\gamma_2 - \gamma_3|) \bar{\phi}_{j_0}(\gamma_0) \bar{\phi}_{j_1}(\gamma_1) \bar{\phi}_{j_2}(\gamma_2) \bar{\phi}_{j_3}(\gamma_3)$$
$$\mathcal{F}_1 \mathcal{F}_2 \left(w_2(|\gamma_1 - \gamma_2|) \begin{vmatrix} \theta_{j_0}(\gamma_0) & \mathcal{F}^{-1}\theta_{j_0}(\gamma_1) & \mathcal{F}^{-1}\theta_{j_0}(\gamma_2) & \theta_{j_0}(\gamma_3) \\ \theta_{j_1}(\gamma_0) & \mathcal{F}^{-1}\theta_{j_1}(\gamma_1) & \mathcal{F}^{-1}\theta_{j_1}(\gamma_2) & \theta_{j_1}(\gamma_3) \\ \theta_{j_2}(\gamma_0) & \mathcal{F}^{-1}\theta_{j_2}(\gamma_1) & \mathcal{F}^{-1}\theta_{j_2}(\gamma_2) & \theta_{j_2}(\gamma_3) \\ \theta_{j_3}(\gamma_0) & \mathcal{F}^{-1}\theta_{j_3}(\gamma_1) & \mathcal{F}^{-1}\theta_{j_3}(\gamma_2) & \theta_{j_3}(\gamma_3) \end{vmatrix} \right) d\gamma_0 d\gamma_1 d\gamma_2 d\gamma_3 . \tag{2.25}$$

2.7. Plain antisymmetric inner products. The analysis of the inner product $\langle \tilde{\psi}, \psi \rangle_{\mathcal{A}}$ is a simpler version of the analysis for $\langle \tilde{\psi}, \mathcal{W}\psi \rangle_{\mathcal{A}}$, so we omit the details. In Table 2.3 we summarize the cases that result.

TABLE 2.3
Summary of the structures and cases from $\langle \tilde{\psi}, \psi \rangle_{\mathcal{A}}$.

	Base cases		Expanded	
Structure	#	Mult	Symmetry group	#
●–●–●–●	1	1	{(0123),(0132),(1023),(1032)}	10
●–●–● ●	1	4	{(012)}	6
●–● ●–●	1	2	{(01),(10)}	2
Total: 3	3			18

3. Computational techniques and costs. In this section we consider how to actually compute the 584 terms that we generated in section 2.5.2 and so compute the inner products $\langle\tilde{\psi}, \mathcal{W}\psi\rangle_{\mathcal{A}}$ and $\langle\tilde{\psi}, \mathcal{V}\psi\rangle_{\mathcal{A}}$. Our goals in this analysis are to provide an algorithm for determining the best way to compute each term, assess the total cost of all the computations, provide insight into the most costly terms, archive the formulas so that the skeptical reader can check our claims, and avoid numbing the reader's mind with too many cases or too much detail.

In section 3.1 we consider one term in detail and show how to compute it and the resulting cost. Through this example we introduce a graphical way to present the essential structure of a term. We also demonstrate our first principle for organizing the computation, which is to note common structures in the many terms and perform computations that reduce the term to a simpler term. In the graphical representation this results in the contraction of a graph to a simpler graph.

In section 3.2 we collect the principles from section 3.1 into an algorithm. Applied to our 584 terms, this algorithm reduces the problem to 21 residual graphs. We then introduce one additional element to our graphical representation and an additional reduction that uses it. Incorporating this reduction into the algorithm reduces the number of residual graphs to 14. We list the complexities of these residual graphs here but defer to section 3.4 the somewhat tedious propositions for computing them.

Many of the component computations in various terms are identical and can be cached and reused. In section 3.3 we study how to maximize this reuse and present counts on the number of calculations of each complexity.

3.1. An example computation using reductions. In this section we consider in detail how to compute one term in order to illustrate the general procedure.

We choose to take a term from (2.11), which has the graphical representation  and is generated in section 2.5.1 by the list of tuples $[(0,1),(1,2),(3,4)]$. We choose the term with formula

(3.1)
$$\frac{-1}{2}\frac{|\mathbb{L}|}{N!}\sum_{j_0,j_1,j_2,j_3,j_4}\int\int\int\int\int \bar{w}_1(|\gamma_0-\gamma_1|)w_2(|\gamma_1-\gamma_2|)\bar{w}_3(|\gamma_3-\gamma_4|)\bar{\phi}_{j_0}(\gamma_0)\bar{\phi}_{j_1}(\gamma_1)$$
$$\bar{\phi}_{j_2}(\gamma_2)\bar{\phi}_{j_3}(\gamma_3)\bar{\phi}_{j_4}(\gamma_4)\theta_{j_0}(\gamma_3)\theta_{j_1}(\gamma_0)\theta_{j_2}(\gamma_1)\theta_{j_3}(\gamma_2)\theta_{j_4}(\gamma_4)d\gamma_0 d\gamma_1 d\gamma_2 d\gamma_3 d\gamma_4\,,$$

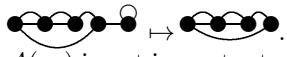
which is generated in the determinant expansion in section 2.5.2 by the list of tuples $[(0,1),(1,2),(2,3),(3,0),(4,4)]$ and has sign -1. To graphically represent this term in the expansion, we use these tuples to define curves between variables and obtain  . Each curve is interpreted as an index that connects the variables at its ends. For example, $(0,1)$ means that $\theta_{j_1}(\gamma_0)$ is present; combining with $\bar{\phi}_{j_1}(\gamma_1)$ yields $\bar{\phi}_{j_1}(\gamma_1)\theta_{j_1}(\gamma_0)$, which has the index j_1 connecting the variables γ_0 and γ_1.

To compute this term we notice that the only terms involving j_4 are $\bar{\phi}_{j_4}(\gamma_4)$ and $\theta_{j_4}(\gamma_4)$ and these involve only the variable γ_4, so we can sum over j_4 with cost NM. We then notice that the only other term involving γ_4 is $\bar{w}_3(|\gamma_3-\gamma_4|)$, so we can then integrate over γ_4 with cost M_*. Thus with cost $NM+M_*$ we compute

$$A(\gamma_3)=\int \bar{w}_3(|\gamma_3-\gamma_4|)\left(\sum_{j_4}\bar{\phi}_{j_4}(\gamma_4)\theta_{j_4}(\gamma_4)\right)d\gamma_4 \tag{3.2}$$

and reduce (3.1) to

$$\frac{-1}{2}\frac{|\mathbb{L}|}{N!}\sum_{j_0,j_1,j_2,j_3}\int\int\int\int \bar{w}_1(|\gamma_0-\gamma_1|)w_2(|\gamma_1-\gamma_2|) \tag{3.3}$$
$$\bar{\phi}_{j_0}(\gamma_0)\bar{\phi}_{j_1}(\gamma_1)\bar{\phi}_{j_2}(\gamma_2)\bar{\phi}_{j_3}(\gamma_3)\theta_{j_0}(\gamma_3)\theta_{j_1}(\gamma_0)\theta_{j_2}(\gamma_1)\theta_{j_3}(\gamma_2)A(\gamma_3)d\gamma_0 d\gamma_1 d\gamma_2 d\gamma_3\,.$$

Graphically we have performed the contraction  . Our new graph does not indicate the presence of $A(\gamma_3)$, but $A(\gamma_3)$ is not important structurally since it can be combined with $\bar{\phi}_{j_3}(\gamma_3)$. This contraction can be stated as a general principle, using  to represent the uninvolved portion of the graph.

