Efficient evaluation of AGP reduced density matrices

Armin Khamoshi,1 Thomas Henderson,2,1 and Gustavo Scuseria2,1
1)Department of Physics and Astronomy, Rice University, Houston, TX 77005-1892
2)Department of Chemistry, Rice University, Houston, TX 77005-1892

(Dated: September 16, 2019)

We propose and implement an algorithm to calculate the norm and reduced density matrices of the antisymmetrized geminal power (AGP) of any rank with polynomial cost. Our method scales quadratically per element of the reduced density matrices. Numerical tests indicate that our method is very fast and capable of treating systems with a few thousand orbitals and hundreds of electrons reliably in double-precision. In addition, we present reconstruction formulae that allows one to decompose higher order reduced density matrices in terms of linear combinations of lower order ones, thereby reducing the computational cost significantly.

Keywords: Antisymmetrized geminal power, reduced density matrix, AGP RDMs polynomial cost

I. INTRODUCTION

In electronic structure theory, a geminal is a wavefunction for two electrons. Geminals are central to the concept of bonding and have a long history in quantum chemistry.1 The geminal creation operator can, in general, be written as

\[ \Gamma^\dagger = \sum_{pq} \eta_{pq} c_p^\dagger c_q^\dagger, \]  

where \( \eta_{pq} \) is antisymmetric, \( 2M \) is the total number of spin-orbitals and \( c_p^\dagger \) is the creation operator of a fermion in spin-orbital \( p \).

The simplest geminal wave function is perhaps the antisymmetrized geminal power (AGP) where all geminals are the same2

\[ |AGP\rangle = \frac{1}{N!} (\Gamma^\dagger)^N |\cdots\rangle. \]  

Here \( \langle \cdots \rangle \) is the physical vacuum containing no electrons, \( N \) is the number of pairs (2N electrons), and the factor \( 1/N! \) is introduced for convenience.

Without loss of generality, we choose to work in the natural orbital basis of the geminal. This is accomplished by applying a unitary transformation that brings \( \eta_{pq} \) into a block diagonal form, wherein the one-body density matrix is diagonal and all spin-orbitals are paired. In this basis, it is mathematically simpler to work with hardcore boson operators, so that

\[ \Gamma^\dagger = \sum_{p=1}^{M} \eta_p P_p^\dagger, \]  

where

\[ P_p^\dagger = c_p^\dagger c_p^\dagger \]  

and

\[ N_p = c_p^\dagger c_p^\dagger + c_p^\dagger c_p^\dagger. \]  

such that \( c_p^\dagger \) is the fermion creation operator in orbital \( p \), and \( \bar{p} \) is the "paired" companion of \( p \). The pair creation and annihilation operators, \( P_p^\dagger \), and \( P_p \), along with \( N_p \) are the generators of a \( su(2) \) algebra

\[ [P_p, P_q^\dagger] = \delta_{pq} (1 - N_p) \]  

\[ [N_p, P_q^\dagger] = 2\delta_{pq} P_q^\dagger. \]

Notice that, written in its natural orbital basis, AGP preserves the so-called seniority symmetry. That is, orbitals with bars and no bars are either both occupied or empty. While in a general Hamiltonian the seniority-defining pairing scheme is arbitrary but can be optimized, in Hamiltonians for which seniority is a symmetry the pairs are naturally defined.

In the nuclear structure and condensed matter physics communities, AGP is known as the number projected Bardeen-Cooper-Schrieffer (PBCS) wavefunction.3,4 The claim to fame of AGP is its ability to describe off-diagonal long range order, a criterion for superconductivity,5 without breaking number symmetry as in BCS.6 AGP is not a great wavefunction per se in quantum chemistry or condensed matter physics, because in most situations electron pairs are very different from each other. However, AGP is potentially an excellent starting point for geminal correlation models because it encompasses a combinatorial number of Slater determinants. Indeed, it is easy to see that

\[ |AGP\rangle = \sum_{1\leq p_1 < \cdots < p_N \leq M} \eta_{p_1}\cdots\eta_{p_N} \left| P_{p_1}^\dagger \cdots P_{p_N}^\dagger \right| \cdots \]  

This is a superposition of all possible seniority zero (paired) determinants involving \( N \) pairs in \( M \) orbitals but the coefficients are factorized by \( \eta_p \) rather than being a tensor. In the latter case, when coefficients are general rather than factorized, the wavefunction is known as doubly occupied configuration interaction (DOCI).7-10 DOCI is obtainable by exact diagonalization over the space of seniority zero determinants with combinatorial cost as a function of \( M \) and \( N \). However, AGP (PBCS) can be solved variationally and optimized using symmetry breaking and restoration techniques with mean-field \( O(M^2) \) cost.11,12 Other classes of geminal theories like Bethe ansatz (BA) for solving
the integrable Richardson-Gaudin Hamiltonians or APIG (antisymmetrized product of interacting geminals) wherein all geminals are different are very interesting models but have combinatorial cost when applied to general (non-integrable) Hamiltonians.

AGP is qualitatively correct for attractive pairing interactions at all correlation regimes where many other models have serious difficulties. This makes AGP an attractive starting point for more accurate correlated geminal theories. Recent work in our research group has shown that an AGP-based configuration interaction (CI) is a promising step in this direction. For AGP-based correlated theories to be computationally affordable for large systems, it is crucial that the AGP reduced density matrices (RDMs) can be expressed in terms of lower rank RDMs, using what we call reconstruction formulae.

