Chemical Physics Letters 584 (2013) 182–187

Contents lists available at ScienceDirect

Chemical Physics Letters

journal homepage: www.elsevier.com/locate/cplett

On spectral quadrature for linear-scaling Density Functional Theory

Phanish Suryanarayana

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

ARTICLE INFO

Article history:
Received 14 March 2013
In final form 13 August 2013
Available online 23 August 2013

ABSTRACT

We provide a unified description of the Fermi operator expansion and recursion methods within the technique of spectral quadrature. Through rigorous error estimates, we prove that this approach is linear-scaling, stable and exponentially convergent. We use this analysis to determine the influence of smearing, band-gap, position of Fermi energy, and spectral width of the Hamiltonian on the convergence rates obtained in practical calculations. Additionally, we establish that super-geometric convergence can be achieved when the *erfc* function is used for smearing. We validate the spectral quadrature method and the accuracy of our analysis by means of selected examples.

1. Introduction

Formulations of Density Functional Theory (DFT) which solve for the orbitals have an $\mathcal{O}(N^3)$ scaling with respect to the number of atoms [1,2]. To overcome this restrictive scaling, there has been great emphasis on the development of $\mathcal{O}(N)$ methods [3,4]. One such technique is the Fermi operator expansion (FOE), where the density matrix is expanded in terms of polynomials [5–7]. Such an expansion can be achieved particularly efficiently through the purification method [8–16]. Another related approach is the recursion method developed by Haydock et al. [17–19]. In this method, the projected density of states (PDOS) is determined through a continued fraction representation, and the quantities of interest are then obtained by integration over it.

A variant of the recursion method was proposed by Nex [20], where instead of calculating the PDOS, the integrals over it are directly evaluated via Gauss quadrature. An improved algorithm for the calculation of the Gauss quadrature nodes and weights was presented by Bai et al. [21], wherein they focussed on the calculation of the partial eigenvalue sum i.e. band structure energy. However, because of the expense associated with generating the Gauss quadrature rules, they proposed the use of a Monte-Carlo simulation technique to approximate the band structure energy. Recently, Suryanarayana et al. [22] utilized the Gauss quadrature technique to coarse-grain DFT, whereby the computational effort to study crystal defects is significantly reduced. Henceforth, we shall refer to the recursion method and its variants mentioned above as the Gauss spectral quadrature method.

In this letter, we analyze the class of methods wherein integrals over the PDOS are approximated via interpolatory quadrature, an approach which we refer to as spectral quadrature (SQ). Gauss SQ represents a particular variant of the SQ method. Similarly, the Clenshaw-Curtis/Fejér SQ method, which is equivalent to the FOE method in terms of Chebyshev polynomials, is another variant of the SQ method. Here, we derive rigorous error estimates to establish the convergence, stability and convergence rate of the SQ method. We utilize this analysis to discuss the scaling and performance of the approach, which is also applicable to the FOE and recursion methods. However, from the perspective of implementation, the SQ and FOE methods have some notable differences. The SQ method does not involve the calculation of the complete density matrix, and therefore utilizes matrix–vector multiplications rather than matrix–matrix multiplications. Unlike FOE, the Fermi energy is not required as input to the method, and quantities like the entropic contribution due to smearing can be evaluated without any additional effort. This limitation of FOE can be overcome by storing the intermediate matrices (e.g. Chebyshev), though at the cost of substantial computer memory requirement. Finally, the SQ method can achieve variable resolution in the simulation domain by spatially varying the order of the quadrature rule.

One of the common approaches to solve the DFT problem is through self-consistent field (SCF) method [1]. In each iteration of the SCF method, the electron density

$$\rho(\mathbf{x}) = 2\sum_n g_n |\psi_n(\mathbf{x})|^2 \tag{1}$$

is calculated by solving the linear eigenvalue problem

$$\mathcal{H}\psi_n(\mathbf{x}) = \lambda_n \psi_n(\mathbf{x}) \tag{2}$$

for the orbitals $\psi_n(\mathbf{x})$, subject to the constraint on the total number of electrons $N_e = 2\sum_n g_n$. We denote the Hamiltonian by $\mathcal{H}$, and λ_n are its eigenvalues corresponding to $\psi_n(\mathbf{x})$. The orbital occupations are given by the Fermi–Dirac distribution $g_n = (1+\exp((\lambda_n - \lambda_f)/\sigma))^{-1}$, where σ is the smearing and λ_f is the Fermi energy. Once the self-consistent solution has been obtained, the evaluation of the ground-state free energy involves the calculation of the band

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

0009-2614/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.cplett.2013.08.035

structure energy U, and the entropic contribution S of the fractional orbital occupations. They can be expressed as [23]

$$U = 2\sum_n g_n \lambda_n, \tag{3}$$

$$S = 2\sigma\sum_n [g_n \log(g_n) + (1-g_n)\log(1-g_n)]. \tag{4}$$

In this Letter, for a given Hamiltonian, we focus on the $\mathcal{O}(N)$ calculation of ρ, U and S using the SQ method. This combined with $\mathcal{O}(N)$ approaches for the electrostatics [24–26] enables the $\mathcal{O}(N)$ solution of the complete DFT problem.

