Chin. Ann. Math.
30B(6), 2009, 729–742
DOI: 10.1007/s11401-009-0201-7

Chinese Annals of Mathematics, Series B
© The Editorial Office of CAM and Springer-Verlag Berlin Heidelberg 2009

Pole-Based Approximation of the Fermi-Dirac Function****

Lin LIN* Jianfeng LU* Lexing YING** Weinan E***

(*Dedicated to Professor Andrew Majda on the Occasion of his* 60*th Birthday*)

Abstract Two approaches for the efficient rational approximation of the Fermi-Dirac function are discussed: one uses the contour integral representation and conformal mapping, and the other is based on a version of the multipole representation of the Fermi-Dirac function that uses only simple poles. Both representations have logarithmic computational complexity. They are of great interest for electronic structure calculations.

Keywords Contour integral, Fermi-Dirac function, Rational approximation
2000 MR Subject Classification 41A20, 65F30

1 Introduction

Given an effective one-particle Hamiltonian $\boldsymbol{H}$, the inverse temperature $\beta = \frac{1}{k_{\mathrm{B}}T}$ and the chemical potential μ, the finite temperature single-particle density matrix of the system is given by the Fermi operator

$$\boldsymbol{\rho} = 2(1 + \exp(\beta(\boldsymbol{H} - \mu)))^{-1} = 1 - \tanh\Big(\frac{\beta}{2}(\boldsymbol{H} - \mu)\Big), \tag{1.1}$$

where tanh is the hyperbolic tangent function.

In the last decade or so, the development of accurate and numerically efficient representations of the Fermi operator has attracted a great deal of attention in the quest for linear scaling electronic structure methods based on effective one-electron Hamiltonians. These approaches have a numerical cost that scales linearly with N, the number of electrons, and thus hold the promise of making electronic structure analysis of large systems feasible. Achieving linear scaling for realistic systems is very challenging. Formulations based on the Fermi operator are appealing since this operator gives directly the single particle density matrix without the need for diagonalizing the Hamiltonian.

Manuscript received June 6, 2009. Published online August 18, 2009.
*Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ 08544, USA. E-mail: linlin@math.princeton.edu jianfeng@math.princeton.edu
**Department of Mathematics and ICES, University of Texas at Austin, 1 University Station/C1200, Austin, TX 78712, USA. E-mail: lexing@math.utexas.edu
***Department of Mathematics and PACM, Princeton University, Princeton, NJ 08544, USA. E-mail: weinan@math.princeton.edu
****Project supported by the Department of Energy (No. DE-FG02-03ER25587), the Office of Naval Research (No. N00014-01-1-0674), an Alfred P. Sloan Research Fellowship and a startup grant from University of Texas at Austin.

From a computational viewpoint, one main issue is that the right hand side of (1.1) is an operator-valued function. To evaluate this function, we have to replace or approximate it by something which can be computed directly without diagonalization. Obvious candidates are polynomial or rational approximations. Such an approach was first introduced by Baroni and Giannozzi [1] and Goedecker and co-workers [4, 5] (see also the review article [6]). Several improvements have been made since then, for example, in [2, 3, 9–11, 14]. These are put broadly under the umbrella of "Fermi operator expansion" (abbreviated as FOE).

From the viewpoint of efficiency, a major concern is the cost for representing the Fermi operator as a function of $\beta\Delta E$ (for finite temperature) or $\frac{\Delta E}{E_g}$ (for gapped systems) where β is the inverse temperature, ΔE is the spectral width of the discretized Hamiltonian matrix and E_g is the spectrum gap of the Hamiltonian around the chemical potential. Consider a finite temperature gapless system for example, the cost of the FOE proposed by Goedecker et al scales as $\beta\Delta E$. The fast polynomial summation technique introduced by Head-Gordon et al [10, 11] reduces the cost to $(\beta\Delta E)^{\frac{1}{2}}$. The cost of the hybrid algorithm proposed by Parrinello et al in a recent preprint [2] scales as $(\beta\Delta E)^{\frac{1}{3}}$. The cost was brought down to logarithmic scaling $\ln(\beta\Delta E)$ in [12] by using a multipole representation of the Matsubara expansion of the Fermi-Dirac function.

The purpose of this article is to introduce two alternative rational expansions of the Fermi-Dirac function that use only simple poles and have computational cost that scales logarithmically. The first strategy is to use the contour integral and conformal mapping idea proposed recently in [8]. This will be presented in the next section. The other strategy is to borrow ideas from [15] and use a version of multipole expansion [12] that only involves simple poles. This will be discussed in Section 3. Numerical examples illustrating the efficiency and accuracy of the representations are discussed in Section 4.

2 Rational Expansions Based on Contour Integral

Our first approach is an adaptation of the ideas proposed recently in [8] based on contour integral representation and conformal mapping. Let us first briefly recall the main idea of [8]. Consider a function f that is analytic in $\mathbb{C} \setminus (-\infty, 0]$ and an operator $\boldsymbol{A}$ with spectrum in $[m, M] \subset \mathbb{R}^+$. One wants to evaluate $f(\boldsymbol{A})$ using a rational expansion of f by discretizing the contour integral

$$f(\boldsymbol{A}) = \frac{1}{2\pi \mathrm{i}} \int_\Gamma f(z)(z - \boldsymbol{A})^{-1} \mathrm{d}z. \tag{2.1}$$

The innovative technique in [8] was to construct a conformal map that maps the stripe $S = [-K, K] \times [0, K']$ to the upper half (denoted by Ω^+) of the domain $\Omega = \mathbb{C} \setminus ((-\infty, 0] \cup [m, M])$. This special map from $t \in S$ to $z \in \Omega^+$ is given by

$$z = \sqrt{mM}\Big(\frac{k^{-1} + u}{k^{-1} - u}\Big), \quad u = \mathrm{sn}(t) = \mathrm{sn}(t|k), \quad k = \frac{\sqrt{\frac{M}{m}} - 1}{\sqrt{\frac{M}{m}} + 1}. \tag{2.2}$$

Here $\mathrm{sn}(t)$ is one of the Jacobi elliptic functions and the numbers K and K' are complete elliptic integrals whose values are given by the condition that the map is from S to Ω^+.