In order to meet this goal, we introduce two techniques in this paper. First, we develop an efficient way of calculating the individual elements of RDMs to all ranks. For this, we formulate all RDMs in terms of the elementary symmetric polynomials and use the sumESP algorithm to compute the sum efficiently. We argue that, with appropriate normalization, this is a reliable, fast, and stable method capable of treating large systems. Secondly, we rigorously prove that all AGP RDMs are expressible in terms of linear combinations of lower rank RDMs, using what we call reconstruction formulae. This is the most significant and novel contribution of this paper and has important theoretical and numerical implications. The significance of this relies in the fact that all correlated theories require high rank RDMs—often as high as 5- and 6-body. Therefore the ability to breakdown high-rank RDMs makes AGP a good starting point for correlated methods. Indeed, our result is reminiscent of Hartree-Fock theory wherein all high rank RDMs can be written as products of 1-body RDMs. From a computational perspective, the reconstruction formulae presented here reduce the cost and scaling of correlated AGP calculations significantly as demonstrated in Sec. V.

II. BASIC EXPRESSIONS

The analytic expressions for the norm and RDMs of AGP already exist in the literature and have been expressed in many different forms. For the sake of completeness, and to familiarize the reader with our notation, we present our own version of these derivations. In so doing, we introduce a Lie algebraic approach to understanding how $P^\dagger_p$, $P_p$, and $N_p$ act on AGP. The AGP state and its excitations describe a general (non-integrable) Hamiltonian.

A. Norm of the AGP wavefunction

The norm of the AGP wavefunction corresponding to $N$ pairs and $M$ orbitals can be obtained by calculating the contractions explicitly. From the commutation relations in Eq. (6) and the fact that $(P^\dagger_p)^2 = 0$, it follows that

$$\langle AGP|AGP \rangle = \sum_{1 \leq p_1 < \ldots < p_N \leq M} \eta_{p_1} \ldots \eta_{p_N} \langle -|P_{p_1}^\dagger P_{p_N}^\dagger P_{p_1} \ldots P_{p_N} | - \rangle$$

$$= \sum_{1 \leq p_1 < \ldots < p_N \leq M} \eta_{p_1}^2 \ldots \eta_{p_N}^2 S^M_N,$$

(8)

where $S^M_N$ is the elementary symmetric polynomial (ESP) of degree $N$ with $M$ variables associated with the vector $\{\eta_1^2, \ldots, \eta_M^2\}$.

It is often convenient and numerically better posed to work with normalized AGP, i.e. $\langle AGP|AGP \rangle = 1$. One can easily verify that the following choice does the job:

$$\eta_p \rightarrow \frac{\eta_p}{(S^N_N)^{\frac{1}{N}}}.$$  

(9)

B. Differential Representation

Before we derive the expressions for the RDMs, we need to understand how $P^\dagger_p$, $P_p$, and $N_p$ act on AGP. We formulate all RDMs in terms of the integrable Richardson-Gaudin Hamiltonians and its excitations. And, we purposefully formulate all the matrix elements in such a way that they can be directly computed by the elementary symmetric polynomials.

The AGP state and its excitations, $\{P_{\{1\ldots N\}} | N \}$, are expressible in terms of lower rank RDMs, using what we call reconstruction formulae. This results in

$$\left[ N_p, (\Gamma^\dagger)^N \right] = 2N\eta_p (\Gamma^\dagger)^{N-1} P^\dagger_p,$$

(10)

from which we get

$$N_p | N \rangle = 2\eta_p P^\dagger_p | N - 1 \rangle,$$

(11)

where $| N \rangle$ is the shorthand notation for $| AGP \rangle$ with $N$ pairs. On the other hand, using explicit derivatives with respect to $\eta_p$ on $| \Gamma^\dagger \rangle$, one obtains

$$\frac{\partial}{\partial \eta_p} | N \rangle = P^\dagger_p | N - 1 \rangle.$$  

(12)

This results in

$$N_p | N \rangle = 2\eta_p \frac{\partial}{\partial \eta_p} | N \rangle.$$  

(13)

A similar derivation for $P_p$ yields

$$\left[ P_{\{1\ldots N\}} | (\Gamma^\dagger)^N \right] = N\eta_p (\Gamma^\dagger)^{N-1} (1 - N_p) - N (N - 1) \eta_p^2 P^\dagger_p (\Gamma^\dagger)^{N-2},$$

(14)
and therefore
\[ P_p |N\rangle = \eta_p |N - 1\rangle - \eta_p^2 P_p^\dagger |N - 2\rangle = (\eta_p - \eta_p^2 \frac{\partial}{\partial \eta_p}) |N - 1\rangle. \tag{15} \]

Eq. (13), and Eq. (15) are the differential representation of operators \(N_p\) and \(P_p\) over AGP. Noting that \(P_p^\dagger\) is the Hermitian conjugate of \(P_p\) and acts to the left, we have the differential representation for all the generators as desired. As a corollary to these equations and the commutations relations of Eq. (6), it is easy to show that \((N_p)^2 = 2N_p\) and \(N_p = 2P_p^\dagger P_p\) over AGP. We frequently use these properties in this paper.