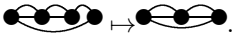
PROPOSITION 3.1. *We can contract* *with cost* $NM + M_*$.

We now notice that the only terms involving γ_3 are $\bar{\phi}_{j_3}(\gamma_3)$, $\theta_{j_0}(\gamma_3)$, and $A(\gamma_3)$. Since these involve the two indexes j_0 and j_3 we can integrate over γ_3 with cost N^2M and obtain a scalar with the two indexes j_0 and j_3. We then notice that the only other term with index j_3 is $\theta_{j_3}(\gamma_2)$, so we can sum over j_3 with cost N^2M and obtain a function of γ_2 with index j_0. Thus with cost $2N^2M$ we compute

$$B_{j_0}(\gamma_2) = \sum_{j_3}\theta_{j_3}(\gamma_2)\left(\int \bar{\phi}_{j_3}(\gamma_3)\theta_{j_0}(\gamma_3)A(\gamma_3)d\gamma_3\right) \tag{3.4}$$

and reduce (3.3) to

$$\frac{-1}{2}\frac{|\mathbb{L}|}{N!}\sum_{j_0,j_1,j_2}\int\int\int \bar{w}_1(|\gamma_0-\gamma_1|)w_2(|\gamma_1-\gamma_2|) \tag{3.5}$$
$$\bar{\phi}_{j_0}(\gamma_0)\bar{\phi}_{j_1}(\gamma_1)\bar{\phi}_{j_2}(\gamma_2)\theta_{j_1}(\gamma_0)\theta_{j_2}(\gamma_1)B_{j_0}(\gamma_2)d\gamma_0 d\gamma_1 d\gamma_2\,.$$

Graphically we have performed the contraction  . In our new graph $B_{j_0}(\gamma_2)$ plays the role of $\theta_{j_0}(\gamma_2)$. This contraction can be stated as a general principle.

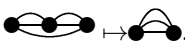
PROPOSITION 3.2. *We can contract*  *with cost* $2N^2M$.

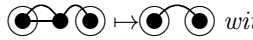
Next we notice that the only terms involving γ_2 are $w_2(|\gamma_1-\gamma_2|)$, $\bar{\phi}_{j_2}(\gamma_2)$, and $B_{j_0}(\gamma_2)$. Since these involve the indexes j_0 and j_2, we can integrate over γ_2 with cost N^2M_* and obtain a function of γ_1 with indexes j_0 and j_2. We then notice that the only other term that involves j_2 is $\theta_{j_2}(\gamma_1)$ and it involves only the same variable γ_1 as our new term. Thus we can sum over j_2 with cost N^2M and obtain a function of γ_1 with index j_0. Thus with cost $N^2M_* + N^2M$ we compute

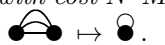
$$C_{j_0}(\gamma_1) = \sum_{j_2}\theta_{j_2}(\gamma_1)\left(\int w_2(|\gamma_1-\gamma_2|)\bar{\phi}_{j_2}(\gamma_2)B_{j_0}(\gamma_2)d\gamma_2\right) \tag{3.6}$$

and reduce (3.5) to

$$\frac{-1}{2}\frac{|\mathbb{L}|}{N!}\sum_{j_0,j_1}\int\int \bar{w}_1(|\gamma_0-\gamma_1|)\bar{\phi}_{j_0}(\gamma_0)\bar{\phi}_{j_1}(\gamma_1)\theta_{j_1}(\gamma_0)C_{j_0}(\gamma_1)d\gamma_0 d\gamma_1\,. \tag{3.7}$$

Graphically we have performed the contraction  . This contraction can be stated as a general principle.

PROPOSITION 3.3. *We can contract*  *with cost* $N^2M_* + N^2M$.

We can now apply Proposition 3.3 again to contract  . With cost $N^2M_* + N^2M$ we compute

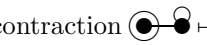
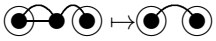
$$D_{j_0}(\gamma_0) = \sum_{j_1} \theta_{j_1}(\gamma_0) \left(\int \bar{w}_1(|\gamma_0 - \gamma_1|) \bar{\phi}_{j_1}(\gamma_1) C_{j_0}(\gamma_1) d\gamma_1 \right) \tag{3.8}$$

and reduce (3.7) to

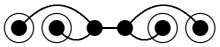
$$\frac{-1}{2} \frac{|\mathbb{L}|}{N!} \sum_{j_0} \int \bar{\phi}_{j_0}(\gamma_0) D_{j_0}(\gamma_0) d\gamma_0 \,. \tag{3.9}$$

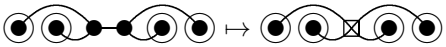
We can then integrate over γ_0 with cost NM and then sum over j_0 with cost N to complete our calculation. In total, our cost for this term is $2N^2M_* + M_* + 4N^2M + 2NM + N$.

3.2. Reduction procedure and the residual terms. Based on our experience for the example in section 3.1, our first algorithm to compute each of our 584 terms is as follows:

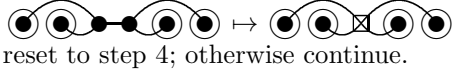
1. Attempt to apply the contraction  in Proposition 3.1. If it applies, then reset to step 1; otherwise continue.
2. Attempt to apply the contraction  in Proposition 3.2. If it applies, then reset to step 1; otherwise continue.
3. Attempt to apply the contraction  in Proposition 3.3. If it applies, then reset to step 1; otherwise continue.

This algorithm does not complete the computation of any term but instead reduces each term to one of 21 irreducible cases, which are shown in the first columns of Tables 3.1 and 3.2.

Eleven of these irreducible cases have the common element  . Integrating over one of these variables costs N^2M_* and yields a function of the other variable with two indexes. Integrating over the other variable then costs N^4M and yields a scalar with four indexes. To indicate a scalar in our graphical language, we introduce the element  . We thus have the following contraction proposition.