Applying the trapezoidal rule with Q equally spaced points in $(-K + \frac{\mathrm{i}K'}{2}, K + \frac{\mathrm{i}K'}{2})$,

$$t_j = -K + \frac{\mathrm{i}K'}{2} + 2\frac{(j - \frac{1}{2})K}{Q}, \quad 1 \le j \le Q, \tag{2.3}$$

we get the quadrature rule (denote $z_j = z(t_j)$)

$$f_Q(\boldsymbol{A}) = \frac{-4K\sqrt{mM}}{\pi Q k} \operatorname{Im} \sum_{j=1}^{Q} \frac{f(z_j)(z_j - \boldsymbol{A})^{-1} \operatorname{cn}(t_j)\operatorname{dn}(t_j)}{(k^{-1} - \operatorname{sn}(t_j))^2}. \tag{2.4}$$

Here cn and dn are the other two Jacobi elliptic functions in standard notation and the factor $\operatorname{cn}(t_j)\operatorname{dn}(t_j)(k^{-1} - \operatorname{sn}(t_j))^{-2}\frac{\sqrt{mM}}{k}$ comes from the Jacobian of the function $z(t)$.

It is proved in [8] that the convergence is exponential in the number of quadrature points Q and the exponent deteriorates only logarithmically as $\frac{M}{m} \to \infty$:

$$\|f(\boldsymbol{A}) - f_Q(\boldsymbol{A})\| = \mathcal{O}(\mathrm{e}^{-\pi^2 Q(\log(\frac{M}{m})+3)^{-1}}). \tag{2.5}$$

To adapt the idea to our setting with the Fermi-Dirac function or the hyperbolic tangent function, we face with two differences: First, the tanh function has singularities on the imaginary axis. Second, the operator we are considering, $\beta(\boldsymbol{H} - \mu)$, has spectrum on both the negative and positive axis.

2.1 Gapped case

We first consider the case when the Hamiltonian $\boldsymbol{H}$ has a gap in its spectrum around the chemical potential μ, such that $\operatorname{dist}(\mu, \sigma(\boldsymbol{H})) = E_g > 0$. Physically, this will be the case when the system is an insulator.

Let us consider $f(z) = \tanh(\frac{\beta}{2} z^{\frac{1}{2}})$ acting on the operator $\boldsymbol{A} = (\boldsymbol{H} - \mu)^2$. Now, $f(z)$ has singularities only on $(-\infty, 0]$ and the spectrum of $\boldsymbol{A}$ is contained in $[E_g^2, E_M^2]$, where

$$E_M = \max_{E \in \sigma(\boldsymbol{H})} |E - \mu|.$$

We note that obviously $E_M \leq \Delta E$. Hence we are back in the same scenario as considered in [8] except that we need to take care of different branches of the square root function when we apply the quadrature rule.

More specifically, we construct the contour and quadrature points z_j in the z-plane by using parameters $m = E_g^2$ and $M = E_M^2$. Denote $g(\xi) = \tanh(\frac{\beta\xi}{2})$, $\xi_j^{\pm} = \pm z_j^{\frac{1}{2}}$ and $\boldsymbol{B} = \boldsymbol{H} - \mu$. The quadrature rule is then given by

$$\begin{aligned} g_Q(\boldsymbol{B}) = \frac{-2K\sqrt{mM}}{\pi Q k} \operatorname{Im}\Big(& \sum_{j=1}^{Q} \frac{g(\xi_j^+)(\xi_j^+ - \boldsymbol{B})^{-1} \operatorname{cn}(t_j)\operatorname{dn}(t_j)}{\xi_j^+ (k^{-1} - \operatorname{sn}(t_j))^2} \\ & + \sum_{j=1}^{Q} \frac{g(\xi_j^-)(\xi_j^- - \boldsymbol{B})^{-1} \operatorname{cn}(t_j)\operatorname{dn}(t_j)}{\xi_j^- (k^{-1} - \operatorname{sn}(t_j))^2} \Big), \end{aligned} \tag{2.6}$$

where the factors $\xi_j^{\pm}$ in the denominator come from the Jacobian of the map from z to ξ. The number of poles to be inverted is $N_{\text{pole}} = 2Q$. After applying (2.5), we have a similar error estimate for $g(\boldsymbol{B})$

$$\|g(\boldsymbol{B}) - g_Q(\boldsymbol{B})\| = \mathcal{O}(\mathrm{e}^{-\pi^2 Q(2\log(\frac{E_M}{E_g})+3)^{-1}}). \tag{2.7}$$

In Figure 1, a typical configuration of the quadrature points is shown. The x-axis is taken to be $E - \mu$. We see that in this case the contour consists of two loops, one around the spectrum below the chemical potential and the other around the spectrum above the chemical potential.

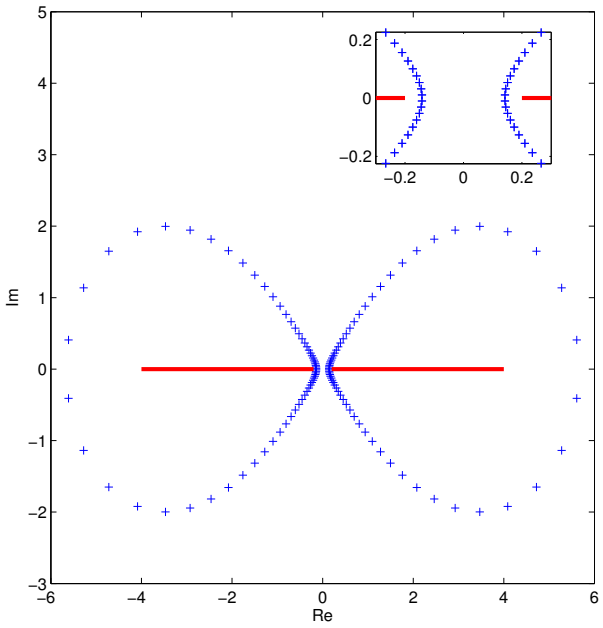

Figure 1 A typical configuration of the poles on a two-loop contour. $Q = 30$, $E_g = 0.2$, $E_M = 4$ and $\beta = 1000$. The red line indicates the spectrum. The inset shows the poles close to the origin. The x-axis is $E - \mu$ with E the eigenvalue of $\boldsymbol{H}$. The poles with negative imaginary parts are not explicitly calculated.