**C. AGP reduced density matrices**

Consider a many-body system of fermions. To evaluate the energy or other observables thereof, one needs to compute the many-body RDMs. Since AGP preserves seniority and the total number of pairs is fixed, only two kinds of contractions are nonzero
\[ \langle AGP|...c_{p_1}...c_{p_p}|AGP\rangle \tag{16a} \]
\[ \langle AGP|...c_{p_1}^\dagger c_{p_1}...c_{q_1}...c_{q_q}|AGP\rangle. \tag{16b} \]

Clearly, all terms like that of Eq. (16b) can be written as \(P_p^\dagger\) and \(P_p\) by definition; and all like that of Eq. (16a) can be replaced by \(N_p/2\) as all electrons come in pairs. Therefore, all electron density matrices can be written as linear combinations of terms like \(\langle P_p^\dagger...N_q...P_r...\rangle\). Recall that \(N_p = 2P_p^\dagger P_p\) over AGP; therefore, we can ultimately normal order everything such that all \(P_p^\dagger\) are on the left and all \(P\) are on the right, i.e. (\(P_p^\dagger...P_q...\)). We refer to these density matrices as pair RDMs.

We set out by getting the expression for the \(n = 1\) pair RDM
\[ \gamma_{pq} = \langle AGP|P_p^\dagger P_q|AGP\rangle. \tag{17} \]

Notice that
\[ P_q|AGP\rangle = (\eta_q - 2^2 \frac{\partial}{\partial \eta_q}) |N - 1\rangle, \]
\[ = \eta_q \sum_{p_1 < ... < p_{N-1} \neq q} \eta_{p_1}...\eta_{p_{N-1}}P_{p_1}^\dagger...P_{p_{N-1}}^\dagger | - \rangle. \tag{18} \]

and from this it follows
\[ \gamma_{pq} = \eta_p \eta_q \sum_{1 \leq p_1 < ... < p_{N-1} \neq q} \eta_{p_1}^2...\eta_{p_{N-1}}^2 = \eta_p \eta_q S_{N-1}^{M-2}(\eta_p^2, \eta_q^2), \tag{19} \]

where \(S_{N-1}^{M-2}(\eta_p^2, \eta_q^2)\) denotes ESP over \(\{\eta_1^2,...,\eta_{M-2}^2\}\) such that \(\eta_p^2\) and \(\eta_q^2\) are omitted. Obviously if \(p = q\), then
\[ \gamma_{pp} = \eta_p^2 S_{N-1}^{M-1}(\eta_p^2). \]

Generalization of this to higher rank RDMs follows the same reasoning. By induction on \(n\), one can show that the matrix elements of the \(n\)-pair RDM, \(\gamma^{(n)}\), is
\[ \gamma_{p_1...p_n} = \langle AGP|P_{p_1}^\dagger...P_{p_n}^\dagger P_{q_1}...P_{q_n}|AGP\rangle = \prod_{i=1}^n \eta_{p_i} \eta_{q_i} S_{N-n}^{M-d}(\eta_{p_1}^2,...,\eta_{p_n}^2), \tag{20} \]

where \(d\) is the number of unique indices among \(\{p_i\} \cup \{q_i\}\). Note that \(\gamma_{p_1...p_n}\) is symmetric with respect to permutations of \(p_i’s\) and \(q_i’s\). If \(p_i = p_j\), or \(q_i = q_j\) for some \(i, j\), then the corresponding matrix element is instead zero. Therefore, in practice, one needs to compute only the unique elements, e.g. \(\gamma_{p_1...p_n}\), which requires at most \(\binom{M}{n}\) elements—as opposed to \(M^{2n}\). There is an extra symmetry in Eq. (20); \(p_i\) is interchangeable with \(q_j\) for some \(i, j\), if and only if \(p_i \neq \{q_1...q_{j-1}p_{j+1}...q_n\}\) and \(q_j \neq \{p_1...p_{i-1}p_{i+1}...p_n\}\). We make use of this property later in Sec. IV.

There is a special case of \(\gamma_{p_1...p_n}\) that we need later in this paper; that is when \(p_i = q_i\), for all \(i\). Again, by \(N_p = 2P_p^\dagger P_p\), this can be written in terms of \(\langle N_{p_1}...N_{p_2}...N_{p_n}\rangle\). We refer to this as a number RDM. Formally, we define the number RDM of rank \(n\), \(\nu^{(n)}\), as
\[ \nu_{p_1...p_n} = \langle AGP|N_{p_1}N_{p_2}...N_{p_n}|AGP\rangle, \tag{21} \]

where all the indices are assumed to be different—otherwise it would be equal to a lower rank number RDM by \(N_p^2 = 2N_p\). From Eq. (20) this can be computed by
\[ \nu_{p_1...p_n} = \prod_{i=1}^n \eta_{p_i}^2 S_{N-n}^{M-2}(\eta_{p_1}^2,...,\eta_{p_n}^2). \tag{22} \]

**III. NUMERICAL ALGORITHM**

From Sec. II it is clear that computing ESP efficiently is imperative to using AGP for any realistic system. This is because there are \(\binom{M}{N}\) summands in every \(S_N^M\) and a straightforward summation could grow combinatorially with system size. In this section, we introduce the sumESP algorithm which calculates ESP with polynomial cost. First, we briefly review some important remarks about the error analysis that were extensively studied in Ref.22,23 In Sec. IIIB and III C, we study the time scales of various computations involving the norm of AGP and the corresponding RDMs as a function of \(N\) and \(M\). The computer environments used for the runtime measurements can be found in Appendix C.