PROPOSITION 3.4. *We can contract*  *with cost* $N^4M + N^2M_*$.

We can then add another step to our algorithm:

4. Attempt to apply the contraction  in Proposition 3.4. If it applies, then reset to step 4; otherwise continue.

This additional step reduces 11 cases to 4, thus leaving us with 14 cases. One could perhaps reduce to fewer cases by introducing additional contractions and graphical elements, but it does not appear worthwhile to do so.

In Table 3.1 we list the the 10 cases that do not use this additional contraction, their computational complexity, and the proposition in section 3.4 that demonstrates this complexity. In Table 3.2 we list the 11 cases that do use this additional contraction, the 4 cases to which they reduce, their complexity, and the proposition used to compute them. The listed complexities do not include the cost of any reductions used to obtain the residual cases.

TABLE 3.1
End cases without scalar element and their costs.

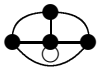
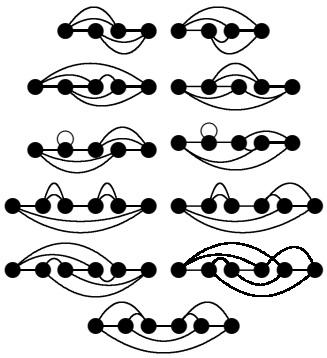
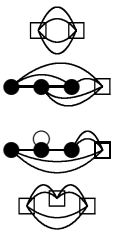
Graph	Cost	Proposition
	$\mathcal{O}(NM)$	3.5
	$\mathcal{O}(N^2M_*)$	3.6
	$\mathcal{O}(MM_*)$	3.7
	$\mathcal{O}(MM_* + NM^2)$	3.8
	$\mathcal{O}(N^2MM_*)$	3.9
	$\mathcal{O}(N^2M_*)$	3.10
	$\mathcal{O}(N^4M_*)$	3.11
	$\mathcal{O}(N^4M_*)$	3.12
	$\mathcal{O}(N^4M_*)$	3.13
	$\mathcal{O}(N^3M + N^2M_*)$	3.14

TABLE 3.2
End cases with scalar element and their costs.

Graphs	After 3.4	Cost	Proposition
		$\mathcal{O}(N^4)$	3.15
		$\mathcal{O}(N^4M + N^2M_*)$	3.16
		$\mathcal{O}(N^4M + N^2M_*)$	3.17
		$\mathcal{O}(N^6)$	3.18

3.3. Maximizing reuse of computations. To analyze the algorithm further, we have implemented a "dry run" version. All functions are empty objects that only know how many indexes and variables they have. These are then wrapped by objects that know which indexes and variable these are. A list of these wrapper objects then defines a term. For example, the integrand $\bar{w}_1(|\gamma_0 - \gamma_1|)\bar{\phi}_{j_0}(\gamma_0)\bar{\phi}_{j_1}(\gamma_1)\theta_{j_1}(\gamma_0)C_{j_0}(\gamma_1)$ from (3.7) is stored as

$$[(\bar{w}_1(0,2,\text{“convolution”}),(),(0,1)),(\bar{\phi}(1,1),(0,),(0,)),(\bar{\phi}(1,1),(1,),(1,)),(\theta(1,1),(1,),(0,)),(C(1,1),(0,),(1,))]\,. \tag{3.10}$$

TABLE 3.3
Total operations sorted by complexity. The entry gives the number of operations of the complexity given by the product of the row and column headings.

	1	M	M_*	M^2	MM_*
1		6	496	6	4
N	716	1097		10	
N^2	193	1358	1054	2	2
N^3	132	14			
N^4	130	260	12		
N^5	12				
N^6	12				

When we sum over an index or integrate over a variable we remove involved terms from the list and introduce a new term to represent the result. We keep track of the cost of each operation and present the total cost for all 584 case in Table 3.3. There are 2758 sums and 2758 integrations for 5516 total operations. (The initialization cost $2N^2M + \mathcal{O}(N^3)$ to construct $\mathbb{L}$, $\mathbb{L}^{-1}$, and $\mathbf{\Theta}$ in section 2.2 is not included.)

Many of these operations are duplicates. For example, the computation (3.2) will occur both in  and in  . By caching and reusing the results we can reduce our overall computational cost. The analysis is complicated by various choices that do not affect the cost of individual terms but may affect the sharing of computations between terms. Specifically, in the algorithm in section 3.2 we may be able to apply Propositions 3.1, 3.2, and 3.3 in a different default order or apply one of them to more than one place in the graph. In the propositions in section 3.4 we give one order of operations to achieve the claimed cost, but other orders are sometimes possible.

Finding the optimal choices to minimize the total cost appears to have combinatorial complexity, so we use the following heuristic:

1. By inspection, identify 50 common computations and include them in the cache.
2. For each term,
   (a) generate many possible ways to compute each terms while keeping the cost unchanged;
   (b) intersect these ways to see what computations are unavoidable and include these in the cache.
3. Loop through the terms (arbitrary order):
   (a) Identify which of the possible ways to compute it requires the fewest computations not already in the cache (breaking ties arbitrarily).
   (b) Include these in the cache.

We verified that the resulting cache is minimal in the sense that removing any element from it prevents at least one term from being fully computed. With this procedure we obtain 967 sums and 575 integrations for 1542 total operations. Table 1.1 in section 1.2 shows how many computations we have of each complexity. By saving the results of this optimized dry run, we obtain autogenerated code to use for real runs.

3.4. Propositions. In this section we collect propositions to compute the residual cases from section 3.2. For each case we treat one instance in which it arises. Permutations of $\{w_1, w_2, w_3\}$ and the presence of V or factors created by earlier contractions lead to additional instances, but they are all treated the same way. We neglect overall constants such as $|\mathbb{L}|$. These propositions consist simply in specifying

the order of operations and counting the cost of each sum and integral. To specify the order, we simply arrange the terms and use large parentheses. We indicate the cost of a summation or integration by putting it in brackets (e.g., $[N^2 M_*]$) as the upper limit on the $\sum$ or $\int$ symbol. All sums actually go to N and all integrals are over all space and the two spin values.

