Optimized purification for density matrix calculation

Phanish Suryanarayana

College of Engineering, Georgia Institute of Technology, GA 30332, USA

ARTICLE INFO

ABSTRACT

We present an optimized linear-scaling purification method for calculation of the density matrix. Traditional purification polynomials, including those proposed by McWeeny and Holas, are monotonic and have stable fixed points at 0 and 1. We relax these conditions and develop optimized purification polynomials which achieve maximum reduction in the LUMO eigenvalue and maximum increase in the HOMO eigenvalue, while heading towards idempotency. We demonstrate that optimized purification achieves appreciable speedup over traditional purification, which increases with decreasing band-gap. We also show improvement over non-monotonic purification proposed by Rubensson, while having identical performance for polynomials of degree 3.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Electronic structure calculations, including Hartree–Fock (HF) and Density Functional Theory (DFT), typically solve for the orbitals through diagonalization [1–11]. However, this approach results in cubic-scaling with respect to the number of atoms. To overcome this restrictive scaling, there has been great emphasis on the development of linear-scaling methods [12–20]. One of the common approaches to achieve linear-scaling is through the polynomial expansion of the density matrix, with truncation based on its decay [15,19]. A particularly efficient method for this expansion is through recursive purification [19], which is the focus of this letter.

The idempotent density matrix **P** can be expressed as **P** = $\theta(\mu\mathbf{I} - \mathbf{H})$, where θ is the step function, **H** is the Hamiltonian matrix, μ is the Fermi energy and **I** is the identity matrix. In the purification method, a polynomial approximation to the density matrix is developed through the iterative procedure [19]

$$\mathbf{X}_0 = p_0(\mathbf{H}),$$
$$\mathbf{X}_t = p_t(\mathbf{X}_{t-1}) \quad t = 1, 2, \ldots, T. \tag{1}$$

p_0 is the initial linear transformation which maps the eigenvalues of **H** into the interval $[0, 1]$. p_t are purification polynomials which drive the occupied eigenvalues of **H** towards 1 and the unoccupied eigenvalues towards 0, with no change to the eigenvectors. As a consequence, $\mathbf{P} = \lim_{T \to +\infty} \mathbf{X}_T$. In practice, the iteration is stopped once the matrix is sufficiently idempotent.

The efficiency of the purification method is a consequence of its recursive nature which allows for polynomial expansions of extremely high degree in comparison to serial polynomial expansions

[21]. The degree of the polynomial required to accurately represent the density matrix scales as $\mathcal{O}(1/\zeta)$, where ζ is the ratio of the band gap to the spectral width of **H** [15]. By spectral width, we refer to the magnitude of the difference between the largest and smallest eigenvalues. Therefore, in a serial polynomial expansion, the number of matrix–matrix multiplications and consequently the cost typically scale as $\mathcal{O}(1/\zeta)$. However, due to the recursive nature of the purification method, its cost scales as $\mathcal{O}(\log 1/\zeta)$ [22]. This makes the purification method highly efficient for studying systems with small values of ζ, a situation commonly encountered in HF and DFT, where the number of basis functions are orders of magnitude larger than the number of atoms.

The idea of purification was first introduced by McWeeny [23], who used the mapping $p(\mathbf{X}) = 3\mathbf{X}^2 - 2\mathbf{X}^3$ to make a sufficiently idempotent matrix more idempotent. Palser and Manolopoulos [24] proposed the use of matrix purification as a technique to calculate the density matrix from the Hamiltonian matrix without diagonalization. This has motivated the development of numerous purification techniques, which aim to reduce the number of matrix–matrix multiplications required to calculate the density matrix. Holas [25] generalized the McWeeny purification by developing polynomials of higher degree, to enable higher rates of convergence. However, this comes at the price of larger number of matrix–matrix multiplications per purification step. Mazziotti [26,27] developed generalizations of McWeeny's formula within the formalism of the particle-hole duality from the theory of reduced density matrices. He also suggested the use of asymmetric polynomials for purification. Niklasson and coworkers [28,29] introduced a trace-correcting algorithm using asymmetric purification polynomials which does not require *a priori* knowledge of the Fermi energy.

The purification polynomials proposed in the above cited references share some common features. Specifically, they are

E-mail address: phanish.suryanarayana@ce.gatech.edu

monotonic in the interval $[0, 1]$ and have stable fixed points at $0, 1$. We refer to these polynomials as traditional purification polynomials. These conditions have been relaxed recently, with the goal of further increasing the efficiency of the purification method. Kim and Jung [30] proposed the use of generalized nonpurifying functions in the first few iterations of the purification to enhance convergence. Rubensson [31] proposed a scaled purification technique in which he used non-monotonic purification polynomials. He demonstrated significant speedup over traditional purification polynomials, which increases with decreasing band gap.