**A. sumESP and error analysis**

The algorithm we use to calculate the ESP is one initially proposed by Fischer.21,28 This is the same summation algorithm used today in MATLAB’s poly function.
Following the notation of Ref. 23 we call it \text{sumESP}. The \text{sumESP} algorithm takes advantage of the following property of ESP
\[ S_N^M = S_N^{M-1}(\eta_p^2) + \eta_p^2 S_N^{M-1}(\eta_p^2). \] (23)
Intuitively, Eq. (23) says that we can split any ESP into two sums such that one contains some arbitrary term \( \eta_p^2 \) in all of its summands and one that does not. In the context of quantum chemistry, Eq. (23) is mentioned in Ref. 29. However, there has been substantial progress in the computational and applied mathematics community to better understand and craft the algorithm. In particular, Ref. 22 performed an error analysis and proved the stability of the algorithm. Ref. 23 made a slight improvement to the roundoff error bound and introduced new ways of calculating \text{sumESP} with enhanced precision. From these analyses it follows that \text{sumESP} is highly accurate and stable for positive summands\textsuperscript{22} as is the case for the overlaps of AGP.

Quantitatively, one can say that, given \( \{\eta_1^2, \ldots, \eta_M^2\} \) as a set of floating-point numbers, the “worst case” error due to \text{sumESP} is bounded above as follows\textsuperscript{23}
\[ \left| \frac{S_N^M - S_N^{M}}{S_N^{M}} \right| \leq \frac{2(M-1)\epsilon}{1 - 2(M-1)\epsilon}, \] (24)
where \( S_N^{M} \) is the exact value of ESP and \( S_N^{M} \) is computed using \text{sumESP} in floating-point arithmetic; \( \epsilon \) denotes machine epsilon (unit roundoff). The error bound assumes that there are no numerical overflow and underflow occurring anywhere in the calculation.\textsuperscript{22} In practice, however, poor choice of normalization of \( \eta_p \) could lead to overflow and/or underflow. For example, when \( \eta_p > 1 \), the sum could overflow even for small values of \( M \). Generally, \text{sumESP} is better conditioned when \( 0 < \eta_p < 1 \); this is because the underlying summations in the algorithm prevents the summands from getting too small. Nevertheless, the sum is prone to underflow when \( N \) is close to \( M \) and \( N \) is large. (Fortunately, this is not a concern in physical problems since the number of orbitals is always much greater than the number of electrons.) And it is prone to overflow when \( 1 \ll N < M/2 \) and \( M \) is large.

To make this analysis more quantitative, for \( 0 < \eta_p < 1 \), we have plotted the magnitude of \text{sumESP} (the same as the overlap of AGP) as a function of \( M \) in Fig. 1. In Fig. 2 (the right axes), we fix \( M = 10^2, 10^3 \) and vary \( N \). It follows from the plots that, in very large systems, one must take extra caution when performing calculations away from \( N \ll M \) or half-filling. But in moderate or small systems overflow and underflow should not be a concern in physical problems. It is noteworthy that, often times, the overflow and underflow is not due to \text{sumESP} per se as the exact values of the ESP itself is not expressible in double-precision. Issues of this kind can generally be avoided by scaling \( \eta \)’s appropriately for a certain regime.

### B. Runtime cost of an individual matrix element

The cost of calculating a single matrix element of \( n \)-pair RDM is bounded above by the cost of the norm of AGP, i.e. \( \langle AGP|AGP \rangle \). This is because all higher order RDMs require evaluation of lower degree ESPs by Eq. (20) and the cost of the prefactor is negligible. As such, to get the upper bound of the cost, we only report the elapsed time for evaluating the norm of AGP, which we refer to as the overlap. The theoretical cost of the overlap as a function of \( N \) and \( M \) grows as \( N(1 - N + M) - 1 \), which is the total number of iterations in the loops of the \text{sumESP} algorithm.

The left axes in Fig. 2 (the blue dots) illustrate the elapsed time of the overlap as a function of number of pairs, \( N \). Every point in the plot is the sample mean of \( 10^3 \) observations with \( \eta_p \sim \text{Unif}(0,1) \). By inspection, the most expensive iterations occurs when \( N \approx M/2 \) which is expected since \( \left(\frac{M}{N}\right) \) is maximum when \( N = M/2 \). The shorter elapsed times in \( N > M/2 \) is due to fewer summations in the algorithm as \( N \) approaches \( M \).

To find the asymptotic scaling with system size, \( M \), we fix \( N = M/2 \) and vary \( M \) from \( 10^2 \) to \( 10^3 \). This gives the most expensive elapsed time for every value of \( M \). The results are shown in Fig 3 in which every point is the sample mean of \( 10^3 \) observations with \( \eta_p \sim \text{Unif}(0,1) \). A linear fit to the log-log plot indicates that the asymptotic time scales quadratically, \( T[s] \propto M^{1.97} \), with the system size, in line with the theoretical result.

### C. Runtime cost of \( n \)-pair RDMs

Here, we report the maximum time needed for calculating all matrix elements of \( n \)-pair RDMs. At this juncture,
we remind the reader that it is sufficient to merely compute and store \( \gamma_{q_1 \ldots q_n}^{p_1 \ldots p_n} \). Since the calculations of the matrix elements are independent from each other, this is highly parallelizable. (See Appendix A for the pseudocode used to calculate the matrix elements.)

The theoretical cost of constructing an \( n \)-pair RDM on a single core is

\[
\text{cost}\{\text{n-pair RDM}\} \leq \binom{M}{n}^2 \text{cost}\{\langle AGP|AGP\rangle\},
\]

where \( \binom{M}{n}^2 \) is the number of matrix elements needed to construct an \( n \)-pair RDM with \( M \) levels. Recall that the asymptotic cost of the norm of AGP is \( O(N(M-N)) \).