PROPOSITION 3.5. *We can compute*  *with cost* $NM + N$ *via the ordering*

$$\sum_{j_0}^{[N]} \left(\int^{[NM]} \bar{\phi}_{j_0}(\gamma_0) \theta_{j_0}(\gamma_0) d\gamma_0 \right) . \tag{3.11}$$

PROPOSITION 3.6. *We can compute*  *with cost* $2N^2 M_* + N^2 M + NM + N^2 + N$ *via the ordering*

$$\sum_{j_0, j_2}^{[N^2]} \left(\int^{[N^2 M]} \left(\int^{[N^2 M_*]} \bar{w}_1(|\gamma_0 - \gamma_1|) \bar{\phi}_{j_0}(\gamma_0) \theta_{j_2}(\gamma_0) d\gamma \right) \left(\sum_{j_1}^{[NM]} \bar{\phi}_{j_1}(\gamma_1) \theta_{j_1}(\gamma_1) \right) \left(\int^{[N^2 M_*]} w_2(|\gamma_1 - \gamma_2|) \bar{\phi}_{j_2}(\gamma_2) \theta_{j_0}(\gamma_2) d\gamma_2 \right) d\gamma_1 \right) . \tag{3.12}$$

PROPOSITION 3.7. *We can compute*  *with cost* $MM_* + M^2 + 3NM + M$ *via the ordering*

$$\int^{[M]} \left(\int^{[M^2]} w_2(|\gamma_1 - \gamma_2|) \left(\sum_{j_2}^{[NM]} \bar{\phi}_{j_2}(\gamma_2) \theta_{j_2}(\gamma_2) \right) \left(\sum_{j_1}^{[NM]} \bar{\phi}_{j_1}(\gamma_1) \theta_{j_1}(\gamma_1) \right) \left(\int^{[MM_*]} \left(\sum_{j_0}^{[NM]} \bar{\phi}_{j_0}(\gamma_0) \theta_{j_0}(\gamma_0) \right) \bar{w}_1(|\gamma_0 - \gamma_1|) w_3(|\gamma_2 - \gamma_0|) d\gamma_0 \right) d\gamma_1 \right) d\gamma_2 . \tag{3.13}$$

The cost $\mathcal{O}(MM_*)$ *comes from noting that we can consider* $\int A(\gamma_0) \bar{w}_1(|\gamma_0 - \gamma_1|) w_3(|\gamma_2 - \gamma_0|) d\gamma_0$ *as performing a convolution* $\gamma_0 \to \gamma_1$ *for each fixed value of* γ_2.

PROPOSITION 3.8. *We can compute*  *with cost* $MM_* + 2NM^2 + M^2 + NM + M$ *via the ordering*

$$\int^{[M]} \left(\int^{[M^2]} w_2(|\gamma_1 - \gamma_2|) \left(\sum_{j_2}^{[NM^2]} \bar{\phi}_{j_2}(\gamma_2) \theta_{j_2}(\gamma_1) \right) \left(\sum_{j_1}^{[NM^2]} \bar{\phi}_{j_1}(\gamma_1) \theta_{j_1}(\gamma_2) \right) \left(\int^{[MM_*]} \left(\sum_{j_0}^{[NM]} \bar{\phi}_{j_0}(\gamma_0) \theta_{j_0}(\gamma_0) \right) \bar{w}_1(|\gamma_0 - \gamma_1|) w_3(|\gamma_2 - \gamma_0|) d\gamma_0 \right) d\gamma_1 \right) d\gamma_2 . \tag{3.14}$$

PROPOSITION 3.9. *We can compute*  *with cost* $N^2MM_* + N^2M^2 + 2NM^2 + M^2 + M$ *via the ordering*

$$\int^{[M]} \left(\int^{[M^2]} \left[\sum_{j_1}^{[NM^2]} \bar{\phi}_{j_1}(\gamma_1) \left(\sum_{j_0}^{[NM^2]} \bar{\phi}_{j_0}(\gamma_0)\theta_{j_0}(\gamma_1)\bar{w}_1(|\gamma_0-\gamma_1|) \right) \right.\right. \left.\left. \left(\sum_{j_2}^{[N^2M^2]} \theta_{j_2}(\gamma_0) \left(\int^{[N^2MM_*]} w_2(|\gamma_1-\gamma_2|)w_3(|\gamma_2-\gamma_0|)\bar{\phi}_{j_2}(\gamma_2)\theta_{j_1}(\gamma_2)d\gamma_2 \right) \right) \right] d\gamma_0 \right) d\gamma_1 \,. \tag{3.15}$$

PROPOSITION 3.10. *We can compute*  *with cost* $3N^2M_* + N^2M + 2NM + N^2 + N$ *via the ordering*

$$\sum_{j_0,j_3}^{[N^2]} \left(\int^{[N^2M]} \left(\int^{[N^2M_*]} \bar{w}_1(|\gamma_0-\gamma_1|)\bar{\phi}_{j_0}(\gamma_0)\theta_{j_3}(\gamma_0)d\gamma_0 \right) \left(\sum_{j_1}^{[NM]} \bar{\phi}_{j_1}(\gamma_1)\theta_{j_1}(\gamma_1) \right) \left[\int^{[N^2M_*]} \left(\int^{[N^2M_*]} w_3(|\gamma_2-\gamma_3|)\bar{\phi}_{j_3}(\gamma_3)\theta_{j_0}(\gamma_3)d\gamma_3 \right) \left(\sum_{j_2}^{[NM]} \bar{\phi}_{j_2}(\gamma_2)\theta_{j_2}(\gamma_2) \right) w_2(|\gamma_1-\gamma_2|)d\gamma_2 \right] d\gamma_1 \right) . \tag{3.16}$$

PROPOSITION 3.11. *We can compute*  *with cost* $N^4M_* + 2N^2M_* + N^4M + N^4 + N^3 + N^2 + N$ *via the ordering*

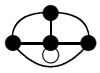
$$\sum_{j_0,j_1,j_2,j_3}^{[N^4]} \left(\int^{[N^4M]} \bar{\phi}_{j_1}(\gamma_1)\theta_{j_2}(\gamma_1) \left(\int^{[N^2M_*]} \bar{w}_1(|\gamma_0-\gamma_1|)\bar{\phi}_{j_0}(\gamma_0)\theta_{j_3}(\gamma_0)d\gamma_0 \right) \left[\int^{[N^4M_*]} \left(\int^{[N^2M_*]} w_3(|\gamma_2-\gamma_3|)\bar{\phi}_{j_3}(\gamma_3)\theta_{j_0}(\gamma_3)d\gamma_3 \right) w_2(|\gamma_1-\gamma_2|)\bar{\phi}_{j_2}(\gamma_2)\theta_{j_1}(\gamma_2)d\gamma_2 \right] d\gamma_1 \right) . \tag{3.17}$$

