

Accelerated Purification Using Generalized Nonpurifying Intermediate Functions for Large-Scale Self-Consistent Field Calculations

Jaehoon Kim and Yousung Jung*

Graduate School of EEWS (WCU) and KI for Nanocentury, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 305-701, Korea

ABSTRACT: Purification is a widely used technique to calculate idempotent density matrices from a Hamiltonian in large-scale electronic structure calculations. However, the initial guess of a density matrix usually contains large errors, which require many iterations to remove them, using standard recursive schemes such as those derived by McWeeny or Holas. In this Letter, we propose a way to obtain a converged density matrix much more rapidly by removing the stability conditions that the functions have fixed points and vanishing derivatives at 0 and 1, assumptions usually made in most traditional purification methods. That is, by extending the recursive function space, which gives the approximated step function via the generalized nonpurifying intermediate functions, and optimizing them, we reduce the purification cost approximately by a factor of 1.5 compared to grand canonical purification algorithms for the linear alkanes, diamondoid, and a protein endothelin that has a very small band gap.

1. INTRODUCTION

In self-consistent field calculations such as Hartree–Fock (HF) or density functional theory (DFT), the overall computational cost is determined by the formation of the effective Hamiltonian matrix and calculation of the new density matrix from it. For the formation of Fock matrix, many efficient linear scaling algorithms have been developed, for example, continuous fast multipole methods (CFMM) and Fourier transform Coulomb (FTC) methods for Coulomb interactions^{1–6} and LinK, ONX, and multipole-accelerated methods for the HF exchange matrices.^{7–12} For an effective Hamiltonian in DFT, efficient linear scaling numerical quadrature methods for the exchange–correlation potential have also been developed.^{13,14}

The other important step in self-consistent field (SCF) calculations, updating the density matrix using the previous Fock matrix just constructed, is then usually done by diagonalizing the Fock matrix with $O(N^3)$ cost. For small- to medium-sized systems, diagonalization is not a major bottleneck due to a small prefactor of the cubic scaling. For this reason, diagonalization is a preferred choice of density update in almost all quantum chemistry calculations employing localized Gaussian basis today. For large systems, however, the cubic scaling of diagonalization becomes important and, eventually, completely dominates the overall computational cost due to a higher scaling compared to the Fock build.

Linear scaling diagonalization-free algorithms have been developed. Since there are only $O(N)$ significant matrix elements in \mathbf{P} for large enough systems, direct minimization of the energy functional with respect to the density matrix elements leads to a linear scaling algorithm.¹⁵ Alternatively, one can use the fact that the density matrix (\mathbf{P}) commutes with the Hamiltonian (\mathbf{H}) and so can be a direct function of \mathbf{H} (e.g., polynomials). In this, one essentially expresses the density matrix as a Heaviside step function of the Hamiltonian matrix:

$$\mathbf{P} = h(\mu\mathbf{I} - \mathbf{H}) \quad (1)$$

where μ is the chemical potential, \mathbf{I} is the identity matrix, and $h(\cdot)$ is Heaviside step function. One can then approximate this Heaviside step function as Fermi–Dirac¹⁶ or complementary error functions¹⁷ that can be evaluated efficiently with the Chebyshev polynomials of the Hamiltonian matrix of finite order.¹⁸ This expansion and computation only requires matrix–matrix multiplications (matmuls) and additions which can be performed with linear scaling efforts due to the locality and sparsity of the matrices involved. In addition to the direct computation of eq 1 using matrix polynomials of very high order, one can also use some recurrence formula, such as that of McWeeny¹⁹ as in eq 2, to purify an appropriately chosen initial density matrix to satisfy idempotency and ensure occupation numbers of 0's or 1's to approximate eq 1:

$$\mathbf{P}_{n+1} = 3\mathbf{P}_n^2 - 2\mathbf{P}_n^3 \quad (2)$$

We determine below that the different diagonalization-free algorithms described above (minimization, expansion, or purification) have different convergence behaviors. For the Fermi operator expansion (FOE) method, error decreases as $O(\epsilon^m)$, where m is the number of required matmuls related to the degree of polynomial, while for an improved FOE algorithm with a fast summation technique, the error decreases more quickly as $O(\epsilon^{m^2/4})$. We find a similar convergence rate to the FOE method, i.e., $O(\epsilon^m)$ for some c , derived for the Curvy-step (CS) method²⁰ since it uses Baker–Hausdorff expansion, which is also a type of matrix polynomial expansion, but with a reduced cost due to a better initial guess. For purification algorithms, if we write the density matrix at the n th step as $\mathbf{P}_n = \mathbf{P} + \epsilon_n$ where \mathbf{P} is the exact density matrix, then $\mathbf{P}_{n+1} = \mathbf{P} + (3\mathbf{I} - 6\mathbf{P})\epsilon_n^2 + O(\epsilon_n^3)$ with \mathbf{P} and ϵ_n commuting, where $O(\epsilon_n^3)$ indicates a matrix whose elements are on the same order of magnitude as those of ϵ_n^3 (see appendix B in ref 21). With the initial guess of density matrix specially chosen as eq 3, the absolute value of eigenvalues of ϵ_n is always lower than 0.5, and thus $\mathbf{P}_\infty = \mathbf{P}$. This initial guess \mathbf{P}_0 of eq 3 is