In Fig. 4 we report the average elapsed time of \( n \)-pair RDM for \( n = 1, 2, \) and 3, with \( M \leq 100 \). The plots show the sample average of 100 observations for \( n = 1, 2 \) and 10 observations for \( n = 3 \) in which \( \eta \sim \text{Unif}(0,1) \).

IV. RECONSTRUCTION FORMULAE

In Sec. IIIIC we argued that the asymptotic cost of constructing an \( n \)-pair RDM is \( O(N M^{2n+1}) \). Here, we show a way of cutting down the cost by expressing higher order RDMs as a linear combination of lower order ones. This is an important ingredient for developing efficient correlated theories based on AGP, as most techniques require evaluation of high order RDMs. For example, in AGP based configuration interaction (AGP-CI) calculations done in our group up to 5-body density matrices were needed (the Hamiltonian and the correlators contain 2-body operators each, and using the killers one can introduce a commutator to reduce the rank by 1). Therefore, a systematic way of reducing the scaling and the cost of calculating many-body RDMs is of great interest. Rosina mathematically anticipated that the 2-RDMs of AGP determine all of its higher order RDMs. As it turns out, we here show that 1-RDM occupation numbers RDMs \( \nu^{(1)} \) and \( \eta \)'s are sufficient to determine all RDMs of AGP.

Note that the decomposition method presented here differs from the decomposition of PBCS in which the density matrix is a weighted integral over the projection grid of a factorized transition density matrix.\(^{12}\)

A. Direct decomposition

Our goal, here, is to express the nonzero elements of any \( n \)-pair RDM, \( \gamma^{(n)} \), as a linear combination of \( \nu^{(n')} \) where \( n \leq n' \leq 2n \); then we show that we can further decompose each \( \nu^{(n')} \) into a sum of \( \nu^{(1)} \), hence \( \gamma^{(n)} \rightarrow \ldots\).
Consider an $n$-pair RDM, $\gamma_{q_1\ldots q_n}^{p_1\ldots p_n}$. In general, it can be that $p_i = q_j$ for some $i, j$. Since $N_p = 2^\sum p_i P_p$, we can write
\[
\gamma_{q_1\ldots q_n}^{p_1\ldots p_n} = 2^{-k} \langle AGP|P_{p_1}\ldots P_{p_{n-k}}N_{r_1}\ldots N_{r_k}P_{q_1}\ldots P_{q_{n-k}}|AGP\rangle,
\]
where $k$ is the number of common indices among $\{p_i\}$ and $\{q_j\}$. Written in this manner, we can assume all the remaining indices are different; otherwise the element is zero by construction. By this and hermiticity of $\gamma^{(n)}$, and the fact that the top and lower indices are permutable, we can further express
\[
\gamma_{q_1\ldots q_n}^{p_1\ldots p_n} = \frac{1}{2^n} \langle AGP|N_{r_1}\ldots N_{r_k}(P_{p_1} P_{q_1} + P_{q_1} P_{p_1})
\ldots(P_{p_{n-k}} P_{q_{n-k}} + P_{q_{n-k}} P_{p_{n-k}})|AGP\rangle.
\]
Now, by manipulating the killer of AGP, i.e. $K_{pq}|AGP\rangle = 0$ reported in Ref.\textsuperscript{20} we can write
\[
P_{p} P_{q} + P_{q} P_{p} = \frac{1}{\eta_p^2 + \eta_q^2} (K_{pq}^\dag + K_{pq}) + \eta_p \eta_q (N_p + N_q - N_p N_q).
\]
By plugging Eq. (28) into Eq. (27) we arrive at an expression for $\gamma^{(n)}$ that is written purely as a linear combination of $\nu^{(1)}$ as desired. Now, we decompose $\nu^{(n)}$ into a sum of $\nu^{(1)}$ and factors of $\eta_n$. For this, we introduce a closed form expression which we prove in Appendix B
\[
\nu_{p_1\ldots p_n} = \frac{n}{\nu^{(1)} (n)} \left( \prod_{j=1}^{n} \frac{2\eta_{p_j}^2}{\eta_{p_j}^2 + \eta_q^2} \right) \nu_{p_i}.
\]
To make this discussion concrete, consider $\gamma^{(2)}$ as an example, whose non-zero elements can be written as
\[
\gamma_{r,s}^{pq} = \begin{cases} \langle P_{p} P_{q} P_{r} P_{s} \rangle & \text{if } p, q \neq r, s \\ \frac{1}{2} (P_{p} N_{r} P_{s}) & \text{if } p \neq r \text{ and } q = s \\ \frac{1}{2} (N_{p} N_{q}) & \text{if } p = r \text{ and } q = s \end{cases}
\]
For each of the three cases, we obtain an expression entirely in terms of the number RDMs and $\eta$’s using Eq. (28):
\[
\gamma_{r,s}^{pq} = \left( \frac{\eta_p \eta_q \eta_{rs}}{4(\eta_p^2 + \eta_q^2)(\eta_{rs}^2 + \eta_{rs}^2)} \right) \left( \langle N_{p} + N_{r} - N_{p} N_{r} \rangle \right. \\
\left. \frac{\eta_{p} \eta_{rs}}{4(\eta_p^2 + \eta_{rs}^2)} \langle N_{p} N_{q} \rangle + \langle N_{q} N_{r} \rangle - \langle N_{p} N_{q} N_{r} \rangle \right)
\]
Then each of the $\langle N_{p} N_{q} \rangle$, $\langle N_{p} N_{q} N_{r} \rangle$, etc., can be inserted in Eq. (29) to produce an expression in terms of a linear combination of $\langle N_{p} \rangle$ and $\eta$’s only, as desired.