PROPOSITION 3.12. *We can compute*  *with cost* $N^4M_* + 2N^2M_* + N^4M + N^4 + N^3 + N^2 + N$ *via the ordering*

$$\sum_{j_0,j_1,j_2,j_3}^{[N^4]} \left(\int^{[N^4M]} \bar{\phi}_{j_1}(\gamma_1)\theta_{j_3}(\gamma_1) \left(\int^{[N^2M_*]} \bar{w}_1(|\gamma_0-\gamma_1|)\bar{\phi}_{j_0}(\gamma_0)\theta_{j_2}(\gamma_0)d\gamma_0 \right) \left[\int^{[N^4M_*]} \left(\int^{[N^2M_*]} w_3(|\gamma_2-\gamma_3|)\bar{\phi}_{j_3}(\gamma_3)\theta_{j_1}(\gamma_3)d\gamma_3 \right) w_2(|\gamma_1-\gamma_2|)\bar{\phi}_{j_2}(\gamma_2)\theta_{j_0}(\gamma_2)d\gamma_2 \right] d\gamma_1 \right) . \tag{3.18}$$

PROPOSITION 3.13. *We can compute*  *with cost* $N^4 M_* + 2N^2 M_* + N^4 M + N^4 + N^3 + N^2 + N$ *via the ordering*

$$\sum_{j_0,j_1,j_2,j_3}^{[N^4]} \left(\int^{[N^4 M]} \bar{\phi}_{j_1}(\gamma_1)\theta_{j_2}(\gamma_1) \left(\int^{[N^2 M_*]} \bar{w}_1(|\gamma_0 - \gamma_1|)\bar{\phi}_{j_0}(\gamma_0)\theta_{j_3}(\gamma_0) d\gamma_0 \right) \right. \\ \left[\int^{[N^4 M_*]} \left(\int^{[N^2 M_*]} w_3(|\gamma_2 - \gamma_3|)\bar{\phi}_{j_3}(\gamma_3)\theta_{j_1}(\gamma_3) d\gamma_3 \right) \right. \\ \left. \left. w_2(|\gamma_1 - \gamma_2|)\bar{\phi}_{j_2}(\gamma_2)\theta_{j_0}(\gamma_2) d\gamma_2 \right] d\gamma_1 \right) . \tag{3.19}$$

PROPOSITION 3.14. *We can compute*  *with cost* $3N^2 M_* + N^3 M + NM + N^3 + N^2 + N$ *via the ordering*

$$\sum_{j_1,j_2,j_3}^{[N^3]} \left(\int^{[N^3 M]} \left(\sum_{j_0}^{[NM]} \bar{\phi}_{j_0}(\gamma_0)\theta_{j_0}(\gamma_0) \right) \right. \\ \left(\int^{[N^2 M_*]} \bar{w}_1(|\gamma_0 - \gamma_1|)\bar{\phi}_{j_1}(\gamma_1)\theta_{j_3}(\gamma_1) d\gamma_1 \right) \left(\int^{[N^2 M_*]} w_2(|\gamma_1 - \gamma_2|)\bar{\phi}_{j_2}(\gamma_2)\theta_{j_1}(\gamma_2) d\gamma_2 \right) \\ \left. \left(\int^{[N^2 M_*]} w_3(|\gamma_1 - \gamma_3|)\bar{\phi}_{j_3}(\gamma_3)\theta_{j_2}(\gamma_3) d\gamma_3 \right) d\gamma_0 \right) . \tag{3.20}$$

PROPOSITION 3.15. *We can compute*  *with cost* $N^4 + N^3 + N^2 + N$ *via the ordering*

$$\sum_{j_0,j_1,j_2,j_3}^{[N^4]} A_{j_0 j_1 j_2 d} B_{j_0 j_1 j_2 j_3} . \tag{3.21}$$

PROPOSITION 3.16. *We can compute*  *with cost* $2N^2 M_* + N^4 M + N^3 M + N^4 + N^3 + N^2 + N$ *via the ordering*

$$\sum_{j_0,j_1,j_3,j_4}^{[N^4]} \left(\int^{[N^4 M]} \left[\sum_{j_2}^{[N^3 M]} \left(\int^{[N^2 M_*]} \bar{w}_1(|\gamma_0 - \gamma_1|)\bar{\phi}_{j_0}(\gamma_0)\theta_{j_2}(\gamma_0) d\gamma_0 \right) \right. \right. \\ \left. \left. \left(\int^{[N^2 M_*]} w_2(|\gamma_1 - \gamma_2|)\bar{\phi}_{j_2}(\gamma_2)\theta_{j_3}(\gamma_2) d\gamma_2 \right) \right] \bar{\phi}_{j_1}(\gamma_1)\theta_{j_4}(\gamma_1) d\gamma_1 \right) A_{j_0 j_1 j_3 j_4} . \tag{3.22}$$

PROPOSITION 3.17. *We can compute*  *with cost* $2N^2M_* + N^4M + NM + N^4 + N^3 + N^2 + N$ *via the ordering*

$$\sum_{j_0,j_2,j_3,j_4}^{[N^4]} \left[\int^{[N^4M]} \left(\int^{[N^2M_*]} \bar{w}_1(|\gamma_0 - \gamma_1|)\bar{\phi}_{j_0}(\gamma_0)\theta_{j_3}(\gamma_0)d\gamma_0 \right) \left(\sum_{j_1}^{[NM]} \bar{\phi}_{j_1}(\gamma_1)\theta_{j_1}(\gamma_1) \right) \left(\int^{[N^2M_*]} w_2(|\gamma_1 - \gamma_2|)\bar{\phi}_{j_2}(\gamma_2)\theta_{j_4}(\gamma_2)d\gamma_2 \right) d\gamma_1 \right] A_{j_0j_2j_3j_4}\,. \tag{3.23}$$

PROPOSITION 3.18. *We can compute*  *with cost* $N^6 + N^5 + N^4 + N^3 + N^2 + N$ *via the ordering*

$$\sum_{j_0,j_1,j_2,j_3}^{[N^4]} A_{j_0j_1j_2j_3} \left(\sum_{j_4,j_5}^{[N^6]} B_{j_2j_3j_4j_5} C_{j_4j_5j_0j_1} \right). \tag{3.24}$$

4. Further considerations.

4.1. Using localization. In this section we assume w_1 and w_2 are localized in space in a region of "volume" $K < M$. (Since $w_3(|\gamma - \gamma'|) = (\|\mathbf{r} - \mathbf{r}'\|)^{-1}$ we cannot assume it is local.)