Published: October 28, 2011

specially designed and scaled for the eigenvalues to lie between the $[0,1]$ interval in reverse order as part of the purification procedure to calculate the matrix step function. Therefore, the $\text{Tr}(\mathbf{P}_0)$ does not mean that it is the true number of electrons in the system, but rather will be converged to the correct number of electrons during the purification iterations.

$$\mathbf{P}_0 = \frac{\lambda}{2}(\mu\mathbf{I} - \mathbf{H}) + \frac{1}{2}\mathbf{I} \quad (3)$$

$$\lambda = \min\left(\frac{1}{H_{\max} - \mu}, \frac{1}{\mu - H_{\min}}\right) \quad (4)$$

where H_{\max} and H_{\min} are the upper and lower bounds of the eigenvalue spectrum of the Hamiltonian.²¹ The errors in the purification algorithm of McWeeny, for example, decrease as $\mathbf{P} - \mathbf{P}_n = O(\epsilon^{2^n}) = O(\epsilon^{2^{m/2}})$.

The computational cost to achieve a given accuracy can then be written as

$$\text{cost} = O(f^{-1}(\text{thresh})) \quad (5)$$

where the error is $O(\epsilon^{f(m)})$ and thresh defines the accuracy with $\text{maxerror} = 10^{-\text{thresh}}$. For the curvy-step and FOE methods, the cost is linearly dependent on thresh. For an improved FOE method with the fast summation technique, $\text{cost} = O((\text{thresh})^{1/2})$, and for purification methods, $\text{cost} = O(\log(\text{thresh}))$. Consideration of these different scaling properties of error reduction for various density matrix computation methods indicates that purification is a computationally more efficient way to calculate the density matrix from the Hamiltonian to achieve a given accuracy compared to other indirect diagonalization approaches. A similar and more detailed comparison of various density matrix calculation methods has also been given recently, leading to a similar conclusion that the recursive polynomial expansion is generally more efficient than other minimization methods.²² This Letter is about reducing the computational cost of existing purification algorithms even further by lifting one of the usual assumptions traditionally made.

In purification, one can of course go beyond third-order in eq 2 for faster convergence (fewer iterations) toward the true density matrix. When using the higher-order functions, however, one must consider the cost to compute the polynomial itself at each iteration. Recently, we have shown²³ that the optimal purification scheme is obtained when the fifth-order polynomial of Holas²⁴ is repeatedly applied throughout all iterations, and either the use of McWeeny or higher-order polynomials throughout or the mixed use of different-order polynomials at different iterations results in higher cost.

Regardless of the forms and degrees of polynomials employed in the purification algorithm, however, nearly all traditional purification methods generally assume a convention that at each iteration the polynomials are required to purify the density matrix; that is, the polynomials have stable fixed points at 0 and/or 1. However, this procedure of using the true purification functions at all iterations is not a necessary condition since only the final converged density matrix needs to be idempotent. In other words, if some generalized polynomials, even if they are nonpurifying functions, can reduce the large errors in density matrix more rapidly than the traditional true purification polynomials, one can consider using these nonpurifying functions during early and intermediate stages of iterations and apply the true purification functions, such as those of McWeeny or Holas, only at the last stages to ensure idempotency.

Table 1. Optimized Parameters for Nonpurifying Intermediate Functions Defined in eq 6 for Each Iteration in OP₅^a

	iteration 1	iteration 2	iteration 3	iteration 4
a_1	16.05230180	18.01787054	10.80499078	7.284154647
a_2	0.486483492	0.394957751	0.417919928	0.315415352
c	0.3515			

^a c in the last row determines the usable boundary of input eigenspectrum defined in eqs 7–9.

We call, in this paper, these generalized nonpurifying functions specially designed and optimized to accelerate the convergence “intermediate functions” (MFs), to be distinguished from a true purification function (PF). To this end, we here suggest an optimized purification (OP) algorithm that can reduce the total purification cost in terms of matmuls approximately by a factor of 1.5 by combining the generalized nonpurifying MFs and purifying PFs, as compared to the conventional schemes that usually use PFs repeatedly until the convergence is reached, such as canonical purification (CP) or grand canonical purification (GCP).