Note further that as the temperature goes to zero, the Fermi-Dirac function converges to the step function

$$\eta(\xi) = \begin{cases} 2, & \xi \le 0, \\ 0, & \xi > 0. \end{cases} \tag{2.8}$$

Therefore, the contribution of the quadrature points ξ_j^+ on the right half plane ($\mathrm{Re}\,\xi_j^+ > 0$) is negligible when β is large. In particular, for the case of zero temperature, one may choose only the quadrature points on the left half plane. The quadrature formula we obtain then becomes

$$\eta_Q(\boldsymbol{B}) = \frac{-4K\sqrt{mM}}{\pi Q k} \mathrm{Im}\Big(\sum_{j=1}^{Q} \frac{(\xi_j^- - \boldsymbol{B})^{-1} \mathrm{cn}(t_j)\mathrm{dn}(t_j)}{\xi_j^- (k^{-1} - \mathrm{sn}(t_j))^2} \Big). \tag{2.9}$$

The number of poles to be inverted is then $N_{\mathrm{pole}} = Q$.

We show in Figure 2 a typical configuration of the set of quadrature points. Only one loop is required compared with Figure 1.

2.2 Gapless case

The more challenging case is when the spectrum of $\boldsymbol{H}$ does not have a gap, i.e., $E_g = 0$. Physically, this corresponds to the case of metallic systems. In this case, the construction discussed in the last subsection does not work.

To overcome this problem, we note that the hyperbolic tangent function $\tanh(\frac{\beta}{2}z)$ is analytic except at poles $(2l-1)\frac{\pi}{\beta \mathrm{i}}$, $l \in \mathbb{Z}$ on the imaginary axis. Therefore, we could construct a contour around the whole spectrum of $\boldsymbol{H}$ which passes through the imaginary axis on the upper half

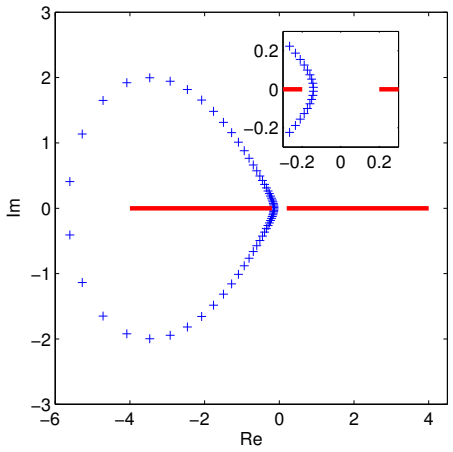

Figure 2 A typical configuration of the poles for zero temperature ($\beta = \infty$). $Q = 30$, $E_g = 0.2$ and $E_M = 4$. The red line indicates the spectrum. The inset zooms into the poles that is close to the origin. The x-axis is $E - \mu$ with E the eigenvalue of $\boldsymbol{H}$. The poles with negative imaginary parts are not explicitly calculated.

plane between the origin and $\frac{\pi}{\beta \mathrm{i}}$ and also on the lower half plane between the origin and $-\frac{\pi}{\beta \mathrm{i}}$. Thus, we will have a dumbbell-shaped contour as shown in Figure 3.

To be more specific, let us first construct the contour and quadrature points z_j in the z-plane as in the last subsection by using parameters $m = \frac{\pi^2}{\beta^2}$ and $M = E_M^2 + \frac{\pi^2}{\beta^2}$. Denote $\xi_j^{\pm} = \pm(z_j - \frac{\pi^2}{\beta^2})^{\frac{1}{2}}$, $g = \tanh(\frac{\beta\xi}{2})$ and $\boldsymbol{B} = \boldsymbol{H} - \mu$. The quadrature rule takes the following form

$$g_Q(\boldsymbol{B}) = \frac{-2K\sqrt{mM}}{\pi Q k} \operatorname{Im}\Big(\sum_{j=1}^{Q} \frac{g(\xi_j^+)(\xi_j^+ - \boldsymbol{B})^{-1}\mathrm{cn}(t_j)\mathrm{dn}(t_j)}{\xi_j^+(k^{-1} - \mathrm{sn}(t_j))^2} + \sum_{j=1}^{Q} \frac{g(\xi_j^-)(\xi_j^- - \boldsymbol{B})^{-1}\mathrm{cn}(t_j)\mathrm{dn}(t_j)}{\xi_j^-(k^{-1} - \mathrm{sn}(t_j))^2} \Big). \tag{2.10}$$

When applying the quadrature formula, the number of poles to be inverted is $N_{\mathrm{pole}} = 2Q$. Figure 3 shows a typical configuration of quadrature points for $Q = 30$. The map $\xi(z) = (z - \frac{\pi^2}{\beta^2})^{\frac{1}{2}}$ maps the circle in the z-plane to a dumbbell-shaped contour (put two branches together).

Actually, what is done could be understood as follows. Similarly to [8], we have constructed a map from the rectangular domain $[-3K, K] \times [0, K']$ to the upper half of the domain

$$U = \{z \mid \operatorname{Im} z \geq 0\} \setminus \Big([-E_M, E_M] \cup \mathrm{i}\Big[\frac{\pi}{\beta}, \infty\Big)\Big).$$

The map is carried out in three steps, shown in Figure 4. The first two steps use the original map constructed in [8], however with extended domain of definition. First, the Jacobi elliptic function

$$u = \mathrm{sn}(t) = \mathrm{sn}(t|k), \quad k = \frac{\sqrt{\frac{M}{m}} - 1}{\sqrt{\frac{M}{m}} + 1} \tag{2.11}$$