Consider the difficult integral

$$A_{j_1,j_2}(\gamma_0,\gamma_1) = \int w_2(|\gamma_1 - \gamma_2|)w_3(|\gamma_2 - \gamma_0|)\bar{\phi}_{j_2}(\gamma_2)\theta_{j_1}(\gamma_2)d\gamma_2 \tag{4.1}$$

from (3.15), which in general costs N^2MM_*. For fixed values of γ_1, j_1, and j_2, the function $w_2(|\gamma_1 - \gamma_2|)\bar{\phi}_{j_2}(\gamma_2)\theta_{j_1}(\gamma_2)$ is localized by w_2 to a region of volume K. Convolving with $w_3(|\gamma_2 - \gamma_0|)$, which is also localized, costs K^2 and results in a function localized to a region of volume proportional to K. (The volume is at most $27K$ but we neglect the factor of 27 for clarity.) The cost to do (4.1) is thus K^2N^2M. The function $A_{j_1,j_2}(\gamma_0,\gamma_1)$ is "banded" and costs KM to represent. Inserting it in (3.15), we obtain

$$\sum_{j_0}^{[N]} \left(\int^{[NM]} \bar{\phi}_{j_0}(\gamma_0) \left[\int^{[NKM]} \bar{w}_1(|\gamma_0 - \gamma_1|)\theta_{j_0}(\gamma_1) \left(\sum_{j_1}^{[NKM]} \bar{\phi}_{j_1}(\gamma_1) \left(\sum_{j_2}^{[N^2KM]} \theta_{j_2}(\gamma_0)A_{j_1,j_2}(\gamma_0,\gamma_1) \right) \right) d\gamma_1 \right] \right) d\gamma_0\,. \tag{4.2}$$

Our total cost is thus $K^2N^2M + KN^2M + 2KNM + NM + N$.

If $K^2 < M_*$, then localization was helpful. Similar arguments apply to the other cyclic terms in (3.13) and (3.14). If $K^2M < N^2M_*$, then the cyclic terms are no longer dominant in cost. The next most costly terms are  ,  , and  with dominant cost N^4M_*, which localization cannot significantly improve.

4.2. RI approximation. The RI technique is widely used for integrals when geminals are present; see the discussion in the review [27]. In this section we apply it to our most costly cyclic case, which costs $\mathcal{O}(N^2 M M_*)$ using (3.15).

Suppose we can find an orthonormal set of functions $\{u_q(\gamma)\}_{q=1}^{Q}$ such that for all values of j_1, j_2, and γ_0 there is a sufficiently good approximation

$$\begin{aligned}(4.3)\quad &\int w_2(|\gamma_1-\gamma_2|)w_3(|\gamma_2-\gamma_0|)\bar{\phi}_{j_2}(\gamma_2)\theta_{j_1}(\gamma_2)d\gamma_2\\ &\approx \sum_{q=1}^{Q}\left(\int w_2(|\gamma_1-\gamma_3|)u_q(\gamma_3)d\gamma_3\right)\left(\int u_q(\gamma_2)w_3(|\gamma_2-\gamma_0|)\bar{\phi}_{j_2}(\gamma_2)\theta_{j_1}(\gamma_2)d\gamma_2\right).\end{aligned}$$

(The name "resolution of the identity" comes from considering $\sum_{q=1}^{Q} u_q(\gamma)u_q(\gamma')$ as an approximation of the identity operator as an integral operator.) Substituting (4.3) into (3.15) and rearranging, we obtain

$$\begin{aligned}(4.4)\quad &\sum_{q=1}^{Q}\left(\int^{[QM]}\left(\sum_{j_0,j_1}^{[QN^2M]}\bar{\phi}_{j_0}(\gamma_0)\right.\right.\\ &\left[\int^{[QN^2M_*]}\bar{\phi}_{j_1}(\gamma_1)\theta_{j_0}(\gamma_1)\bar{w}_1(|\gamma_0-\gamma_1|)\left(\int^{[QM_*]}w_2(|\gamma_1-\gamma_3|)u_q(\gamma_3)d\gamma_3\right)d\gamma_1\right]\\ &\left.\left.\left[\sum_{j_2}^{[QN^2M]}\theta_{j_2}(\gamma_0)\left(\int^{[QN^2M_*]}u_q(\gamma_2)w_3(|\gamma_2-\gamma_0|)\bar{\phi}_{j_2}(\gamma_2)\theta_{j_1}(\gamma_2)d\gamma_2\right)\right]\right)d\gamma_0\right).\end{aligned}$$

The dominant cost changed from $N^2 M M_*$ to $QN^2 M_*$, which means we have replaced a factor of M with Q. The significance of this reduction is a matter of interpretation. On the one hand, the review [27] is enthusiastic about this technique, thus implying useful approximations (4.3) can be obtained with $Q \ll M$. On the other hand, the requirement that the approximation hold for all values of j_1, j_2, and γ_0 suggests $Q \approx M$; otherwise one could change the representation of functions of γ to reduce M to Q. Some gain can be had from the approximate nullspace of convolving with w_2, but if w_2 is localized and has a cusp, then this gain may not be significant. Note that these considerations are not specific to our method but apply whenever RI is used.

4.3. Using Gaussian geminals. In this section we explore the computational gain from choosing Gaussian geminals. Since the geminal would be a Gaussian (of possibly different exponent) for each spin variable and the integrals would be computed separately, we will neglect spin in our discussion. Thus we assume $w_p(\|\mathbf{r}_1-\mathbf{r}_2\|) = \exp(-\tau_p\|\mathbf{r}_1-\mathbf{r}_2\|^2)$ for $p=1,2$. (Since $w_3(|\gamma-\gamma'|) = (\|\mathbf{r}-\mathbf{r}'\|)^{-1}$ we cannot assume it is a Gaussian.)

In the dominant cyclic cases, the difficulty is the computation of integrals of the form

$$(4.5)\qquad \int A(\gamma_2)w_1(|\gamma_0-\gamma_2|)w_2(|\gamma_1-\gamma_2|)d\gamma_2\,,$$

which reduces now to

$$(4.6)\qquad \int A(\mathbf{r}_3)\exp(-\tau_1\|\mathbf{r}_1-\mathbf{r}_3\|^2)\exp(-\tau_2\|\mathbf{r}_2-\mathbf{r}_3\|^2)d\mathbf{r}_3\,.$$

By multiplying out and completing the square, we obtain

(4.7)
$$\exp\left(-\frac{\tau_1\tau_2}{\tau_1+\tau_2}\|\mathbf{r}_1-\mathbf{r}_2\|^2\right)\int\exp\left(-(\tau_1+\tau_2)\left\|\mathbf{r}_3-\frac{\tau_1\mathbf{r}_1+\tau_2\mathbf{r}_2}{\tau_1+\tau_2}\right\|^2\right)A(\mathbf{r}_3)d\mathbf{r}_3\,.$$