2. THEORY

The key aspect of our approach is that we remove the stability condition that the functions have stable fixed points at 0 and 1 at each iteration, and we adjust functions differently for each iteration. That is, we extend the recursive function space, which gives the approximated step function via the parametrized functions. We restrict the function space to be explored using three constraints:

1. maintaining symmetry with respect to the point (0.5,0.5)
2. using the fifth-order polynomials throughout all iterations
3. passing through the two fixed points (1,1) and (0,0).

We use constraint 1 to conserve symmetry during iterations since the shifted step function has a symmetry around (0.5,0.5) and the number of matmuls required to evaluate the functions can be reduced due to symmetry. The reduction in cost due to symmetry arises since the polynomial with symmetry can be expanded with either an odd or even function.²⁴ Constraint 2 is imposed to minimize the computational cost of polynomial evaluation, motivated by our recent analytical proof²³ that the repeated application of the fifth-order Holas polynomial throughout all iterations is an optimal scheme for the density matrix purification when polynomials have stable fixed points at 0 and 1. Here, we extend this theorem to the nonpurifying polynomials. Constraint 3 is needed to ensure that, once purified to 0's and 1's, they remain as 0's and 1's even with the application of nonpurifying functions. Since all of the MFs must be applied in sequence as one “quanta” in OP, without constraint 3, application of many nonpurifying functions may yield the matrix that diverges.

The functional form that satisfies all three conditions described above is

$$f(x, a_1, a_2) = x + a_1 x(x-1) \left(x - \frac{1}{2} \right) (x^2 - x - a_2) \quad (6)$$

Since we have only two adjustable parameters, a_1 and a_2 , we search for the optimized functions at each iteration directly, as described below. If the fifth-order Holas purification process

1. Input : Hamiltonian (**H**)
2. Upper and lower bounds of the eigenspectrum of Hamiltonian
3. Calculate Initial guess,

$$\mathbf{P}_0 = \frac{\lambda}{2}(\mu\mathbf{I} - \mathbf{H}) + \frac{1}{2}\mathbf{I} \text{ with } \lambda = \min\left(\frac{1.7030}{H_{\max}-\mu}, \frac{1.7030}{\mu-H_{\min}}\right)$$

4. Compute the intermediate functions in sequence, i.e.,

$$\mathbf{P}_4 = F_4\left(F_3\left(F_2\left(F_1(\mathbf{P}_0)\right)\right)\right).$$

5. If (convergence I)

If (convergence II)

Stop (Purification completed)

Else

Purify density matrix using the true purification function.

$$\mathbf{P}_{k+1} = \mathbf{P}_k^3 [10 - 15\mathbf{P}_k + 6\mathbf{P}_k^2], k \geq 4$$

6. Else go to step 4 (\mathbf{P}_0 is replaced by \mathbf{P}_4)

$$\text{Convergence criterion I : } c_1 = \frac{\|\mathbf{P}_4 - \mathbf{P}_0\|_F^2}{\sqrt{N}} = \sqrt{\frac{\text{Tr}((\mathbf{P}_4 - \mathbf{P}_0)^2)}{N}} < 10^{-2}$$

$$\text{Convergence criterion II : } c_2 = \max[(\mathbf{P}_{k+1} - \mathbf{P}_k)_{ij}] < 10^{-9}$$

Figure 1. Flowchart of the optimized purification (OP) algorithm. $\|\cdot\|_F$ is a Frobenius norm defined as $\|\mathbf{A}\|_F = (\text{Tr}(\mathbf{A}^T \mathbf{A}))^{1/2}$, where N is a dimension of the density matrix and $\text{Tr}()$ denotes the trace of a matrix.

were indeed optimal as in ref 23, one should recover $(a_1, a_2) = (6, 1/3)$ for all iterations in this extended function space when the parameters are optimized to produce the best approximate step function. The fact that the optimal values of (a_1, a_2) in eq 6 turn out to be different from those of Holas (as shown in Table 1) demonstrates that Holas' polynomial is not optimal if we generalize the function space to nonpurifying functions during iterations. A substantial improvement in convergence rate is indeed obtained with this generalization, as shown below. We abbreviate the proposed optimized purification scheme using the fifth-order polynomials as OP₅.

3. OPTIMIZATION

The number of intermediate purification functions (MFs), M , is chosen considering three factors: (i) The more efficient sequence of polynomials may be obtained with the larger number of MFs that are optimized; however, (ii) since all of the optimized MFs must be applied in order as one packet or one "quanta" (see algorithmic flowchart), the smaller number of M is more flexible to apply, and (iii) optimization of the parameters themselves becomes exponentially more difficult as M increases. On the basis of this trade-off between the optimized performance and flexibility, and also since a purification process for calculating a density matrix typically requires at least four iterations, we set the number of optimized MFs to be $M = 4$. In principle, one can go beyond this, if necessary. We comment on this aspect at the end of the present Letter.