In this letter, we develop a linear-scaling purification method for calculation of the density matrix. We first define the set of generalized symmetric purification polynomials where the conditions of monotonicity and stable fixed points at $0, 1$ are not enforced. Next, we obtain polynomials from this set which achieve maximum reduction in the LUMO eigenvalue and maximum increase in the HOMO eigenvalue, while heading towards idempotency. Subsequently, we propose a purification method which utilizes these polynomials to calculate the density matrix. Finally, we validate the accuracy and efficacy of the proposed purification method through examples.

2. Theory

Consider a Hamiltonian matrix $\mathbf{H}$ with ordered (increasing) eigenvalues λ_m, $m = 1, 2, \ldots, M$. Since the eigenvectors of $\mathbf{H}$ are unchanged throughout, the purification method can be analyzed through the behavior of the eigenvalues. Let $\lambda_f < \mu < \lambda_{f+1}$, where μ is the Fermi energy, λ_f and λ_{f+1} are the HOMO and LUMO eigenvalues respectively. Further, let $\lambda_M - \mu \geqslant \mu - \lambda_1$, which is typically the case in electronic structure calculations. First, the transformation

$$\widehat{\mathbf{H}} = \frac{\hat{\mu}}{\lambda_M - \mu}(\mu\mathbf{I} - \mathbf{H}) + \hat{\mu}\mathbf{I} \tag{2}$$

is used to map the eigenvalues of $\mathbf{H}$ into the interval $[0, 1]$. We denote the eigenvalues of $\widehat{\mathbf{H}}$ by $\hat{\lambda}_m$, $m = 1, 2, \ldots, M$. Through this transformation, μ is mapped to $\hat{\mu}$, the unoccupied part of the spectrum is mapped to $[0, \hat{\mu})$, and the occupied part of the spectrum is in $(\hat{\mu}, 1]$. Next, the iteration given by Eq. (1) is used to calculate the density matrix. The purification polynomials p (dropping the subscript t) typically have the following properties. First, they have stable fixed points at $0, 1$. Second, they have an unstable fixed point at $\hat{\mu}$. Third, they are monotonic such that for $0 < \hat{\lambda} < \hat{\mu}$, $0 < p(\hat{\lambda}) < \hat{\lambda}$ and for $\hat{\mu} < \hat{\lambda} < 1$, $\hat{\lambda} < p(\hat{\lambda}) < 1$. As a consequence, repeated applications of these polynomials will drive the eigenvalues $\hat{\lambda}_m < \hat{\mu}$ towards 0 and eigenvalues $\hat{\lambda}_m > \hat{\mu}$ towards 1. In this letter, we focus on the so called symmetric purification polynomials which satisfy $\hat{\mu} = 0.5$ and $p(\hat{\lambda}) + p(1 - \hat{\lambda}) = 1$ for $0 \leqslant \hat{\lambda} \leqslant 1$.

Consider a matrix having unoccupied eigenvalues in the interval $[0, \hat{\lambda}_c]$ and occupied eigenvalues in $[1 - \hat{\lambda}_c, 1]$. For this matrix, we define the set of symmetric purification polynomials of degree $2K + 1$ as

$$\mathcal{C}_{2K+1} = \{p(\hat{\lambda}) \in \mathcal{P}_{2K+1} : \quad 0 \leqslant p(\hat{\lambda}) \leqslant p(\hat{\lambda}_c) < \hat{\lambda}_c \text{ for } 0 \leqslant \hat{\lambda} \leqslant \hat{\lambda}_c < \hat{\mu};$$
$$1 - \hat{\lambda}_c < p(1 - \hat{\lambda}_c) \leqslant p(\hat{\lambda}) \leqslant 1 \text{ for } \hat{\mu} < 1 - \hat{\lambda}_c \leqslant \hat{\lambda} \leqslant 1;$$
$$p(\hat{\mu}) = \hat{\mu}; \quad \min_{[0,1]} p(\hat{\lambda}) = 0; \quad \max_{[0,1]} p(\hat{\lambda}) = 1\}, \tag{3}$$

where $\mathcal{P}_{2K+1}$ is the space of polynomials of degree $2K + 1$. In the special case when $\hat{\lambda}_c = 0$, we have $p(0) = 0$ and $p(1) = 1$. The first constraint in $\mathcal{C}_{2K+1}$ is to ensure that the unoccupied eigenvalues remain unoccupied and the second constraint ensures that the occupied eigenvalues remain occupied. The last three constraints ensure that $\hat{\mu} = 0.5$ is a fixed point, the minimum value attained is 0 and the maximum value attained is 1. Overall, the constraints guarantee that the purified eigenvalues occupy the intervals $[0, p(\hat{\lambda}_c)]$ and

$[p(1 - \hat{\lambda}_c), 1]$, where $p(\hat{\lambda}_c) < \hat{\lambda}_c$ and $p(1 - \hat{\lambda}_c) > 1 - \hat{\lambda}_c$. As a consequence, repeated applications of these polynomials results in idempotency.

