Partial fraction decomposition of the Fermi function

Alexander Croy and Ulf Saalmann
Max-Planck-Institute for the Physics of Complex Systems, Nöthnitzer Str. 38, 01187 Dresden, Germany
(Received 8 April 2009; revised manuscript received 5 June 2009; published 7 August 2009)

A partial fraction decomposition of the Fermi function resulting in a finite sum over simple poles is proposed. This allows for efficient calculations involving the Fermi function in various contexts of electronic-structure or electron-transport theories. The proposed decomposition converges in a well-defined region faster than exponential and is thus superior to the standard Matsubara expansion.

DOI: 10.1103/PhysRevB.80.073102 PACS number(s): 71.15.Dx, 72.10.Bg, 02.30.Mv

I. INTRODUCTION

Many problems in electronic-structure and electron-transport calculations involve the evaluation of integrals containing the Fermi function. These are in general difficult to compute and therefore several approximation schemes have been developed.[1–7] Among them the Sommerfeld expansion[1] and the Matsubara expansion[2] being the most prominent ones. While the former is by construction useful for low temperatures, the latter provides, in principle, a way to cover the range from low to high temperatures. Moreover, it turns out that the expansion in a (finite) sum of simple poles is particularly suitable for evaluating the integrals by means of contour integration using the residue theorem. For example, finite temperature charge-density calculations only require the evaluation of a Green's function at a finite set of energies[8,9] given by the poles of the expansion. Recently the same concept was used for the auxiliary density-matrix propagation in the context of time-resolved electron transport in molecular wires.[10]

The major disadvantage of the Matsubara expansion consists in its poor convergence behavior, the error decreasing only linearly with the number of terms in the expansion. To overcome this issue other approximative expressions for the Fermi function have been studied[4–7] which also yield an expansion in terms of poles in the complex plane. For the resulting approximate function one can discriminate between band-Fermi functions with finite support[4–6] and functions which essentially cover the whole real line.[2,7] Here we derive an expansion of the Fermi function in terms of simple poles with particularly simple coefficients. We will show that it converges very rapidly with increasing order of the expansion in a well-defined region which is found to increase linearly with the order.

For the following discussion it is convenient to write the Fermi function $f(\varepsilon)$ in terms of a dimensionless variable x,

$$f(x) = \frac{1}{1+e^{x}} \quad \text{with} \quad x = \frac{\varepsilon-\mu}{kT}, \tag{1}$$

where μ is the chemical potential, T is the temperature, k is the Boltzmann factor, and ε denotes the energy. The expansion consists in finding a partial fraction decomposition (PFD) with simple poles of the form

$$f(x) = \frac{1}{2} - \sum_{p=-\infty}^{\infty} \frac{A_p}{x-x_p}, \tag{2}$$

where A_p are expansion coefficients and x_p are (possibly complex) poles. For practical purposes the sum over p is truncated and the Fermi function is approximated by $f(x) \approx f_N(x)$ with N being the number of terms in the expansion.

For example, the well-known Matsubara expansion[2] is given in terms of the purely imaginary zeros x_n of the denominator in Eq. (1), $x_p = \imath\pi(2p-1)$, which yields coefficients $A_p=1$ and gives

$$f_N(x) = \frac{1}{2} - \sum_{p=1}^{N} \left[\frac{1}{x+\imath\pi(2p-1)} + \frac{1}{x-\imath\pi(2p-1)} \right]. \tag{3}$$

For $N \to \infty$ in Eq. (3) the expansion becomes exact. However, the convergence is very slow, which renders the application of this expansion impractical especially for low temperatures. Specifically, denoting by

$$\delta f_N(x) = f(x) - f_N(x) \tag{4}$$

the deviation of the finite expansion from the exact function, one finds for the Matsubara expansion $\delta f_N(x) \propto 1/N$. For the band-Fermi function[4] this can be improved[5] to $\delta f_N(x) \propto 1/N^2$.