The region of adjustable parameter space explored is set to be inside a rectangular space whose edges are $(-6, 0)$ and $(18, 2/3)$. This parameter space has a center at $(6, 1/3)$ which recovers Holas' fifth-order formula. To find the global minimum for (a_1, a_2) using the error function defined below, we explored the entire parameter space using a dual-resolution, cruder spacing $(1.5, 1/9)$ and a finer spacing $(10^{-10}, 10^{-10})$. All $(17 \times 7)^4$ possibilities for

$M = 4$ were considered to find the local minima at the lower resolution, which were then subsequently searched for a global minimum with a finer grid.

We define the error function to determine the optimal parameters as follows:

$$e[F(x), \alpha] = \frac{\int_{0.5}^{1+c} |F(x) - h(x - 0.5)|^\alpha dx}{0.5 + c} \quad (7)$$

where $F(x)$ is a generalized nonpurifying intermediate function (MF) that reduces the errors in the density matrix rapidly, i.e.,

$$F_0(x) = x \quad (8a)$$

$$F_i(x) = f(F_{i-1}(x), a_1(i), a_2(i)) \text{ for } i = 1 \sim 4 \quad (8b)$$

$$F(x) = F_4(x) \quad (8c)$$

The α can be any positive real number, and c is a parameter that extends the purification region beyond $[0, 1]$ and is determined by the minimal positive number that satisfies

$$\min_c : [F(1 + c) - 1]^2 > z \quad (9)$$

In eq 9, z is a tolerance parameter that sets a small allowed variation of occupancy around the exact occupancy, 0 or 1. We used $z = 0.003$. For α , $\alpha = 2$ may be the most natural choice in a least-squares sense; however, we regarded this exponent of the error function also as a variable for a finer optimization. By generating uniformly distributed random numbers for α in the interval $[1, 3]$ and finding an α that minimizes the error function in eq 7, we obtained and used $\alpha = 1.7199$.

The desirable properties of the optimized MFs would include (i) a large slope near the chemical potential, (ii) a wideness of eigenvalue spectrum (input) that can be covered by the chosen intermediate functions, and (iii) a narrowness of the eigenvalue spectrum (output) of the processed density matrix. The proposed error function, eq 7, is designed to meet these three conditions as much as possible. First, for i, the larger slope at the chemical potential would better approximate the step function and, hence, would require a smaller number of iterations (in particular for smaller band gap problems). Second, for ii, since $F(\mathbf{P})$ is a polynomial of \mathbf{P} , its approximation of the step function is only applicable for a limited region of the eigenvalue spectrum, and the parameter c in eq 9 determines this boundary. In our approach, the approximate step function has an extended usable interval $[-c, 1 + c]$ instead of $[0, 1]$.

4. RESULTS

The optimized nonpurifying MFs are summarized in Table 1, and the proposed accelerated purification algorithm is summarized in Figure 1 as a flowchart. Purification begins by applying the optimized nonpurifying MFs to the appropriately chosen initial density matrix (four iterations in one sweep) followed by the first convergence check. If the errors are still large, e.g., if the Frobenius norm of the difference between the density matrix and its square is more than $N \times (10^{-2})^2$, where N is a dimension of the density matrix, one goes back and keeps applying multiple sweeps of MFs until the criterion of convergence I is met. Since the functions are designed to make the spectral norm of errors for the density matrix near 10^{-2} , the repeated iteration of our intermediate functions must meet convergence I. Once