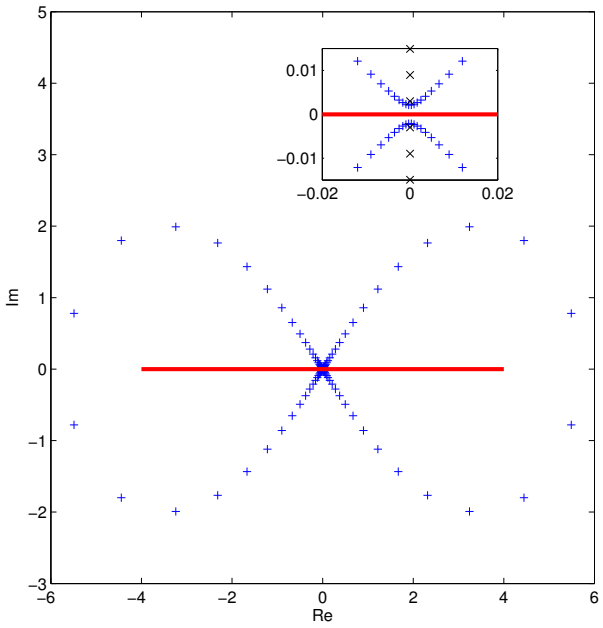

Figure 3 A typical configuration of the poles on a dumbbell-shaped contour. $Q = 30$, $E_g = 0$, $E_M = 4$ and $\beta = 1000$. The inset zooms into the part close to the origin. The red line indicates the spectrum. The black crosses indicate the positions of the poles of tanh function on the imaginary axis. The poles with negative imaginary parts are not explicitly calculated.

maps the rectangular domain to the complex plane, with the ends mapping to $[1, k^{-1}]$ and the middle vertical line $-K + \mathrm{i}[0, K']$ to $[-k^{-1}, -1]$. Then, the Möbius transformation

$$z = \sqrt{mM}\Big(\frac{k^{-1} + u}{k^{-1} - u}\Big) \tag{2.12}$$

maps the complex plane to itself in such a way that $[-k^{-1}, -1]$ and $[1, k^{-1}]$ are mapped to $[0, m]$ and $[M, \infty]$, respectively. Finally, the shifted square root function

$$\xi = (z - m)^{\frac{1}{2}} \tag{2.13}$$

maps the complex plane to the upper-half plane (we choose the branch of the square root such that the lower-half plane is mapped to the second quadrant and the upper-half plane is mapped to the first quadrant), in such a way that $[0, m]$ is sent to $\mathrm{i}[0, \sqrt{m}\,]$ and $[M, \infty)$ is sent to $(-\infty, -\sqrt{M - m}\,] \cup [\sqrt{M - m}\,, \infty)$. The map can be extended to a map from $[-7K, K] \times [0, K']$ to the whole U. In this case, the z-plane becomes a double-covered Riemann surface with branch point at m.

Since the function g is analytic in the domain U, the composite function $g(t) = g(\xi(z(u(t))))$ is analytic in the stripe in the t-plane, and therefore, the trapezoidal rule converges exponentially fast. Using a similar analysis that leads to (2.5), it can be shown that

$$\|g(\boldsymbol{B}) - g_Q(\boldsymbol{B})\| = \mathcal{O}(\mathrm{e}^{-CQ\log^{-1}(\beta E_M)}), \tag{2.14}$$

where C is a constant.

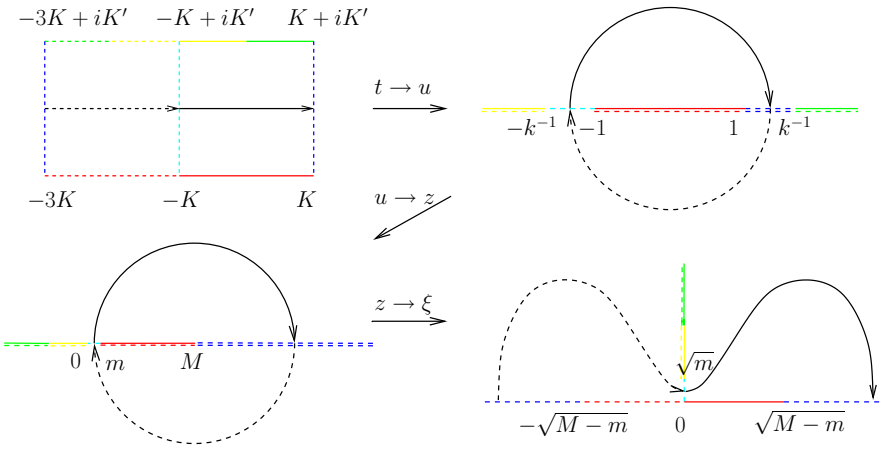

Figure 4 The map from the rectangular domain $[-3K, K] \times [0, K']$ to the upper-half of the domain U. The map is constructed in three steps: $t \to u \to z \to \xi$. The boundaries are shown in various colors and line styles.

We remark that the construction proposed in this subsection also applies to the gapped case. In practice, if the temperature is high (so that β is small) or the gap around the chemical potential is small (in particular, for gapless system), the contour passing through the imaginary axis will be favorable; otherwise, the construction in the last subsection will be more efficient.

3 Rational Approximations Based on the Multipole Expansion

Another strategy for obtaining an efficient rational approximation for the Fermi-Dirac function for finite temperature is based on the multipole expansion, proposed recently in [12]. Let us first recall the construction of the multipole representation.

Using the Matsubara representation (pole expansion) of the Fermi-Dirac function, the density matrix is given by

$$\boldsymbol{\rho} = 1 - 4\,\mathrm{Re}\sum_{l=1}^{\infty} \frac{1}{\beta(\boldsymbol{H}-\mu) - (2l-1)\pi\mathrm{i}}. \tag{3.1}$$

The summation in (3.1) can be seen as a summation of residues contributed from the poles $\{(2l-1)\pi\mathrm{i}\}$, with l a positive integer, on the imaginary axis. This suggests looking for a multipole expansion of the contributions from the poles, as was done in the fast multipole method (FMM, see [7]). To do so, we use a dyadic grouping of the poles, in which the n-th group contains terms from $l = 2^{n-1}$ to $l = 2^n - 1$, for a total of 2^{n-1} terms. We decompose the summation in (3.1) accordingly. Let $x = \beta(\boldsymbol{H} - \mu)$. Then

$$\sum_{l=1}^{\infty} \frac{1}{x-(2l-1)\pi\mathrm{i}} = \sum_{n=1}^{\infty} \sum_{l=2^{n-1}}^{2^n-1} \frac{1}{x-(2l-1)\pi\mathrm{i}} = \sum_{n=1}^{\infty} S_n. \tag{3.2}$$