This result is profound as it implies that we can obtain all higher rank pair RDMs by merely computing $\nu^{(1)}$ whose cost grows asymptotically as $O(M^2 N)$ and a prefactor of $n(\lambda - 1)$ due to the direct decomposition. (For all practical purposes $n \ll N, M$; thus the cost of the prefactor is manageable.) However, we pay the price of introducing $1/(\eta_p^2 - \eta_q^2)$ factors that can be numerically ill-posed when $\eta_p^2 \approx \eta_q^2$ or when $\eta_p, \eta_q \to 0$. In the regime that these factors are not problematic, the decomposition $\gamma^{(n)} \rightarrow \sum \nu^{(1)}$ is a major improvement over computing all the matrix elements of an $n$-pair RDM. We must note that in our own implementation of these equations for practical problems (e.g. the attractive pairing Hamiltonian in Ref.\textsuperscript{20}) we have not observed these potential numerical issues.

B. Stepwise decomposition

In practice, there are situations in which it is more advantageous to break down high rank RDMs in terms of “slightly” lower rank ones. This also makes the issue of having too many $1/(\eta_p^2 - \eta_q^2)$ factors less severe. To this
end, we need a new notation for $\langle P^1\ldots N\ldots P^r\rangle$. Define

$$Z^{(k,m)}_{\nu,pq,r,s,t} = \langle AGP|P^1_p,\ldots P^1_p|P^4_p,\ldots P^4_p|AGP\rangle,$$

where $k$ is the number of $N_p$ operators in the middle, and $m$ is the total number of indices such that $k \leq m$. And define $t = (m-k)/2$. For example,

$$Z^{(0,6)}_{pqrstu} = \langle AGP|P^1_p,\ldots P^3_p,\ldots P^3_p|P^4_p,\ldots P^4_p|AGP\rangle,$$

$$Z^{(1,3)}_{pqrst} = \langle AGP|P^1_p,\ldots P^4_p,\ldots P^4_p|AGP\rangle,$$

$$Z^{(2,4)}_{pqrstu} = \langle AGP|P^1_pN_rP_s|AGP\rangle,$$

$$Z^{(3,3)}_{pq} = \langle AGP|N_pN_r|AGP\rangle.$$

Notice that the subscript indices on each $Z^{(k,m)}$ are all different; otherwise they are either zero or reducible to some other $Z^{(k,m')}$ by $N_p = 2P^1_pP^3_p$ and $N_p = 2N_p$. Obviously all $\gamma^{(n)}$ can be mapped to $Z^{(k,2n-k)}$ for some $k$ and $n$, and vice versa.

Our goal, here, is to show that

$$Z^{(0,2n)} = \sum Z^{(1,2n-1)} \to \ldots \to \sum Z^{(n,n)} = \sum \nu^{(n)} \to \sum \nu^{(1)}.$$  

As we prove in Appendix B, this can be accomplished by using the following formula at every step

$$Z^{(k,m)}_{\nu,pq,r,s,t} \equiv \frac{\eta_{p\nu}^{(k)}\eta_{q\nu}^{(k)}(2\eta_{q\nu}^{(k)} - \eta_{p\nu}^{(k)})}{2(\eta_{q\nu}^{(k)} - \eta_{p\nu}^{(k)})}Z^{(k+1,m-1)}_{\nu,pq,r,s,t} - Z^{(k+1,m-1)}_{\nu,pq,r,s,t} - Z^{(k+1,m-1)}_{\nu,pq,r,s,t},$$

where $i, j \in \{1, 2, \ldots \}$. Similarly, it is easy to show that

$$\nu^{(n)} = \sum_{\nu} \eta^{(n)}_{\nu} \nu^{(n)}_{\nu},$$

where $i, j \in \{1, 2, \ldots \}$.

The advantage of breaking down the density matrices like this is that, at every step, we reduce the dimension by one, thereby reducing the asymptotic scaling of computing it by a factor of $M^2$ and a negligible prefactor. Moreover, we can stop at any step of our choice based on the cost that we are willing to tolerate.

V. RUNTIME OF ENERGY AND AGP-CI

In this section, we benchmark the speed gain when using the reconstruction formulae. Here, we report the runtime measurements of two calculations: (1) the energy of the pairing Hamiltonian (reduced BCS); (2) AGP-CI calculations as reported in Ref. 20. See Appendix C for the computer environments used in these benchmarks.

We start with the pairing Hamiltonian. Recall that the attractive pairing Hamiltonian can be written as follows:

$$H = \sum_p \epsilon_p N_p - G \sum_{pq} P^1_pP^1_q.$$  

The expected value of the energy over AGP in terms of the pair RDMs is

$$E = \sum_p \gamma^p_p(2\epsilon_p - G) - 2G \sum_{p<q}\gamma^p_p\gamma^q_p.$$  

And using the reconstruction formulae, we can get

$$E = \sum_{\nu} \nu^{(n)}(\epsilon_p - G/2) - G \sum_{p<q} \nu^{(n)}_{\nu_p}\nu^{(n)}_{\nu_q}(\nu_p - \nu_q).$$

Fig. 5 shows the elapsed time differences between implementing Eq. (37) and Eq. (38) as a function of number of levels at half-filling. For each case, the corresponding RDM is stored and then called in the calculation. By inspection, it is easy to see that the reconstruction formula speed up the calculations often by an order of magnitude. Obviously, the improvement becomes more noticeable as the number of levels increases.