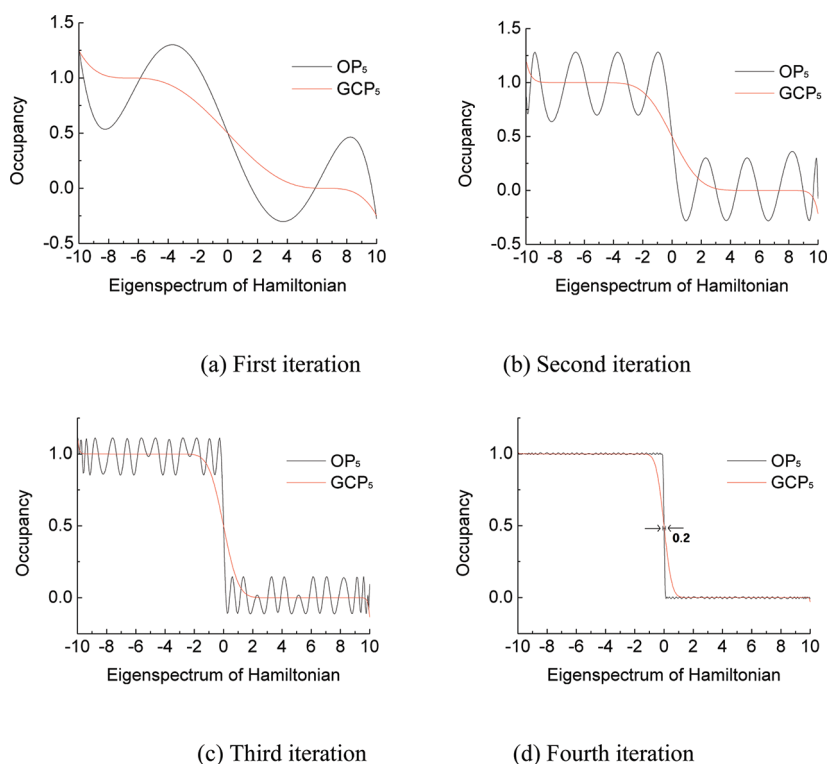


Figure 2. Efficiency of the OP scheme (OP_5) proposed here using the sequence of optimized fifth-order polynomials, compared to that of GCP using Holas' fifth-order function (GCP_5), along the first four iterations. The x -axis is the eigenspectrum of the Hamiltonian for a toy system with the spectral width $\Delta\epsilon = 20$ and the chemical potential $\mu = 0$.

convergence criterion I is met, one then switches to a standard true purification function, i.e., the fifth-order Holas polynomial, and keeps applying the same Holas function until convergence criterion II ($<10^{-9}$) is satisfied.

The numerical results for the evenly distributed eigenvalue spectrum of the Hamiltonian matrix in $[-10, 10]$ are shown in Figure 2, where we compare OP_5 and GCP_5 , where the subscript 5 means that the fifth-order nonpurifying or Holas polynomials are used, respectively. At first, OP_5 appears to have larger errors than GCP_5 during the initial 1–3 iterations, but these nonpurifying MFs in OP_5 play an important role in the rapid reduction of large errors so that the subsequent intermediate function stabilizes small errors more uniformly and effectively. As a result, one can visually see a clear improvement of OP_5 over GCP_5 after four iterations (1 sweep of OP_5). The net effect of applying these generalized optimized functions (that are not intended to purify the density matrix) is a small wiggling of the occupation numbers near 0 and 1, which needs to be purified subsequently. Usually, 1–3 applications of the true purification function are found to be sufficient to reduce the maximum error below the target criterion II.

The proposed OP method has a few advantages over the existing purification schemes using the true purification functions. First, the valid purification region for the fifth-order Holas polynomial is formally $[-0.259, 1.259]$, although traditionally only the limited interval $[0, 1]$ was utilized in density matrix calculation. In OP, the applicable range of occupation numbers is wider, $[-0.3515, 1.3515]$. Second, in a graph similar to Figure 2, the slope or the width of a “rising region” at the chemical potential can serve as an indicator of accuracy for the approximate step function. Here, we use the latter, i.e., a minimal range

where the occupancy is between 10^{-2} and $1 - 10^{-2}$. Under this measure, we reach the width 0.5 ± 0.009 after four iterations in the case of OP, which is effectively 0.5 ± 0.005 when the initial guess is scaled properly. Holas' fifth-order purification, however, yields the effective width of the rising region to be 0.5 ± 0.052 (about a factor of 10 larger than OP) after four iterations and takes eight iterations total to obtain the same accuracy as OP (width being ± 0.005).

We implemented the proposed OP algorithm into the development version of Q-CHEM²⁵ and assessed its practical performance for chemical systems, namely, linear alkane and diamond structures, using the BLYP density functional calculations at the STO-3G and 6-31G** bases. The symmetric transformation, i.e., $S^{-1/2}$, where S is overlap matrix, was used for orthogonalization of the atomic orbital (AO) basis. In Tables 2 and 3, we compare the performance of OP and other expansion and purification methods in the literature in terms of the average number of matmuls required per SCF cycle. For the present calculations, we used the superposition of atomic density for an initial guess and Pulay's DIIS algorithm with five subspaces for the SCF algorithm. As shown in Tables 2 and 3, the OP algorithm is faster than the existing expansion or purification methods by roughly a factor of 1.5.

To test the performance of OP for smaller band gap systems, we additionally considered a protein called endothelin that has 328 atoms (or 171 non-hydrogen atoms). For comparison, $C_{240}H_{482}$ (alkane) and $C_{100}H_{100}$ (diamondoid) have HOMO–LUMO (highest occupied molecular orbital–lowest unoccupied molecular orbital) gaps of 0.877 and 0.310 hartree, respectively, at the BLYP/STO-3G level, while endothelin has a HOMO–LUMO gap of 0.039 hartree at the same level.