The basic idea is to combine the simple poles into a set of multipoles at $l = l_n$, where l_n is taken as the midpoint of the interval $[2^{n-1}, 2^n - 1]$:

$$l_n = \frac{3\cdot 2^{n-1} - 1}{2}. \tag{3.3}$$

Then the S_n term in the above equation can be written as

$$\begin{aligned} S_n &= \sum_{l=2^{n-1}}^{2^n-1} \frac{1}{x-(2l_n-1)\pi\mathrm{i}-2(l-l_n)\pi\mathrm{i}} \\ &= \sum_{l=2^{n-1}}^{2^n-1} \frac{1}{x-(2l_n-1)\pi\mathrm{i}} \sum_{\nu=0}^{\infty} \Big(\frac{2(l-l_n)\pi\mathrm{i}}{x-(2l_n-1)\pi\mathrm{i}}\Big)^{\nu} \\ &= \sum_{l=2^{n-1}}^{2^n-1} \frac{1}{x-(2l_n-1)\pi\mathrm{i}} \sum_{\nu=0}^{P-1} \Big(\frac{2(l-l_n)\pi\mathrm{i}}{x-(2l_n-1)\pi\mathrm{i}}\Big)^{\nu} \\ &\quad + \sum_{l=2^{n-1}}^{2^n-1} \frac{1}{x-(2l-1)\pi\mathrm{i}} \Big(\frac{2(l-l_n)\pi\mathrm{i}}{x-(2l_n-1)\pi\mathrm{i}}\Big)^{P}. \end{aligned} \tag{3.4}$$

Using the fact that x is real, the second term in (3.4) can be bounded by

$$\sum_{l=2^{n-1}}^{2^n-1} \Big|\frac{1}{x-(2l-1)\pi\mathrm{i}}\Big|\Big|\frac{2(l-l_n)\pi\mathrm{i}}{x-(2l_n-1)\pi\mathrm{i}}\Big|^{P} \le \sum_{l=2^{n-1}}^{2^n-1} \frac{1}{|(2l-1)\pi|}\Big|\frac{2(l-l_n)}{2l_n-1}\Big|^{P} \le \frac{1}{2\pi}\frac{1}{3^P}.$$

Therefore, if we approximate the sum S_n by the first P terms, the error decays exponentially fast with P:

$$\Big|S_n(x) - \sum_{l=2^{n-1}}^{2^n-1} \frac{1}{x-(2l_n-1)\pi\mathrm{i}} \sum_{\nu=0}^{P-1} \Big(\frac{2(l-l_n)\pi\mathrm{i}}{x-(2l_n-1)\pi\mathrm{i}}\Big)^{\nu}\Big| \le \frac{1}{2\pi}\frac{1}{3^P}. \tag{3.5}$$

The above analysis is of course standard from the view point of the fast multipole method (see [7]). The overall philosophy is also similar: given a preset error tolerance, one selects the value of P, the number of terms to retain in S_n, according to (3.5).

Moreover, the remainder of the sum in (3.1) from $l = M_{\mathrm{pole}} + 1$ to ∞ has an explicit expression

$$\mathrm{Re} \sum_{l=M_{\mathrm{pole}}+1}^{\infty} \frac{1}{2x-(2l-1)\mathrm{i}\pi} = \frac{1}{2\pi}\,\mathrm{Im}\,\psi\Big(M_{\mathrm{pole}} + \frac{1}{2} + \frac{\mathrm{i}}{\pi}x\Big), \tag{3.6}$$

where ψ is the digamma function $\psi(z) = \frac{\Gamma'(z)}{\Gamma(z)}$.

In summary, we arrive at the following multipole representation for the Fermi operator (see [12]):

$$\begin{aligned} \boldsymbol{\rho} = 1 - 4\,\mathrm{Re} &\sum_{n=1}^{N_G} \sum_{l=2^{n-1}}^{2^n-1} \frac{1}{\beta(\boldsymbol{H}-\mu)-(2l_n-1)\pi\mathrm{i}} \sum_{\nu=0}^{P-1} \Big(\frac{2(l-l_n)\pi\mathrm{i}}{\beta(\boldsymbol{H}-\mu)-(2l_n-1)\pi\mathrm{i}}\Big)^{\nu} \\ &- \frac{2}{\pi}\,\mathrm{Im}\,\psi\Big(M_{\mathrm{pole}} + \frac{1}{2} + \frac{\mathrm{i}}{2\pi}\beta(\boldsymbol{H}-\mu)\Big) + \mathcal{O}\Big(\frac{N_G}{3^P}\Big), \end{aligned} \tag{3.7}$$

where N_G is the number of groups in the multipole representation. $M_{\mathrm{pole}} = 2^{N_G} - 1$ is the number of poles that are effectively represented in the original Matsubara representation. In practice, N_G simple poles are first calculated, and then the $N_G(P-1)$ multipoles can be constructed through matrix-matrix multiplication.

A disadvantage of (3.7) is that one needs to multiply simple poles together to get the multipoles before extracting the diagonal of Fermi operator. This prevents us from being able

to directly apply the fast algorithms for extracting the diagonal of an inverse matrix, such as the one proposed in [13]. Therefore, it is useful to find an expansion similar to (3.7) that uses only simple poles. As we mentioned earlier, the key idea in deriving (3.7) is to combine the poles in each group together to form multipoles as the distance between them and the real axis is large. However, if instead we want an expansion that involves only simple poles, it is natural to revisit the variants of FMM that only use simple poles, for example, the version introduced in [15]. The basic idea there is to use a set of equivalent charges on a circle surrounding the poles in each group to reproduce the effective potential away from these poles.

Specifically, take the group of poles from $l = 2^{n-1}$ to $l = 2^n - 1$ for example. Consider a circle B_n with center $c_n = (3 \cdot 2^{n-1} - 2)\pi\mathrm{i}$ and radius $r_n = 2^{n-1}\pi$. It is clear that the circle B_n encloses the poles considered. Take P equally spaced points $\{x_{n,k}\}_{k=1}^P$ on the circle B_n. Next, one needs to place equivalent charges $\{\rho_{n,k}\}_{k=1}^P$ at these points such that the potential produced by these equivalent charges match with the potential produced by the poles inside B_n away from the circle. This can be done in several ways, for example, by matching the multipole expansion, by discretizing the potential on B_n generated by the poles, and so on. Here we follow the approach used in [15].