2. Theory

In Section 2.1, we reformulate the DFT problem. We derive rigorous error estimates in Section 2.2, which is used to determine the convergence rates in Section 2.3. We discuss the scaling and performance of the SQ method in Section 2.4, and finally discuss alternative forms of smearing in Section 2.5.

2.1. Integral representations

Consider a self-adjoint Hamiltonian $\mathcal{H}$ over a finite-dimensional Hilbert space with a localized orthonormal basis $\{\eta_p\}_{p=1}^M$ and inner product $\langle .,. \rangle$. Let the matrix representation of $\mathcal{H}$ be sparse, with ordered eigenvalues $\{\lambda_m\}_{m=1}^M$. We map the spectrum of $\mathcal{H}$ to the interval $[-1, 1]$ using the transformation $\hat{\mathcal{H}} = (\mathcal{H} - \chi\mathcal{I})/\xi$, where $\xi = (\lambda_M - \lambda_1)/2, \chi = (\lambda_M + \lambda_1)/2$, and $\mathcal{I}$ represents the identity operator. Starting from the PDOS $\sum_{n=1}^M \langle \zeta, \psi_n \rangle \delta(\hat{\lambda} - \hat{\lambda}_n) \langle \psi_n, \zeta \rangle$ [18], we define the integrated PDOS as [22]

$$\mathcal{E}_{\zeta,\zeta}(\hat{\lambda}) = \begin{cases} 0, & \text{if } \hat{\lambda} < \hat{\lambda}_1 \\ \sum_{n=1}^{m}\sum_{p=1}^{M}\sum_{q=1}^{M} \psi_{n,p}\psi_{n,q}\zeta_p\zeta_q, & \text{if } \hat{\lambda}_m \leqslant \hat{\lambda} < \hat{\lambda}_{m+1}, \\ \sum_{n=1}^{M}\sum_{p=1}^{M}\sum_{q=1}^{M} \psi_{n,p}\psi_{n,q}\zeta_p\zeta_q, & \text{if } \hat{\lambda}_M < \hat{\lambda} \end{cases} \tag{5}$$

where $\psi_n(\mathbf{x}) = \sum_{p=1}^M \psi_{n,p}\eta_p(\mathbf{x})$ and $\{\hat{\lambda}_m\}_{m=1}^M$ are the eigenvalues of $\hat{\mathcal{H}}$. Note that the derivative of $\mathcal{E}_{\zeta,\zeta}(\hat{\lambda})$ represents the density of states of $\hat{\mathcal{H}}$ projected onto the function $\zeta(\mathbf{x}) = \sum_{p=1}^M \zeta_p \eta_p(\mathbf{x})$. Let $g(\hat{\lambda}) = \left(1 + \exp\left((\hat{\lambda} - \hat{\lambda}_f)/\hat{\sigma}\right)\right)^{-1}$ denote the scaled and shifted Fermi–Dirac distribution, where $\hat{\sigma} = \sigma/\xi$and $\hat{\lambda}_f = (\lambda_f - \chi)/\xi$. We define $u(\hat{\lambda}) = (\xi\hat{\lambda} + \chi)g(\hat{\lambda})$ and $s(\hat{\lambda}) = \sigma[g(\hat{\lambda})\log g(\hat{\lambda}) + (1 - g(\hat{\lambda}))\log(1 - g(\hat{\lambda}))]$. Using Eq. 5, we arrive at the Riemann–Stieltjes integrals [21,22,27]

$$\begin{aligned} N_e &= 2\sum_{p=1}^M \int_{-1}^1 g(\hat{\lambda})\, \mathrm{d}\mathcal{E}_{\eta_p,\eta_p}(\hat{\lambda}), \\ \rho(\mathbf{x}_0) &= 2\int_{-1}^1 g(\hat{\lambda})\, \mathrm{d}\mathcal{E}_{\zeta_0,\zeta_0}(\hat{\lambda}), \\ U &= 2\sum_{p=1}^M \int_{-1}^1 u(\hat{\lambda})\, \mathrm{d}\mathcal{E}_{\eta_p,\eta_p}(\hat{\lambda}), \\ S &= 2\sum_{p=1}^M \int_{-1}^1 s(\hat{\lambda})\, \mathrm{d}\mathcal{E}_{\eta_p,\eta_p}(\hat{\lambda}), \end{aligned} \tag{6}$$

where $\mathbf{x}_0$ is any chosen point in space and $\zeta_0(\mathbf{x}) = \sum_{p=1}^M \eta_p(\mathbf{x}_0)\eta_p(\mathbf{x})$.