II. PARTIAL FRACTION DECOMPOSITION

The proposed PFD is obtained by first writing Eq. (1) as[6]

$$f(x) = \frac{1}{2} - \frac{1}{2}\tanh(x/2) = \frac{1}{2} - \frac{\sinh(x/2)}{2\cosh(x/2)}, \tag{5}$$

and second by expanding numerator and denominator in a power series, truncating the respective sums, such that the degree of the polynomial in the denominator is larger than the degree of the numerator polynomial. This procedure gives

$$f_N(x) = \frac{1}{2} - \frac{1}{2}\frac{P_{N-1}(x/2)}{Q_N(x/2)}, \tag{6}$$

with polynomials

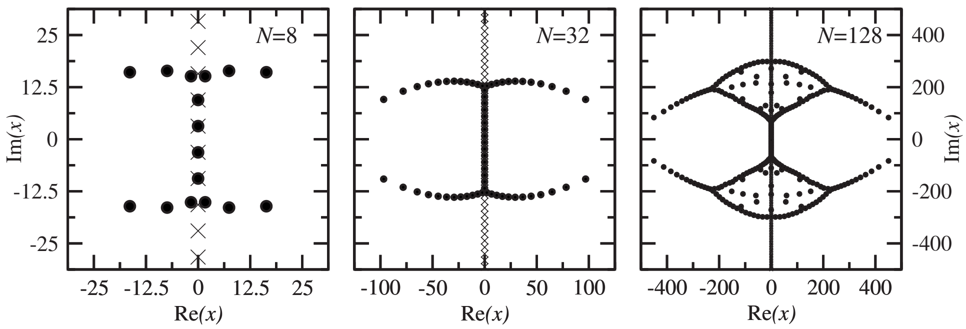

FIG. 1. Poles (symbol ●) of the PFD expansion, i.e., $\pm 2\sqrt{z_p}$, with z_p the eigenvalues of matrix, Eq. (9), for various orders N. For comparison the purely imaginary poles (symbol ×) of the Matsubara expansion, Eq. (3), are shown as well. Note the different scales of the three graphs.

$$P_N(x) = \sum_{m=0}^{N} \frac{x^{2m+1}}{(2m+1)!} \quad \text{and} \quad Q_N(x) = \sum_{m=0}^{N} \frac{x^{2m}}{(2m)!}. \tag{7}$$

This construction allows for a PFD, i.e., an expansion of the form

$$\frac{P_{N-1}(x/2)}{Q_N(x/2)} = \sum_{p=1}^{N} \left(\frac{A_p}{x/2 - x_p} + \frac{B_p}{x/2 + x_p} \right). \tag{8}$$

Here, $\pm x_p$ are the zeros of the polynomial Q_N, which appear in pairs since Q_N contains only even powers of $x/2$. It can be shown that the zeros can be obtained as $x_p = \sqrt{z_p}$, whereby the z_p are the eigenvalues of the following matrix[6] with $i,j = 1,\ldots,N$

$$Z_{ij} = 2i(2i-1)\delta_{j,i+1} - 2N(2N-1)\delta_{iN}. \tag{9}$$

The eigenvalues can be efficiently calculated using standard methods. In Fig. 1 we have plotted the poles of Eq. (8), given as $\pm 2x_p = \pm 2\sqrt{z_p}$, for three sets of eigenvalues z_p for different sizes $N \times N$ of the matrix Z_{ij}. As can be seen from the figure some of the poles $\pm 2x_p$ arising from the PFD are purely imaginary and are close to the Matsubara poles. On the other hand there are also poles with a nonvanishing real part, which display an irregular distribution. These very poles improve considerably the approximation for the Fermi function as we show below.

It remains to determine the corresponding expansion coefficients A_p and B_p in Eq. (8). Multiplying both sides of this equation by $(x/2 - x_k)$ and letting $x \to 2x_k$ leaves on the right side of Eq. (8) only the term A_k, which is thus given as

$$A_k = \lim_{x \to 2x_k} (x/2 - x_k) \frac{P_{N-1}(x/2)}{Q_N(x/2)} = \lim_{\eta \to 0} \frac{\eta P_{N-1}(x_k + \eta)}{Q_N(x_k + \eta)}. \tag{10}$$

By means of the definitions [Eq. (7)] one finds from this limit $A_k \equiv 1$ and similarly $B_k \equiv 1$. Thus we arrive at the main result of this Brief Report: The Fermi function can be approximated by the finite sum

$$f_N(x) = \frac{1}{2} - \sum_{p=1}^{N} \left(\frac{1}{x + 2\sqrt{z_p}} + \frac{1}{x - 2\sqrt{z_p}} \right), \tag{11}$$

with z_p the eigenvalues of matrix, Eq. (9). The formal structure of this approximation is similar to the Matsubara expansion, Eq. (3). However, taking advantage of having complex rather than purely imaginary poles makes the PFD for given order N of the expansion vastly superior to the Matsubara expansion. This can be seen in Fig. 2, where we have shown both expansions for different orders N. Whereas the Matsubara expansion, Eq. (3), in Fig. 2(a) does not give a reasonable representation for any of the orders shown, the PFD expansion, Eq. (11), in Fig. 2(b) improves rapidly with increasing order.