We simply take a bigger concentric circle $\mathcal{B}_n$ outside B_n with radius $R_n = 2^n\pi$ and match the potential generated on $\mathcal{B}_n$ by the poles and by the equivalent charges on B_n. For this purpose, we solve for $\rho_{n,k}$ the equation

$$\sum_{k=1}^{P} \frac{\rho_{n,k}}{y - x_{n,k}} = \sum_{l=2^{n-1}}^{2^n-1} \frac{1}{y - (2l-1)\pi\mathrm{i}}, \quad y \in \mathcal{B}_n. \tag{3.8}$$

Regularization techniques such as Tikhonov regularization are required here since this is a first-kind Fredholm equation.

One can also prove that similar to the original version of the multipole representation, the error in the potential produced by the equivalent charges decays exponentially in P. The details can be found in [15]. Putting these all together, we can write down the following expansion of the Fermi-Dirac function

$$\begin{aligned}\boldsymbol{\rho} = 1 &- 4\,\mathrm{Re}\sum_{n=1}^{N_G}\sum_{k=1}^{P} \frac{\rho_{n,k}}{\beta(\boldsymbol{H}-\mu) - x_{n,k}} \\ &- \frac{2}{\pi}\,\mathrm{Im}\,\psi\Big(M_{\mathrm{pole}} + \frac{1}{2} + \frac{\mathrm{i}}{2\pi}\beta(\boldsymbol{H}-\mu)\Big) + \mathcal{O}\Big(\frac{N_G}{3^P}\Big).\end{aligned} \tag{3.9}$$

The number of poles that are effectively represented in the original Matsubara representation is still $M_{\mathrm{pole}} = 2^{N_G} - 1$. $N_{\mathrm{pole}} = N_G P$ simple poles are now to be calculated in practice.

The tail part can be approximated using a Chebyshev polynomial expansion. Similar to the analysis in [12], it can be shown that the complexity of the expansion is $\mathcal{O}(\log \beta\Delta E)$. As we pointed out earlier, the advantage of (3.9) over (3.7) is that only simple poles are involved in the formula. This is useful when combined with fast algorithms for extracting the diagonal of an inverse matrix (see [13]).

Note that in (3.7) and (3.9), for $2^{n-1} < P$ there would be no savings if we use P terms in the expansion. They are written in this form just for simplicity. In practice the first P simple poles will be calculated separately and the multipole expansion will be used starting from the $(P+1)$-th term and the starting level is $n = \log_2 P + 1$. We show in Figure 5 a typical configuration of the set of poles in the multipole representation type algorithm.

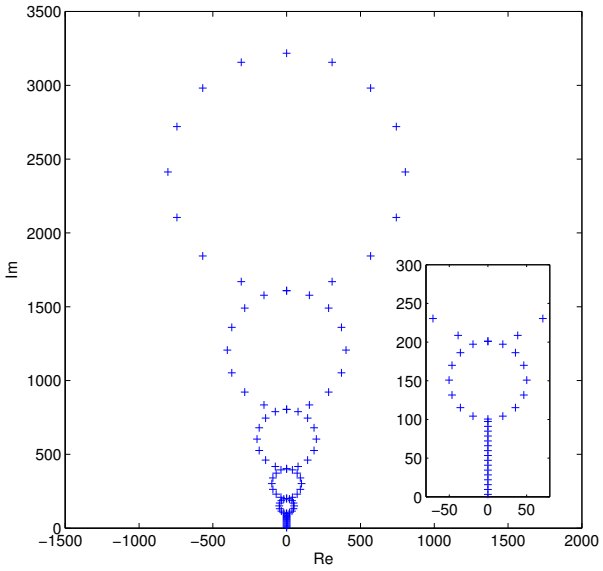

Figure 5 A typical configuration of the poles in the multipole representation type algorithm. $M_{\text{pole}} = 512$ and $P = 16$ is used in this figure. The poles with negative imaginary parts are not explicitly shown. The inset shows the first few poles. The first 16 poles are calculated separately and the starting level is $n = 5$.

4 Numerical Results

We test the algorithms described above by using a two dimensional nearest neighbor tight binding model for the Hamiltonian. The matrix components of the Hamiltonian can be written as (in atomic units),

$$H_{i'j';ij} = \begin{cases} 2 + V_{ij}, & i' = i,\ j' = j, \\ -\dfrac{1}{2} + V_{ij}, & i' = i \pm 1,\ j' = j \text{ or } i' = i,\ j' = j \pm 1. \end{cases} \tag{4.1}$$

The on-site potential energy V_{ij} is chosen to be a uniform random number between 0 and 10^{-3}. The domain size is 32×32 with periodic boundary condition. The chemical potential will be specified later. The accuracy is measured by the L^1 error of the electronic density profile per electron

$$\Delta\rho_{\text{rel}} = \frac{\operatorname{Tr}|\widehat{P} - P|}{N_{\text{Electron}}}. \tag{4.2}$$

4.1 Contour integral representation: gapped case

The error of the contour integral representation is determined by N_{pole}. At finite temperature $N_{\text{pole}} = 2Q$, while at zero temperature $N_{\text{pole}} = Q$, with Q being the quadrature points on one loop of the contour. The performance of the algorithm is studied by the minimum number of N_{pole} such that $\Delta\rho_{\text{rel}}$ (the L^1 error in the electronic density per electron) is smaller than 10^{-6}. For a given temperature, the chemical potential μ is set to satisfy

$$\operatorname{Tr} P = N_{\text{Electron}}. \tag{4.3}$$

In our setup the energy gap $E_g \approx 0.01$ Hartree $= 0.27$ eV and $E_M \approx 4$ Hartree. Therefore, this system can be regarded as a crude model for semiconductor with a small energy gap. The number of $N_{\rm pole}$ and the error $\Delta\rho_{\rm rel}$ are shown in Table 1 with respect to $\beta\Delta E$ ranging between $4,000$ and up to $270,000$. Because of the existence of the finite energy gap, the performance is essentially independent of $\beta\Delta E$, as is clearly shown in Table 1.