Table 2. Average Number of Matmuls Needed to Compute the Density Matrix Per SCF Cycle for Alkane Chains of Varying Size^a

method	STO-3G				6-31G**
	C ₆₀ H ₁₂₂	C ₁₂₀ H ₂₄₂	C ₁₈₀ H ₃₆₂	C ₂₄₀ H ₄₈₂	C ₆₀ H ₁₂₂
CP (ref 20)	33 (6)	33 (5)	33 (5)	33 (5)	64 (7)
CS (ref 20)	34 (7)	34 (7)	35 (7)	35 (7)	49 (9)
FOE (ref 17)	27	27	27	27	47
GCP ₃	30 (6)	30 (5)	30 (5)	30 (5)	40 (7)
GCP ₅	30 (6)	30 (5)	30 (5)	30 (5)	33 (7)
OP ₃	23 (6)	23 (5)	23 (5)	23 (5)	31 (7)
OP ₅	22 (6)	22 (5)	22 (5)	22 (5)	25 (7)

^aFor GCP and OP, the average number of matmuls per SCF cycle is calculated to be the same as the number of matmuls for the last SCF cycle for these molecules. BLYP/STO-3G or BLYP/6-31G** calculations were performed, and the numbers in parentheses are the total number of SCF cycles taken for SCF convergence. For OP, the extra matmuls to calculate the Frobenius norm in convergence criterion I are also taken into account.

Endothelin also has a spectral width of 87.8 hartree, yielding $\xi/\Delta\epsilon = 0.0004$ at the BLYP/STO-3G level, which exceeds the $\xi/\Delta\epsilon = 0.01$ ratio for which OP is originally designed. For this small-gap protein, we find that the number of matmuls required for the purification of the last SCF cycle is 35 for OP₅, still yielding a significant speedup by a factor of 1.6 when compared with 54 matmuls for both GCP₅ and GCP₃. Therefore, it demonstrates the key advantage of OP for a fast error reduction during the early iterations of purification regardless of the band gap. Switching to a true purification function (Figure 1) such as Holas' then allows OP to handle molecular systems whose band gaps are beyond the $\xi/\Delta\epsilon = 0.01$ limit.

To numerically assess the previous mathematical proof (and current extension without proof) that the fifth-order polynomial is optimal for purifying the density matrix, we compared the performance (the number of matmuls taken to achieve a target accuracy) of the OP algorithm proposed here with degrees 3 and 5. For a fair comparison with OP₅ where the four fifth-order MFs were optimized, we used six MFs for the third-order polynomials in OP₃, noting that the application of the four fifth-order MFs and six third-order MFs requires the same number of matmuls to evaluate, namely, 12. The optimized parameters for OP₃ are summarized in Table 4. We used the same optimization scheme to find the optimal third-order polynomials as in the fifth-order functions. The results in Tables 2 and 3 show that the cost of matrix computations for OP₅ is slightly lower than that of OP₃ for selected alkane chain and diamond structures. The number of matmuls for the GCP₃ calculation (using McWeeny) is also compared with that of GCP₅ (using the fifth-order Holas polynomial), whose results show that GCP₅ is indeed marginally more efficient than GCP₃. However, the result for a diamond structure at the STO-3G basis, where GCP₅ required one more matmul than GCP₃ on average, is perhaps due to a fact that three matmuls must be used to evaluate the fifth-order Holas polynomial, while two are required in the case of the McWeeny polynomial.

In the final stage of completing this manuscript, we became aware of Rubensson's scaled purification (SP) method,²⁶ which performs impressively compared to existing purification

Table 3. Average Number of Matrix Multiplications Per SCF Cycle for H-Terminated Diamond Structure C₁₀₀H₁₀₀^a

method	STO-3G	6-31G**
GCP ₃	32 (7)	45 (9)
GCP ₅	33 (7)	43 (9)
OP ₃	23 (7)	37 (9)
OP ₅	22 (7)	35 (9)

^aBLYP/STO-3G or BLYP/6-31G** calculations were performed. For OP, the extra matmuls to calculate the Frobenius norm in convergence criterion I are also taken into account. The average number of matmuls per SCF cycle is calculated to be the same as the number of matmuls for the last SCF cycle for this molecule.

Table 4. Optimized Nonpurifying Intermediate MFs at Each Iteration in OP₃ That Consists of Six Third-Order Polynomials^a

	iteration 1	iteration 2	iteration 3	iteration 4	iteration 5	iteration 6
<i>a</i>	−0.51788	−5.23743	−7.90017	−2.83308	−6.06444	−1.51465

^aThe functional form takes $f(x,a) = a(x-1)(x-1/2)x + x$.