It is clear that traditional purification polynomials, including those proposed by McWeeny and Holas, belong to the set $\mathcal{C}_{2K+1}$. However, the polynomials belonging to $\mathcal{C}_{2K+1}$ are not required to be monotonic and have stable fixed points at 0 and 1. Therefore, $\mathcal{C}_{2K+1}$ represents a generalization of traditional purification polynomials. With this notion of purification, we define the optimized purification polynomial of degree $2K + 1$ to be the solution of the optimization problem

$$p^*(\hat{\lambda}) = \arg \max_{p(\hat{\lambda}) \in \mathcal{C}_{2K+1}} \left(\hat{\lambda}_c - p(\hat{\lambda}_c) \right). \tag{4}$$

Given that the purified eigenvalues occupy $[0, p(\hat{\lambda}_c)]$ and $[p(1 - \hat{\lambda}_c), 1]$, the optimized purification polynomial achieves maximum reduction in the intervals occupied by the eigenvalues, while simultaneously heading towards idempotency.

Since solving Eq. (4) with high accuracy is challenging, we propose that the optimized purification polynomials have derivatives of the form

$$\frac{dp^*(\hat{\lambda})}{d\hat{\lambda}} = c\prod_{k=1}^{K}(\hat{\lambda} - r_k)(\hat{\lambda} + r_k - 1), \tag{5}$$

where $r_K \leqslant r_{K-1} \leqslant \cdots \leqslant r_1 < 0.5 < (1 - r_1) \leqslant \cdots \leqslant (1 - r_{K-1}) \leqslant (1 - r_K)$ are the stationary points of $p^*(\hat{\lambda})$. The constant c is to ensure that the polynomial is bounded by 0 and 1 in the interval $[0, 1]$. The stationary points $\{r_1, r_2, \ldots r_K\}$, scaling factor c and the constant of integration are determined through the constraints

$$p^*(0) = p^*(\hat{\lambda}_c)|0,$$
$$p^*(r_k) = 0, \quad k = 1, 3, \ldots, K|K - 1,$$
$$p^*(r_k) = p^*(\hat{\lambda}_c), \quad k = 2, 4, \ldots K - 1|K, \tag{6}$$
$$p^*(\hat{\mu}) = \hat{\mu}.$$

Above, the values to the left and right of the vertical line ('|') represent the values of the corresponding quantity for odd and even K respectively. We have verified numerically that the optimization problem given by Eq. (4) is indeed equivalent to the problem given by Eqs. (5) and (6).

For the special case $K = 1$ and $\hat{\lambda}_c = 0$, we recover McWeeny's purification polynomial [23]. For $K > 1$ and $\hat{\lambda}_c = 0$, we recover the purification polynomials of higher degree proposed by Holas [25]. In Figure 1, we plot the optimized purification polynomials of degree 3, 5 and 7 obtained for $\hat{\lambda}_c = 0.49$. We note that the optimized purification polynomials as defined by Eq. (4) are in general non-monotonic and do not have stable fixed points at 0 and 1.

Next, we study the efficacy of the optimized purification polynomials. We use the parameter $\eta = (\hat{\lambda}_c - p(\hat{\lambda}_c))/(\hat{\lambda}_c - \overline{p}(\hat{\lambda}_c))$ to judge the efficacy of the purification polynomial $p(\hat{\lambda})$ relative to the traditional purification polynomial $\overline{p}(\hat{\lambda}_c)$ of the same degree. In Figure 2, we plot η versus $\hat{\lambda}_c$ for optimized purification and compare it with the non-monotonic purification proposed by Rubensson [31]. For $K = 1$, we have chosen $\overline{p}(\hat{\lambda}_c)$ to be McWeeny's polynomial and for $K > 1$, we have chosen $\overline{p}(\hat{\lambda}_c)$ to be Holas's polynomials. We observe that the optimized purification is significantly more efficient compared to traditional purification, particularly when $\hat{\lambda}_c$ nears 0.5. We also note that the non-monotonic purification of Rubensson has identical performance to the optimized purification for $K = 1$. However, for larger K, optimized purification achieves larger values of η, highlighting its efficacy.

In Algorithm 1, we outline the calculation of the density matrix using the optimized purification polynomials. Given the Hamiltonian matrix $\mathbf{H}$ and Fermi energy μ, we proceed to calculate the