Substituting $\mathbf{s}=(\tau_1\mathbf{r}_2+\tau_2\mathbf{r}_2)/(\tau_1+\tau_2)$, we have

$$\exp\left(-\frac{\tau_1\tau_2}{\tau_1+\tau_2}\|\mathbf{r}_1-\mathbf{r}_2\|^2\right)\int\exp\left(-(\tau_1+\tau_2)\,\|\mathbf{r}_3-\mathbf{s}\|^2\right)A(\mathbf{r}_3)d\mathbf{r}_3\,, \tag{4.8}$$

which we can integrate in $\mathbf{r}_3$ at cost $\mathcal{O}(M_*)$ to obtain

$$\exp\left(-\frac{\tau_1\tau_2}{\tau_1+\tau_2}\|\mathbf{r}_1-\mathbf{r}_2\|^2\right)B(\mathbf{s})\,. \tag{4.9}$$

Substituting back for $\mathbf{s}$ we obtain

$$C(\mathbf{r}_1,\mathbf{r}_2)=\exp\left(-\frac{\tau_1\tau_2}{\tau_1+\tau_2}\|\mathbf{r}_1-\mathbf{r}_2\|^2\right)B\left(\frac{\tau_1\mathbf{r}_2+\tau_2\mathbf{r}_2}{\tau_1+\tau_2}\right)\,, \tag{4.10}$$

which costs $\mathcal{O}(M^2)$ to represent. By using this technique we avoid costs with factors of MM_*. Specifically, we reduce the cost of (3.15) from $\mathcal{O}(N^2MM_*)$ to $\mathcal{O}(N^2M^2)$, the cost of (3.13) from $\mathcal{O}(MM_*+NM^2)$ to $\mathcal{O}(NM^2)$, and the cost of (3.14) from $\mathcal{O}(MM_*)$ to $\mathcal{O}(M^2)$. Since these costs include factors of M^2, these cases are still the most expensive.

Acknowledgments. I would like to thank John Linser, Vusi Mpendulo Magagula, Son Nguyen, Benigno Parra, Dhinali Peiris, and Chelsie Wollett, who were research assistants on this project while students at Ohio University.

REFERENCES

[1] R. J. Bartlett, *Many-body perturbation theory and coupled cluster theory for electron correlation in molecules*, Ann. Rev. Phys. Chem., 32 (1981), pp. 359–401.

[2] G. Beylkin and M. J. Mohlenkamp, *Numerical operator calculus in higher dimensions*, Proc. Natl. Acad. Sci. USA, 99 (2002), pp. 10246–10251.

[3] G. Beylkin and M. J. Mohlenkamp, *Algorithms for numerical analysis in high dimensions*, SIAM J. Sci. Comput., 26 (2005), pp. 2133–2159.

[4] G. Beylkin, M. J. Mohlenkamp, and F. Pérez, *Approximating a wavefunction as an unconstrained sum of Slater determinants*, J. Math. Phys., 49 (2008), 032107.

[5] F. A. Bischoff, E. F. Valeev, W. Klopper, and C. L. Janssen, *Scalar relativistic explicitly correlated R12 methods*, J. Chem. Phys., 132 (2010), 214104.

[6] S. F. Boys, *The integral formulae for the variational solution of the molecular many-electron wave equations in terms of Gaussian functions with direct electronic correlation*, Proc. Roy. Soc. London Ser. A Math., 258 (1960), p. 402.

[7] R. Bukowski, B. Jeriorski, and K. Szalewicz, *Gaussian geminals in explicitly correlated coupled cluster theory including single and double excitations*, J. Chem. Phys., 110 (1999), pp. 4165–4183.

[8] H.-J. Bungartz and M. Griebel, *Sparse grids*, Acta Numer., 13 (2004), pp. 147–269.

[9] P. Cársky, J. Paldus, and J. Pittner, eds., *Recent progress in coupled cluster methods: Theory and applications*, Challenges and Advances in Computational Chemistry and Physics 11, Springer-Verlag, Berlin, 2010.

[10] S. R. Chinnamsetty, M. Espig, B. N. Khoromskij, W. Hackbusch, and H.-J. Flad, *Tensor product approximation with optimal rank in quantum chemistry*, J. Chem. Phys., 127 (2007), 084110.

[11] S. Fournais, M. Hoffmann-Ostenhof, T. Hoffmann-Ostenhof, and T. Østergaard Sørensen, *Sharp regularity results for Coulombic many-electron wave functions*, Comm. Math. Phys., 255 (2005), pp. 183–227.
[12] I. P. Gavrilyuk, W. Hackbusch, and B. N. Khoromskij, *Hierarchical tensor-product approximation to the inverse and related operators for high-dimensional elliptic problems*, Computing, 74 (2005), pp. 131–157.
[13] L. Grasedyck, *Hierarchical singular value decomposition of tensors*, SIAM J. Matrix Anal. Appl., 31 (2010), pp. 2029–2054.
[14] M. Griebel and J. Hamaekers, *Sparse grids for the Schrödinger equation*, M2AN Math. Model. Numer. Anal., 41 (2007), pp. 215–247.
[15] W. Hackbusch and B. N. Khoromskij, *Low-rank Kronecker-product approximation to multi-dimensional nonlocal operators.* I. *Separable approximation of multi-variate functions*, Computing, 76 (2006), pp. 177–202.
[16] W. Hackbusch and B. N. Khoromskij, *Low-rank Kronecker-product approximation to multi-dimensional nonlocal operators.* II. *HKT representation of certain operators*, Computing, 76 (2006), pp. 203–225.
[17] W. Hackbusch, B. N. Khoromskij, and E. E. Tyrtyshnikov, *Hierarchical Kronecker tensor-product approximations*, J. Numer. Math., 13 (2005), pp. 119–156.
[18] P. Hatzikonstantinou and J. M. Irvine, *Method of correlated basis functions and FHNC theory*, J. Phys. A, 15 (1982), pp. 1715–1731.
[19] T. Helgaker, P. Jørgensen, and J. Olsen, *Molecular Electronic-Structure Theory*, John Wiley and Sons, New York, 2000.
[20] J. G. Hill, S. Mazumder, and K. A. Peterson, *Correlation consistent basis sets for molecular core-valence effects with explicitly correlated wave functions: The atoms B-Ne and Al-Ar*, J. Chem. Phys., 132 (2010).
[21] E. A. Hylleraas, *Über den Grundzustand des Heliumatoms*, Z. Phys., 48 (1928), pp. 469–494.
[22] R. Jastrow, *Many-body problems with strong forces*, Phys. Rev., 98 (1955), pp. 1479–1483.
[23] T. Kato, *On the eigenfunctions of many-particle systems in quantum mechanics*, Comm. Pure Appl. Math., 10 (1957), pp. 151–177.
[24] S. Kedzuch, O. Demel, J. Pittner, S. Ten-no, and J. Noga, *Multireference F*12 *coupled cluster theory: The Brillouin-Wigner approach with single and double excitations*, Chem. Phys. Lett., 511 (2011), pp. 418–423.
[25] W. Klopper, *R*12 *methods, Gaussian geminals*, in Modern Methods and Algorithms of Quantum Chemistry, J. Grotendorst, ed., John von Neumann Institute for Computing, Jülich, Germany, 2000, pp. 153–201.
[26] W. Klopper, R. A. Bachorz, C. Haettig, and D. P. Tew, *Accurate computational thermochemistry from explicitly correlated coupled-cluster theory*, Theoret. Chem. Acc., 126 (2010), pp. 289–304.
[27] W. Klopper, F. R. Manby, S. Ten-No, and E. F. Valeev, *R*12 *methods in explicitly correlated molecular electronic structure theory*, Int. Rev. Phys. Chemistry, 25 (2006), pp. 427–468.
[28] A. Köhn and D. P. Tew, *Explicitly correlated coupled-cluster theory using cusp conditions.* I. *Perturbation analysis of coupled-cluster singles and doubles (CCSD-F*12*)*, J. Chem. Phys., 133 (2010), 174117.
[29] J. Komasa and J. Rychlewski, *Explicitly correlated Gaussian functions in variational calculations. Microhartree accuracy of the Helium dimer energy*, Chem. Phys. Lett., 249 (1996), pp. 253–256.
[30] J. Komasa, Wojciech, and J. Rychlewski, *Explicitly correlated Gaussian functions in variational calculations: The ground state of the beryllium atom*, Phys. Rev. A, 52 (1995), pp. 4500–4507.
[31] W. Kutzelnigg, *R*12*-Dependent Terms in the Wave-Function as Closed Sums of Partial-Wave Amplitudes for Large-L*, Theoretica Chimica Acta, 68 (1985), pp. 445–469.
[32] W. Kutzelnigg and W. Klopper, *Wave functions with terms linear in the interelectronic coordinates to take care of the correlation cusp.* I. *General theory*, J. Chem. Phys., 94 (1991), pp. 1985–2001.
[33] J. V. L. Longstaff and K. Singer, *The use of Gaussian (exponential quadratic) wave functions in molecular problems.* II. *Wave functions for the ground state of the hydrogen atom and of the hydrogen molecule*, Proc. Roy. Soc. London Ser. A Math., 258 (1960), p. 421.
[34] P.-O. Löwdin, *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 (1955), pp. 1474–1489.
[35] M. J. Mohlenkamp, *A center-of-mass principle for the multiparticle Schrödinger equation*, J. Math. Phys., 52 (2010), 022112.