methods, and we comment on it here. The main idea in SP is to stretch out the eigenspectrum of the initial density matrix and fold back to the desired [0,1] interval to accelerate the error reduction. To compare SP (algorithm 1 of ref 26) and OP, proposed here, we considered a toy model with $\xi/\Delta\epsilon = 0.01$ (where $\Delta\epsilon$ is a spectral width of the Hamiltonian and ξ is a band gap). This band gap is chosen since it is the affordable band gap limit of OP. We compared the errors of OP₅ after four iterations and SP after six iterations, which require the same computational work. OP₅ has an error $\|P - P_4\|_2 \approx 0.008$ after four iterations, where $\|\cdot\|_2$ is the spectral norm, while the SP algorithm shows more wiggles with $\|P - P_6\|_2 \approx 0.008$ after six iterations. Of course, it should also be noted that, due to a greater flexibility of the SP algorithm compared to OP, which must be iterated with the multiples of *m*, i.e., 4 in the present case, the relative performance in practice must yet be seen. The practical performance comparison will also depend on the band gap of a particular system of interest with a certain spectral width. For example, for a smaller gap ratio ($\xi/\Delta\epsilon = 0.005$), the SP algorithm has a smaller error (0.091) than OP (0.165), although it is the beyond the affordable band gap ratio for which OP is designed. In principle, one can systematically design a customized set of optimized intermediate functions in OP for different band gap problems by varying the number of MFs and reoptimizing them depending on the band gap.

5. CONCLUSIONS

We have shown that the computational cost of approximating a matrix step function in density matrix calculations can be substantially reduced by optimizing the recursive polynomials used at each iteration differently in an extended function space. In particular, we proposed the use of generalized nonpurifying intermediate functions that are optimized to reduce large errors in the initial density matrix rapidly, which are then subsequently purified to exact idempotency using a usual purification formula, such as the fifth-order Holas polynomial. Numerical assessments using density functional calculations for the linear alkane,

diamond, and a protein endothelin demonstrate that the cost of purification is reduced by a factor of 1.5 in terms of the total number of matmuls by using the proposed OP scheme as compared to the CP or GCP algorithms, regardless of the band gap. The optimal number of intermediate functions, e.g., four for OP_s, however, may depend on the band gap of a problem of interest. Although our optimized purification is developed mainly for a fast computation of the matrix step function from a given Hamiltonian, since purification is a widely used technique in electronic structure theory in various contexts, exploring its advantage in contexts such as in theories that use purification to guide unconstrained density matrix minimization²⁷ may be an interesting subject of future study.

AUTHOR INFORMATION

Corresponding Author

*E-mail: ysjn@kaist.ac.kr.

ACKNOWLEDGMENT

We are pleased to acknowledge the support of Basic Science Research (2010-0023018) and WCU (R-31-2008-000-10055-0) programs funded by the Ministry of Education, Science, and Technology of Korea. We acknowledge the generous computing time from KISTI (KSC-2011-G1-01).