III. CONVERGENCE PROPERTIES

From Fig. 2 it becomes clear that the PFD is indeed converging faster than the Matsubara expansion. In the following we will quantify the rate of convergence as $N \to \infty$ and give a range for x where this convergence behavior can be expected.[11] Regarding the PFD one makes two observations:

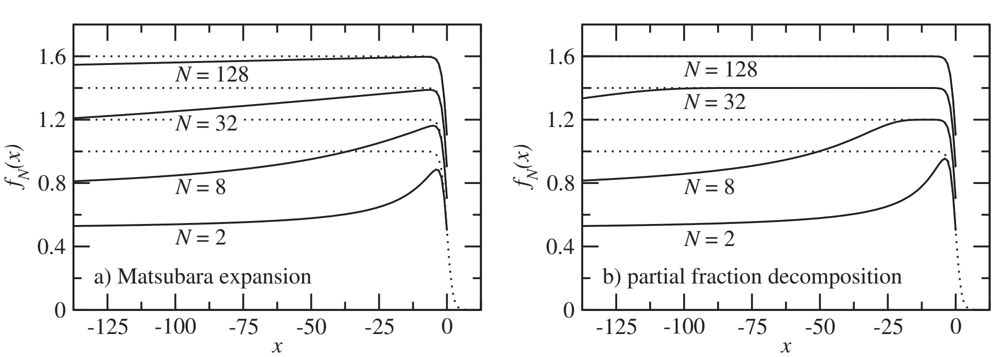

FIG. 2. Approximated Fermi function $f_N(x)$ for expansion orders $N=2,8,32,128$ (full lines). Panel (a): Matsubara expansion, Eq. (3), panel (b): partial fraction decomposition, Eq. (11). The curves are shown for $x<0$ only and are vertically shifted by 0.2 for better visibility. The exact Fermi function, Eq. (1), is denoted by dotted lines.

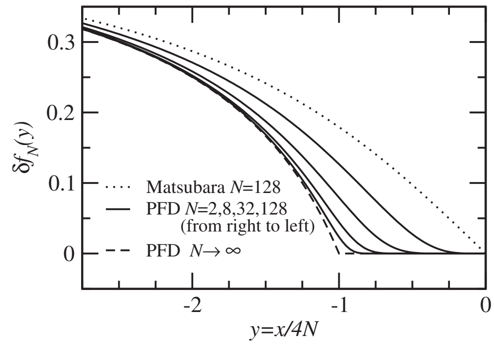

FIG. 3. Deviation δf_N of the approximated Fermi function form the exact one as defined in Eq. (4) as a function of the scaled argument $y=x/4N$ for $N=2,8,32,128$ (solid lines) and the asymptotic expression (dashed line) as given by Eq. (12). The dotted line shows δf_{128} for the Matsubara expansion.

First, in terms of the scaled variable $y=x/4N$ one finds in the limit of large N,

$$\lim_{N\to\infty} \delta f_N(x=y4N) = \begin{cases} 0 & \text{for } \dfrac{x}{4N}=y\geq -1 \\ \dfrac{1}{2}\left(1+\dfrac{4N}{x}\right)=\dfrac{1}{2}\left(1+\dfrac{1}{y}\right) & \text{for } \dfrac{x}{4N}=y\leq -1 \end{cases}. \tag{12}$$

The asymptotic function (12) is shown along with deviations $\delta f_N(y)$ for various finite N in Fig. 3. Second, in the range $-4N\leq x$, i.e., $-1<y$, the rate of convergence is given by the asymptotic expression

$$\delta f_N(x=y4N) \approx \frac{(x/2)^{2N}}{(2N)!}=\frac{(y2N)^{2N}}{(2N)!}, \tag{13}$$

which due to the factorial in the denominator decreases faster than exponential. Equations (12) and (13) are the main results of this section. They corroborate the statement that the PFD is expected to yield a better convergence and allow to estimate the error in actual calculations.