At last, we report the elapsed time of performing AGP-CI calculations with and without the reconstruction formulae in Fig. 6. For these calculations, we stored up to 4-indexed RDMs in memory and calculated all the rank 5 RDMs using the direct decomposition formula Eq. (29). Here, $\eta$’s are optimized beforehand for the pairing Hamiltonian ($G = 2$) for various values of $M$. Similar to the energy calculations, we observe that the reconstruction formulae lead to a substantial improvement.
VI. CONCLUSIONS

Analytic expressions of the norm and reduced density matrices of AGP wavefunction are proportional to ESP. We used the \textit{sumESP} algorithm\textsuperscript{22,23} to efficiently calculate ESPs and argued that, with appropriate normalization, it is well suited for physical problems wherein \(N \ll M\). We have shown that our method can reliably calculate the norm and elements of RDMs to all ranks for systems as large as a few thousand orbitals and hundreds of electrons. Our runtime measurements indicate that the asymptotic cost per element of an \(n\)-pair RDM grows asymptotically as \(\mathcal{O}(NM^{2n+1})\).

However, to reduce the cost of computing the RDMs even further, we have derived reconstruction formulae that allow decomposition of any RDM into a linear combination of lower rank ones. We introduced two methods: (1) Direct decomposition—breaks down a high rank pair RDM in terms of linear combination of rank-1 occupation RDMs; (2) Stepwise decomposition—reduces the dimension of a RDM by one at every step thereby reducing the cost of computing it by a factor of \(M^2\).

We demonstrate the advantage of using our reconstruction formulae by benchmarking it against the energy of the pairing Hamiltonian and AGP-CI calculation without the reconstruction formulae. The numerical results indicate that, indeed, the reconstruction formulae lead to a substantial speed-up, especially in systems with large number of orbitals.

ACKNOWLEDGEMENTS

This work was supported by the U.S. National Science Foundation under Grant No. CHE-1762320; and by the Big-Data Private-Cloud Research Cyberinfrastructure MRI-award funded by NSF under grant CNS-1338099 and by Rice University. G.E.S. is a Welch Foundation Chair (Grant No. C-0036).

Appendix A: Numerical Implementation

Below is a pseudocode for calculating a single matrix element of an \(n\)-pair RDM. Implementation of \textit{sumESP} is taken from Ref.\textsuperscript{23}

\begin{algorithm}
\caption{Matrix elements of \(n\)-pair RDM}
\textbf{input} : \(x = \{\eta_1, \ldots, \eta_M\}\);
\begin{itemize}
  \item \(N\): number of pairs;
  \item \(L = \{p_1, \ldots, p_n, q_1, \ldots, q_n\}\)
\end{itemize}
\textbf{output} : \(\gamma_{\eta_1 \ldots \eta_n}\)
\begin{algorithmic}
  \State \textbf{integer } \(i, j, k, n, M\); \textbf{ real } \(p\)
  \State \textbf{integer array } \((L_i)_{1:2n}\), \((x_i)_{1:M}\), \((S^j_i)_{(i \in 1:M)(j \in 1:N+1)}\)
  \If {\((p_i = p_j \text{ or } q_i = q_j) \text{ or } (n > N)\)}
  \State Output = 0;
  \State End the program;
  \Else
  \State \(k \leftarrow N - \text{size}(L)/2; \quad \text{// size}(x)\) gives the size of \(x\)
  \State \(p \leftarrow (\prod_{i \in L} x_i);\)
  \State \(x \leftarrow \text{eliminate}(x_i \in L)\); \text{// removes } x_i \in L \text{ from } x\)
  \State \(x \leftarrow x^2\)
  \State \text{// Below is the } \textit{sumESP} \text{ algorithm}
  \State \(M \leftarrow \text{size}(x);\)
  \State \(S^j_0 \leftarrow \text{zeros for } j \geq i; S^j_0 \leftarrow 1 \text{ for } 1 \leq i \leq M - 1;\)
  \State \(S^j_1 \leftarrow x_1;\)
  \For {\(i = 2 \text{ to } M\)}
  \State \text{// for } \(j = \max\{1, i + k - n\} \text{ to } \min\{i, k\}\) \Do
  \State \(S^j_i \leftarrow S^j_{i-1} + x_i S^j_{i-1}\)
  \End\}
  \End\}
  \State \text{// End of } \textit{sumESP}\)
  \State Output \leftarrow p \times S^M_{k+1}\)
\end{algorithmic}
\end{algorithm}

One can store any RDM as a 2 dimensional array by linear indexing \(p_i\)'s and \(q_i\)'s such that \(\gamma_{\eta_1 \ldots \eta_n} \rightarrow \gamma'^\mu\). The reader may find the following relation handy in the parallel implementation

\[\mu = p_1 + \sum_{i=2}^{n} \binom{p_i - 1}{i}; \quad 1 \leq p_1 < \ldots < p_n \leq M.\]
Appendix B: Proofs of the reconstruction formulae