[36] J. Noga and W. Kutzelnigg, *Coupled cluster theory that takes care of the correlation cusp by inclusion of linear term in the interelectronic coordinates*, J. Chem. Phys., 101 (1994), pp. 7738–7762.
[37] J. Noga, W. Kutzelnigg, and W. Klopper, *CC-R*12*, a correlation cusp corrected coupled-cluster method with a pilot application to the* Be_2 *potential curve*, Chem. Phys. Lett., 199 (1992), pp. 497–504.
[38] I. V. Oseledets and E. E. Tyrtyshnikov, *Breaking the curse of dimensionality, or how to use SVD in many dimensions*, SIAM J. Sci. Comput., 31 (2009), pp. 3744–3759.
[39] R. T. Pack and W. Byers Brown, *Cusp conditions for molecular wavefunctions*, J. Chem. Phys., 45 (1966), pp. 556–559.
[40] R. Pauncz, *The Symmetric Group in Quantum Chemistry*, CRC Press, Boca Raton, FL, 1995.
[41] B. J. Persson and P. R. Taylor, *Accurate quantum-chemical calculations: The use of Gaussian-type geminal functions in the treatment of electron correlation*, J. Chem. Phys., 105 (1996), pp. 5915–5926.
[42] B. J. Persson and P. R. Taylor, *Molecular integrals over Gaussian-type geminal basis functions*, Theoret. Chem. Acc., 97 (1997), pp. 240–250.
[43] J. Rychlewski, W. Cencek, and J. Komasa, *The equivalence of explicitly correlated Slater and Gaussian functions in variational quantum chemistry computations. The ground state of* H_2, Chem. Phys. Lett., 229 (1994), pp. 657–660.
[44] T. Shiozaki and S. Hirata, *Communications: Explicitly correlated second-order Moller-Plesset perturbation method for extended systems*, J. Chem. Phys., 132 (2010), 151101.
[45] T. Shiozaki and H.-J. Werner, *Communication: Second-order multireference perturbation theory with explicit correlation: CASPT*2*-F*12, J. Chem. Phys., 133 (2010), 141103.
[46] K. Singer, *The use of Gaussian (exponential quadratic) wave functions in molecular problems. I. General formulae for the evaluation of integrals*, Proc. Roy. Soc. London Ser. A Math., 258 (1960), p. 412.
[47] K. Szalewicz and B. Jeziorski, *Explicitly-correlated Gaussian geminals in electronic structure calculations*, Molecular Phys., 108 (2010), pp. 3091–3103.
[48] D. P. Tew, C. Hättig, R. A. Bachorz, and W. Klopper, *Recent Progress in Coupled Cluster Methods: Theory and Applications*, Challenges and Advances in Computational Chemistry and Physics 11, Springer-Verlag, Berlin, 2010.
[49] D. P. Tew and W. Klopper, *Open-shell explicitly correlated* *F*12 *methods*, Molecular Phys., 108 (2010), pp. 315–325.
[50] K. D. Vogiatzis, E. C. Barnes, and W. Klopper, *Interference-corrected explicitly-correlated second-order perturbation theory*, Chem. Phys. Lett., 503 (2011), pp. 157–161.
[51] H.-J. Werner, T. B. Adler, G. Knizia, and F. R. Manby, *Recent progress in coupled cluster methods: Theory and applications*, Challenges and Advances in Computational Chemistry and Physics 11, Springer-Verlag, Berlin, 2010.
[52] H. Yserentant, *On the regularity of the electronic Schrödinger equation in Hilbert spaces of mixed derivatives*, Numer. Math., 98 (2004), pp. 731–759.
[53] H. Yserentant, *Sparse grid spaces for the numerical solution of the electronic Schrödinger equation*, Numer. Math., 101 (2005), pp. 381–389.
[54] H. Yserentant, *The hyperbolic cross space approximation of electronic wavefunctions*, Numer. Math., 105 (2007), pp. 659–690.
[55] H. Yserentant, *The Mixed Regularity of Electronic Wave Functions Multiplied by Explicit Correlation Factors*, Technical report 49, DFG-Schwerpunktprogramm 1324, 2010.
[56] H. Yserentant, *Regularity and approximability of electronic wave functions*, Lecture Notes in Math. 2000, Springer-Verlag, Berlin, 2010.