In the remaining part of this section we will justify and discuss Eqs. (12) and (13). Considering the case $y<-1$, one finds from Eq. (11), that a finite expansion behaves as $f_N(x)\propto 1/x$ for $x\to-\infty$. Since the Fermi function gives $f(x)=1$ for $x\to-\infty$ one expects qualitatively the behavior given in Eq. (12). This holds true for any expansion resulting in a finite sum over simple poles including the Matsubara expansion, which is shown for $N=128$ as dotted line in Fig. 3. In order to verify that this behavior is indeed restricted to $y<-1$, or equivalently to $x\leq -4N$, we write the polynomial Q_N from Eq. (7) explicitly as

$$Q_N(x/2)=Q_N(y2N)=\sum_{m=0}^{N} q_{mN}(y)$$

with

$$q_{mN}(y)=\frac{(2N)^{2m}}{(2m)!}y^{2m}. \tag{14}$$

Assuming $y<-1$, we see that the ratio of two successive terms

$$\frac{q_{mN}(y)}{q_{m-1N}(y)}=y^2\frac{(2N)^2}{2m(2m-1)} \tag{15}$$

is always larger than 1; the terms are monotonically increasing. Thus terms with $m\gg 1$ dominate the sum and we replace the coefficients in q_{mN} by the coefficient from q_{NN}, i.e., instead of the sum [Eq. (14)] we define

$$\tilde{Q}_N(y2N)=\sum_{m=0}^{N}\tilde{q}_{mN}(y)$$

with

$$\tilde{q}_{mN}(y)=\frac{(2N)^{2N}}{(2N)!}y^{2m}. \tag{16}$$

It turns out that in the limit $N\to\infty$ this sum becomes equal to $Q_N(y2N)$, which can be seen by considering the difference of the newly defined terms in Eq. (16) from the original terms in Eq. (14). For $m=N-n$ one gets

$$1-\frac{q_{N-n,N}(y)}{\tilde{q}_{N-n,N}(y)}=1-(2N)^{-2n}\frac{(2N)!}{(2N-2n)!}=\frac{n/2-n^2}{N}+\mathcal{O}\left(\frac{1}{N^2}\right). \tag{17}$$

This expression vanishes for $n=0$ and can be made arbitrarily small by increasing N for all $n\ll N$. Terms with larger n can be neglected because they are exponentially small compared to those with smaller n. Since the sum [Eq. (16)] is a geometric series we obtain for $N\to\infty$

$$Q_N(y2N)=\tilde{Q}_N(y2N)=\frac{y^2}{y^2-1}q_{NN}(y). \tag{18}$$

Analogous considerations for the other polynomial from Eqs. (7) yield

$$P_{N-1}(y2N)\approx\frac{(2N)^{2N-1}}{(2N-1)!}\frac{y^{2N+1}}{y^2-1}=\frac{y^2}{y^2-1}p_{N-1,N}(y) \tag{19}$$

with $p_{mN}(y)=(y2N)^{2m+1}/(2m+1)!$ and we get as an approximation for the ratio, once again using $N\gg 1$,

$$\frac{P_{N-1}(y2N)}{Q_N(y2N)}\approx\frac{1}{y}=\frac{4N}{x}, \tag{20}$$

which explains the asymptotic behavior of $\delta f_N(x)$ in Eq. (12) for $x<-4N$.

Turning now to the case $y>-1$, we first note that there is a crossover for the ratio [Eq. (15)] at $|y|N$; whereas for $m<|y|N$ the terms are increasing, they decrease for $m>|y|N$. Thus, for large N the polynomial expression (6) converges to the exact expression (5) and the deviation $\delta f_N(x)$ vanishes as given by Eq. (12) for $x>-4N$. In order to quantify the rate of convergence it is useful to define complementary sums to P_N and Q_N, namely,

$$\bar{P}_N(x) = \sum_{m=N+1}^{\infty} \frac{x^{2m+1}}{(2m+1)!} \quad \text{and} \quad \bar{Q}_N(x) = \sum_{m=N+1}^{\infty} \frac{x^{2m}}{(2m)!}. \tag{21}$$

Therewith the deviation reads

$$\delta f_N(y4N) = \frac{\sinh(y2N) - \bar{P}_{N-1}(y2N)}{2\cosh(y2N) - 2\bar{Q}_N(y2N)} - \frac{\sinh(y2N)}{2\cosh(y2N)} \approx e^{-y2N}[\bar{Q}_N(y2N) - \bar{P}_{N-1}(y2N)]. \tag{22}$$