REFERENCES

- (1) White, C. A.; Johnson, B. G.; Gill, P. M. G.; Head-Gordon, M. The continuous fast multipole method. *Chem. Phys. Lett.* **1994**, *230*, 8–16.
- (2) Strain, M. C.; Scuseria, G. E.; Frisch, M. J. Achieving Linear Scaling for the Electronic Quantum Coulomb Problem. *Science* **1996**, *271*, 51–53.
- (3) White, C. A.; Johnson, B. G.; Gill, P. M. G.; Head-Gordon, M. Linear scaling density functional calculations via the continuous fast multipole method. *Chem. Phys. Lett.* **1996**, *253*, 268–278.
- (4) White, C. A.; Head-Gordon, M. Rotating around the quartic angular momentum barrier in fast multipole method calculations. *J. Chem. Phys.* **1996**, *105*, 5061.
- (5) Challacombe, M.; Schwegler, E. Linear scaling computation of the Fock matrix. *J. Chem. Phys.* **1997**, *106*, 5526–5536.
- (6) Shao, Y.; Head-Gordon, M. An improved J matrix engine for density functional theory calculation. *Chem. Phys. Lett.* **2000**, *323*, 425–433.
- (7) Schwegler, E.; Challacombe, M. Linear scaling computation of the Hartree-Fock exchange matrix. *J. Chem. Phys.* **1996**, *105*, 2726–2734.
- (8) Burant, J. C.; Scuseria, G. E.; Frisch, M. J. A linear scaling method for Hartree–Fock exchange calculations of large molecules. *J. Chem. Phys.* **1996**, *105*, 8969.
- (9) Schwegler, E.; Challacombe, M.; Head-Gordon, M. Linear scaling computation of the Fock matrix. II. Rigorous bounds on exchange integrals and incremental Fock build. *J. Chem. Phys.* **1997**, *106*, 9708.
- (10) Ochsenfeld, C.; White, C. A.; Head-Gordon, M. Linear and sublinear scaling formation of Hartree–Fock-type exchange matrices. *J. Chem. Phys.* **1998**, *109*, 1663.
- (11) Schwegler, E.; Challacombe, M. Linear scaling computation of the Fock matrix. IV. Multipole accelerated formation of the exchange matrix. *J. Chem. Phys.* **1999**, *111*, 6223.
- (12) Schwegler, E.; Challacombe, M. Linear scaling computation of the Fock matrix. III. Formation of the exchange matrix with permutational symmetry. *Theor. Chem. Acc.* **2000**, *104*, 344–349.
- (13) Pérez-Jorda, J. M.; Yang, W. An algorithm for 3D numerical integration that scales linearly with the size of the molecule. *Chem. Phys. Lett.* **1995**, *241*, 469–476.
- (14) Stratmann, R. E.; Scuseria, G. E.; Frisch, M. J. Achieving linear scaling in exchange-correlation density functional quadratures. *Chem. Phys. Lett.* **1996**, *257*, 213–223.
- (15) Li, X. P.; Nunes, R.; Vanderbilt, D. Density-matrix electronic-structure method with linear system-size scaling. *Phys. Rev. B* **1993**, *47*, 10891.
- (16) Goedecker, S.; Teter, M. Tight-binding electronic-structure calculations and tight-binding molecular dynamics with localized orbitals. *Phys. Rev. B* **1995**, *51*, 9455.
- (17) Liang, W.; Saravanan, C.; Shao, Y.; Baer, R.; Bell, A. T.; Head-Gordon, M. Improved Fermi operator expansion methods for fast electronic structure calculations. *J. Chem. Phys.* **2003**, *119*, 4117.
- (18) Paterson, M. S.; Stockmeyer, L. J. On the Number of Nonscalar Multiplications Necessary to Evaluate Polynomials. *SICOMP* **1973**, *2*, 60.
- (19) McWeeny, R. Some recent advances in density matrix theory. *Rev. Mod. Phys.* **1960**, *32*, 335.
- (20) Shao, Y.; Saravanan, C.; Head-Gordon, M.; White, C. A. Curvy steps for density matrix-based energy minimization: Application to large-scale self-consistent-field calculations. *J. Chem. Phys.* **2003**, *118*, 6144.
- (21) Palser, A.; Manolopoulos, D. Canonical purification of the density matrix in electronic-structure theory. *Phys. Rev. B* **1998**, *58*, 12704–12711.
- (22) Rudberg, E.; Rubensson, E. H. Assessment of density matrix methods for linear scaling electronic structure calculations. *J. Phys.: Condens. Matter* **2011**, *23*, 075502.
- (23) Kim, J.; Jung, Y. On the optimal symmetric purification scheme of the one-particle density matrix. *Chem. Phys. Lett.* **2011**, *511*, 159–160.
- (24) Holas, A. Transforms for idempotency purification of density matrices in linear-scaling electronic-structure calculations. *Chem. Phys. Lett.* **2001**, *340*, 552–558.
- (25) Shao, Y.; Molnar, L. F.; Jung, Y.; Kussmann, J.; Ochsenfeld, C.; Brown, S. T.; Gilbert, A. T. B.; Slipchenko, L. V.; Levchenko, S. V.; O'Neill, D. P.; DiStasio, R. A.; Lochan, R. C.; Wang, T.; Beran, G. J. O.; Besley, N. A.; Herbert, J. M.; Lin, C. Y.; Van Voorhis, T.; Chien, S. H.; Sodt, A.; Steele, R. P.; Rassolov, V. A.; Maslen, P. E.; Korambath, P. P.; Adamson, R. D.; Austin, B.; Baker, J.; Byrd, E. F. C.; Dachsel, H.; Doerksen, R. J.; Dreuw, A.; Dunietz, B. D.; Dutoi, A. D.; Furlani, T. R.; Gwaltney, S. R.; Heyden, A.; Hirata, S.; Hsu, C.; Kedziora, G.; Khalliulin, R. Z.; Klunzinger, P.; Lee, A. M.; Lee, M. S.; Liang, W.; Lotan, I.; Nair, N.; Peters, B.; Proynov, E. I.; Pieniazek, P. A.; Rhee, Y. M.; Ritchie, J.; Rosta, E.; Sherrill, C. D.; Simmonett, A. C.; Subotnik, J. E.; Woodcock, H. L.; Zhang, W.; Bell, A. T.; Chakraborty, A. K.; Chipman, D. M.; Keil, F. J.; Warshel, A.; Hehre, W. J.; Schaefer, H. F.; Kong, J.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M. Advances in methods and algorithms in a modern quantum chemistry program package. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3172–91.
- (26) Rubensson, E. H. Non-monotonic recursive polynomial expansions for linear scaling calculation of the density matrix. *J. Chem. Theory Comput.* **2011**, *7*, 1233–1236.
- (27) Millam, J. M.; Scuseria, G. E. Linear scaling conjugate gradient density matrix search as an alternative to diagonalization for first principles electronic structure calculations. *J. Chem. Phys.* **1997**, *106*, 5569.