Table 1 $N_{\rm pole}$ and L^1 error of electronic density per electron with respect to various $\beta\Delta E$. The energy gap $E_g \approx 0.01$. The contour integral representation for gapped system at finite temperature is used for the calculation. The performance of the algorithm depends weakly on $\beta\Delta E$.

$\beta\Delta E$	$N_{\rm pole}$	$\Delta\rho_{\rm rel}$
$4,208$	40	5.68×10^{-7}
$8,416$	44	3.86×10^{-7}
$16,832$	44	3.60×10^{-7}
$33,664$	44	3.55×10^{-7}
$67,328$	44	3.57×10^{-7}
$134,656$	44	3.47×10^{-7}
$269,312$	44	3.55×10^{-7}

When the temperature is low and therefore when β is large, as discussed before the finite temperature result is well approximated by the zero temperature Fermi operator, i.e., the matrix sign function. In such case the quadrature formula is given by (2.9). Only the contour that encircles the spectrum lower than chemical potential is calculated, and $N_{\rm pole} = Q$.

In order to study the dependence of $\Delta\rho_{\rm rel}$ on the number of poles $N_{\rm pole}$, we tune artificially the chemical potential to reduce the energy gap to 10^{-6} Hartree. Figure 6 shows the exponential decay of $\Delta\rho_{\rm rel}$ with respect to $N_{\rm pole}$. For example, in order to reach the 10^{-6} error criterion, $N_{\rm pole} \approx 50$ is sufficient. The increase in $N_{\rm pole}$ is very small compared to the large decrease of energy gap and this is consistent with the logarithmic dependence of $N_{\rm pole}$ on E_g given by (2.7).

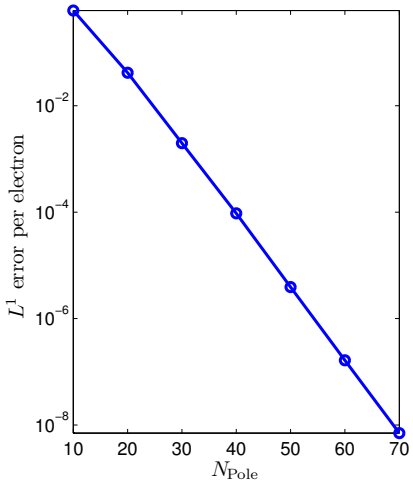

Figure 6 The lin-log plot of the L^1 error of electronic density per electron with respect to $N_{\rm pole}$. The energy gap $E_g \approx 10^{-6}$. The contour integral representation for gapped system at zero-temperature is used for calculation.

4.2 Contour integral representation: gapless case

For gapless systems such as metallic systems, our quadrature formula in (2.10) exploits the effective gap on the imaginary axis due to finite temperature. In the following results the chemical potential is set artificially so that $E_g = 0$. $E_M \approx 4$ Hartree and the error criterion is still 10^{-6} as in the gapped case. Table 2 reports the number of poles N_{pole} and the error $\Delta\rho_{\mathrm{rel}}$ with respect to $\beta\Delta E$ ranging from $4,000$ up to 4 million. These results are further summarized in Figure 7 to show the logarithmic dependence of N_{pole} on $\beta\Delta E$, as predicted in the analysis of (2.14).

Table 2 N_{pole} and L^1 error of electronic density per electron with respect to various $\beta\Delta E$. $E_g = 0$. The contour integral representation for gapless system is used for the calculation.

$\beta\Delta E$	N_{pole}	$\Delta\rho_{\mathrm{rel}}$
$4,208$	58	1.90×10^{-7}
$8,416$	62	5.32×10^{-7}
$16,832$	66	8.28×10^{-7}
$33,664$	72	3.55×10^{-7}
$67,328$	76	3.46×10^{-7}
$134,656$	80	1.69×10^{-7}
$269,312$	84	8.89×10^{-8}
$538,624$	88	7.09×10^{-8}
$1,077,248$	88	8.94×10^{-7}
$2,154,496$	88	4.25×10^{-7}
$4,308,992$	92	3.43×10^{-7}

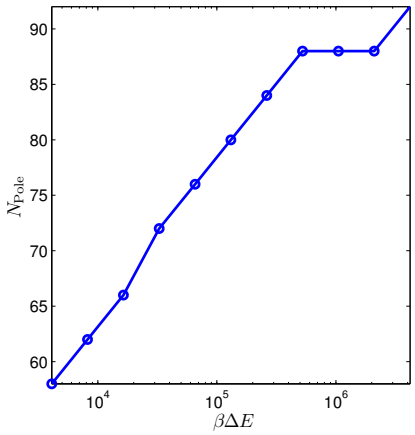

Figure 7 Log-lin plot of N_{pole} with respect to $\beta\Delta E$. The contour integral representation for gapless system is used for the calculation.

4.3 Multipole representation

The approach (3.9) based on the multipole representation has three parts of error: the finite-term multipole expansion, the finite-term Chebyshev expansion for the tail part, and the truncated matrix-matrix multiplication in the Chebyshev expansion.

The error from the multipole expansion is well controlled by P in (3.9). When $P = 16$, $\frac{1}{3^P} \sim \mathcal{O}(10^{-8})$. The number of groups N_G is usually no more than 20, and therefore the

error introduced by multipole expansion is around $\mathcal{O}(10^{-7})$, which is much less than the error criterion 10^{-6}.

The number of terms in the Chebyshev expansion for the tail part N_{Cheb} is $\mathcal{O}(\frac{\beta\Delta E}{M_{\text{pole}}})$, with M_{pole} being the number of poles excluded in the tail part in the pole expansion. The truncation radius for the tail part is $\mathcal{O}(\exp(-C\frac{\beta\Delta E}{M_{\text{pole}}}))$. In order to reach a fixed target accuracy, we set M_{pole} to be proportional to $\beta\Delta E$. Due to the fact that $M_{\text{pole}} \approx 2^{N_G} \approx 2^{\frac{N_{\text{pole}}}{P}}$, N_{pole} grows logarithmically with respect to $\beta\Delta E$.