We set out by first deriving Eq. (34). Notice, by Eq. (20) and Eq. (23), we can write

\[ Z_{P_1 \cdots P_n}^{(k+1,m-1)} = 2^{k+1} \eta_{P_1} \cdot \prod_{s=1}^{k} \eta_{P_s} \left( \prod_{s \geq 1, j} \eta_{P_s \eta_{P_j}} \right) \]

\[ \left( \prod_{s=1}^{k} \eta_{P_s} \right) \left( \prod_{s \geq 1, j} \eta_{P_s \eta_{P_j}} \right) \]

\[ = \frac{1}{2} Z_{P_1 \cdots P_n}^{(k+2,m)} \eta_{P_1} \eta_{P_2} \cdots \eta_{P_n} \]

Similarly, we can state that

\[ Z_{P_1 \cdots P_n}^{(k+1,m-1)} = \frac{1}{2} Z_{P_1 \cdots P_n}^{(k+2,m)} \eta_{P_1} \eta_{P_2} \cdots \eta_{P_n} \]

By subtracting Eq. (B2) from Eq. (B1) and rearranging the terms we get Eq. (34) as desired.

Now we prove Eq. (29) by induction on \(n\). For \(n = 1\), the equality is trivially true. Now, given the induction hypothesis for \(n\), we want to show the \(n+1\) case. Similar to the derivation above, by using Eq. (20) and Eq. (23), we can write

\[ \nu_{P_1 \cdots P_n} = \frac{2 \eta_{P_1} \eta_{P_2} \cdots \eta_{P_n}}{\eta_{P_1} \eta_{P_2} \cdots \eta_{P_n}} \]

Using Eq. (34) and rearranging the terms we get

\[ \nu_{P_1 \cdots P_{n+1}} = \frac{2 \eta_{P_1} \eta_{P_2} \cdots \eta_{P_n}}{\eta_{P_1} \eta_{P_2} \cdots \eta_{P_n}} \nu_{P_1 \cdots P_n} \]

Now, we apply the induction hypothesis to \(\nu_{P_1 \cdots P_n}\) and \(\nu_{P_2 \cdots P_{n+1}}\). Brute force algebra shows that

\[ \nu_{P_1 \cdots P_{n+1}} = \left( \prod_{j=2}^{n+1} \eta_{P_j} \right) \nu_{P_1} + \left( \prod_{j=1}^{n} \eta_{P_1} \eta_{P_j} \right) \nu_{P_{n+1}} \]

\[ + \sum_{i=2}^{n+1} \left( \prod_{j=1}^{i} \eta_{P_j} \right) \left( \eta_{P_i} \right) \left( \eta_{P_{i+1}} \right) \nu_{P_{n+1}} \]

\[ = \prod_{j=1}^{n+1} \eta_{P_j} \nu_{P_1} \]

Appendix C: Computer environments

Here we detail the computer environment used for each runtime test.

Sec. III B and the energy calculations in Sec. V: single core of a workstation with Intel Xeon(R) CPU E3-1270 v6, with 8 cores, each at 3.50GHz on a x86_64 hardware architecture with GNU/Linux operating system. The programs were compiled using GNU Fortran (GCC) 4.8.5 (Red Hat 4.8.5-36) with the default compiler optimization options.

Sec. III C: the computations were carried out in parallel using 16 cores on a single node of a cluster running on the Rice Big Research Data (BiRD) cloud infrastructure. The environment is as follows: Intel(R) Xeon(R) CPU E5-2650 v2 @ 2.60GHz with 16 cores on a x86_64 hardware architecture with GNU/Linux operating system. The compiler is GNU Fortran (GCC) 4.8.5 20 (Red Hat 4.8.5-28) with the default optimization flags.

AGP-CI calculations in Sec. V: performed in parallel on a workstation with Intel Xeon(R) CPU E3-1270 v6, with 8 cores, each at 3.50GHz on a x86_64 hardware architecture with GNU/Linux operating system. PGI-15 compiler with the following flags: ":O4 -Mvect -Mprefetch -Mconcur=allcores -Mcache_align -fast -fastsse4."
M. Degroote, T. M. Henderson, J. Zhao, J. Dukelsky, and G. E. Scuseria, Phys. Rev. B 93, 125124 (2016).
19 Y. Qiu, T. M. Henderson, T. Duguet, and G. E. Scuseria, Phys. Rev. C 99, 044301 (2019).
20 T. M. Henderson and G. E. Scuseria, J. Chem. Phys. 151, 051101 (2019).
21 G. H. Fischer, Einführung in die Theorie psychologischer Tests [Introduction to the theory of psychological tests] (Bern Huber, 1974).
22 R. Rehman and I. Ipsen, SIAM J. Matrix Anal. Appl. 32, 90 (2011).
23 H. Jiang, S. Graillat, R. Barrio, and C. Yang, Applied Mathematics and Computation 273, 1160 (2016).
24 B. Weiner and O. Goscinski, Phys. Rev. A 22, 2374 (1980).
25 J. V. Ortiz, B. Weiner, and Y. Öhrn, International Journal of Quantum Chemistry 20, 113 (1981).
26 J. Cioslowski, ed., Many-Electron Densities and Reduced Density Matrices, Mathematical and Computational Chemistry (Springer US, 2000).
27 R. Gilmore, Lie Groups, Physics, and Geometry: An Introduction for Physicists, Engineers and Chemists (Cambridge University Press, 2008).
28 F. B. Baker and M. R. Harwell, Applied Psychological Measurement 20, 169 (1996).
29 V. N. Staroverov and G. E. Scuseria, J. Chem. Phys. 117, 11107 (2002).