The approximation in the second line applies to large values of N. We can choose, for any given $y<0$, N sufficiently large such that $\exp(-y2N) \gg \exp(+y2N)$. The infinite sums defined in Eq. (21) become small compared to the exponentials, $\bar{Q}_N(y2N) \ll 1 \ll \exp(-y2N)$ and $\bar{P}_{N-1}(y2N) \ll 1 \ll \exp(-y2N)$. It remains to estimate their behavior for large N which can be done in analogy to the considerations for Q_N and P_N, cf. Eqs. (18) and (19). Here the ratio of successive terms as defined in Eq. (15) is always smaller than 1 and the first terms in the sum can be used to estimate the sums. One gets

$$\bar{P}_N(y2N) \approx \frac{p_{N,N}(y)}{1-y^2} \quad \text{and} \quad \bar{Q}_N(y2N) \approx \frac{q_{N+1,N}(y)}{1-y^2}. \tag{23}$$

This directly leads to Eq. (13) and concludes the derivation.

Figure 4 shows this estimate along with the numerically calculated deviation δf_N as a function of the expansion order N for selected values of x. Even for small values of N an overall good agreement is found. Moreover, one sees that the deviation δf_N is of order 1 as long as $N < -x/4$. However, for $N > -x/4$ (this is where the dashed lines start) it decreases very rapidly due to the factorial in the denominator in Eq. (13).

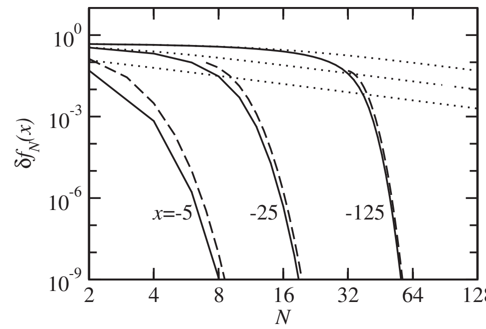

FIG. 4. Deviation δf_N of the approximated Fermi function form the exact one as defined in Eq. (4) as a function of the expansion order N for three arguments $x=-5,-25,-125$. We compare the Matsubara expansion (dotted lines) and PFD (solid lines). For the latter case we show also the asymptotic behavior according to Eq. (13) by dashed lines.

IV. CONCLUSIONS

We have proposed the expansion, Eq. (11), of the Fermi function, Eq. (1), by using a partial fraction decomposition. Its application requires only the diagonalization of a matrix, given in Eq. (9), which has the same dimension N as the expansion. The expansion converges faster than exponential with increasing order N for arguments $|x|<4N$. In other words, the approximation becomes not only more accurate for higher orders, it can also be used for a wider range of arguments. An estimate for the error is explicitly given by Eq. (13). Due to the beneficial convergence properties and the straightforward implementation we expect the PFD to be of great value in any application based on an expansion of the Fermi function as sum of over simples poles. Finally, we would like to notice that an analogous expansion can be found for the Bose-Einstein distribution.

[1] N. W. Ashcroft and N. D. Mermin, *Solid State Physics* (Saunders College, Philadelphia, 1976).
[2] G. D. Mahan, *Many Particle Physics*, 2nd ed. (Plenum, New York, 1990).
[3] S. Goedecker, Phys. Rev. B **48**, 17573 (1993).
[4] D. M. C. Nicholson, G. M. Stocks, Y. Wang, W. A. Shelton, Z. Szotek, and W. M. Temmerman, Phys. Rev. B **50**, 14686 (1994).
[5] D. M. C. Nicholson and X.-G. Zhang, Phys. Rev. B **56**, 12805 (1997).
[6] F. Gagel, J. Comput. Phys. **139**, 399 (1998).
[7] T. Ozaki, Phys. Rev. B **75**, 035123 (2007).
[8] K. Wildberger, P. Lang, R. Zeller, and P. H. Dederichs, Phys. Rev. B **52**, 11502 (1995).
[9] M. J. Watrous, L. Wilets, and J. J. Rehr, Phys. Rev. E **59**, 3554 (1999).
[10] S. Welack, M. Schreiber, and U. Kleinekathöfer, J. Chem. Phys. **124**, 044712 (2006).
[11] In this section we consider only negative arguments $x<0$ but the discussion applies in an analogous manner also to $x>0$.