The target accuracy for the Chebyshev expansion is set to be 10^{-7} and the truncation radius for the tail is set to be 4 for the metallic system under consideration. For $\beta\Delta E = 4208$, M_{pole} is set to be 512 so that the error is smaller than 10^{-6}. For other cases, M_{pole} scales linearly with $\beta\Delta E$. The lin-log plot in Figure 8 shows the logarithmic dependence of N_{pole} with respect to $\beta\Delta E$. For more detailed results, Table 3 measures M_{pole}, N_{pole}, N_{Cheb}, and $\Delta\rho_{\text{rel}}$ for $\beta\Delta E$ ranging from 4000 up to 1 million. For all cases, N_{Cheb} is kept as a small constant. Note that the truncation radius is always set to be a small number 4, and this indicates that the tail part is extremely localized in the multipole representation due to the effectively raised temperature.

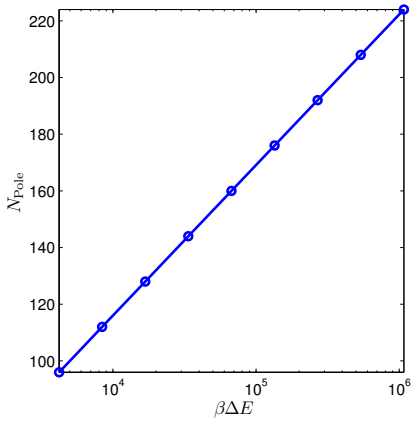

Figure 8 log-lin plot of N_{pole} with respect to $\beta\Delta E$. The multipole representation is used for the calculation.

Table 3 indicates that the error exhibits some slight growth. We believe that it comes from the growth of the number of groups in the multipole representation (3.9) and also the extra log log dependence on $\beta\Delta E$ (see [12] for details). When compared with the results reported in Table 2, we see that for the current application to electronic structure, the contour integral representation outperforms the multipole representation in terms of both the accuracy and the number of poles used.

5 Conclusion

We propose two approaches for the expansion of Fermi operator: a rational approximation based on the contour integral idea introduced in [8] and a variant of the multipole representation in [12] using only simple poles. Both approximations result in logarithmic scaling complexity with respect to $\beta\Delta\epsilon$ with small prefactor. Fast algorithms for electronic structure calculations can be obtained by combining these approaches with the algorithm introduced in [13] for extracting the diagonal of the inverse of a matrix.

Table 3 The number of poles calculated $N_{\rm pole}$, the order of Chebyshev expansion for the tail part $N_{\rm Cheb}$, and the L^1 error of electronic density per electron with respect to various $\beta\Delta E$. The number of poles excluded in the tail part $M_{\rm pole}$ is chosen to be proportional to $\beta\Delta E$.

$\beta\Delta E$	$M_{\rm pole}$	$N_{\rm pole}$	$N_{\rm Cheb}$	$\Delta\rho_{\rm rel}$
4,208	512	96	22	4.61×10^{-7}
8,416	1,024	112	22	4.76×10^{-7}
16,832	2,048	128	22	4.84×10^{-7}
33,664	4,096	144	22	4.88×10^{-7}
67,328	8,192	160	22	4.90×10^{-7}
134,656	16,384	176	22	4.90×10^{-7}
269,312	32,768	192	22	6.98×10^{-7}
538,624	65,536	208	22	3.20×10^{-6}
1,077,248	131,072	224	22	7.60×10^{-6}

Acknowledgement This is a continuation of the work that was done jointly with Roberto Car, to whom the authors are very grateful for many stimulating discussions.

References

[1] Baroni, S. and Giannozzi, P., Towards very large-scale electronic-structure calculations, *Europhys. Lett.*, **17**(6), 1992, 547–552.

[2] Ceriotti, M., Kühne, T. D. and Parrinello, M., A hybrid approach to Fermi operator expansion, 2008, preprint. arXiv:0809.2232v1

[3] Ceriotti, M., Kühne, T. D. and Parrinello, M., An effcient and accurate decomposition of the Fermi operator, *J. Chem. Phys.*, **129**(2), 2008, 024707.

[4] Goedecker, S. and Colombo, L., Efficient linear scaling algorithm for tight-binding molecular dynamics, *Phys. Rev. Lett.*, **73**(1), 1994, 122–125.

[5] Goedecker, S. and Teter, M., Tight-binding electronic-structure calculations and tight-binding molecular dynamics with localized orbitals, *Phys. Rev. B*, **51**(15), 1995, 9455–9464.

[6] Goedecker, S., Linear scaling electronic structure methods, *Rev. Mod. Phys.*, **71**(4), 1999, 1085–1123.

[7] Greengard, L. and Rokhlin, V., A fast algorithm for particle simulations, *J. Comput. Phys.*, **73**(2), 1987, 325–348.

[8] Hale, N., Higham, N. J. and Trefethen, L. N., Computing A^α, $\log(A)$, and related matrix functions by contour integrals, *SIAM J. Numer. Anal.*, **46**(5), 2008, 2505–2523.

[9] Krajewski, F. R. and Parrinello, M., Stochastic linear scaling for metals and nonmetals, *Phys. Rev. B*, **71**(23), 2005, 233105.

[10] Liang, W. Z., Baer, R., Saravanan, C., et al, Fast methods for resumming matrix polynomials and Chebyshev matrix polynomials, *J. Comput. Phys.*, **194**(2), 2004, 575–587.

[11] Liang, W. Z., Saravanan, C., Shao, Y. H., et al, Improved Fermi operator expansion methods for fast electronic structure calculations, *J. Chem. Phys.*, **119**(8), 2003, 4117–4125.

[12] Lin, L., Lu, J., Car, R., et al, Multipole representation of the Fermi operator with application to the electronic structure analysis of metallic systems, *Phys. Rev. B*, **79**(11), 2009, 115133.

[13] Lin, L., Lu, J., Ying, L., et al, Fast algorithm for extracting the diagonal of the inverse matrix with application to the electronic structure analysis of metallic systems, *Comm. Math. Sci.*, 2009, submitted.

[14] Ozaki, T., Continued fraction representation of the Fermi-Dirac function for large-scale electronic structure calculations, *Phys. Rev. B*, **75**(3), 2007, 035123.

[15] Ying, L., Biros, G. and Zorin, D., A kernel-independent adaptive fast multipole algorithm in two and three dimensions, *J. Comput. Phys.*, **196**(2), 2004, 591–626.