2.2. Error estimates

We consider the representative integral

$$I(f) = \int_{-1}^1 f(\hat{\lambda})\, \mathrm{d}\mathcal{E}_{\zeta,\zeta}(\hat{\lambda}), \tag{7}$$

where $f(\hat{\lambda})$ denotes any of the integrands in Eq. 6, and $\mathcal{E}_{\zeta,\zeta}(\hat{\lambda})$ denotes the corresponding integrated PDOS. In the SQ method, we approximate $I(f)$ with a K-point interpolatory quadrature rule $I_K(f)$ having nodes $\{\hat{\lambda}_k^\zeta\}_{k=1}^K$ and weights $\{\hat{w}_k^\zeta\}_{k=1}^K$ i.e.

$$I_K(f) = \sum_{k=1}^K \hat{w}_k^\zeta f(\hat{\lambda}_k^\zeta), \tag{8}$$

whose error

$$E(f) = I(f) - I_K(f). \tag{9}$$

We start by expressing the function $f(\hat{\lambda})$ in terms of a uniformly convergent Chebyshev polynomials (T_k) expansion as

$$f(\hat{\lambda}) = \sum_{k=0}^\infty a_k T_k(\hat{\lambda}), \tag{10}$$

whereby

$$E(f) = \sum_{k=0}^\infty a_k E(T_k). \tag{11}$$

Consider an ellipse ε_r in the complex plane $\mathbb{C}$ within and on which $f(z), z \in \mathbb{C}$ is analytic. Let ε_r have foci at ± 1, semi-major axis a and semi-minor axis $\sqrt{a^2-1}$. Further, let r denote the sum of the semi-major and semi-minor axes i.e. $r = a + \sqrt{a^2-1}$. It can be shown that [28]

$$|a_k| \leqslant \frac{2M_r(f)}{r^k}, \tag{12}$$

where $M_r(f)$ is the maximum value of $|f(z)|$ on ε_r. Consider a quadrature rule which integrates polynomials up to degree $nK-1$ exactly $(1 \leqslant n \leqslant 2)$. We have $E(T_k) = 0, k = 0, 1, \ldots nK-1$. For $k \geqslant nK$,

$$|E(T_k)| \leqslant \left| \int_{-1}^1 T_k(\hat{\lambda}) \mathrm{d}\mathcal{E}_{\zeta,\zeta}(\hat{\lambda}) \right| + |I_K(T_k)| \leqslant C^\zeta + W_K^\zeta, \tag{13}$$

where $C^\zeta = \int_{-1}^1 \mathrm{d}\mathcal{E}_{\zeta,\zeta}(\hat{\lambda})$ and $W_K^\zeta = \sum_{k=1}^K |\hat{w}_k^\zeta|$. Therefore,

$$\begin{aligned} |E(f)| &\leqslant \sum_{k=0}^\infty |a_k||E(T_k)| \leqslant 2M_r(f)\sum_{k=0}^\infty \frac{|E(T_k)|}{r^k} \\ &\leqslant \frac{2(C^\zeta + W_K^\zeta)M_r(f)}{r^{nK-1}(r-1)}. \end{aligned} \tag{14}$$

Specializing this bound for the integrals in Eq. 6, we arrive at the following error estimates

$$\begin{aligned} |N_e - N_{eK}| &\leqslant \frac{4(M + W_K)M_r(g)}{r^{nK-1}(r-1)}, \\ |\rho(\mathbf{x}_0) - \rho_K(\mathbf{x}_0)| &\leqslant \frac{4(C^{\zeta_0} + W_K^{\zeta_0})M_r(g)}{r^{nK-1}(r-1)}, \\ |U - U_K| &\leqslant \frac{4(M + W_K)M_r(u)}{r^{nK-1}(r-1)}, \\ |S - S_K| &\leqslant \frac{4(M + W_K)M_r(s)}{r^{nK-1}(r-1)}. \end{aligned} \tag{15}$$

Above, we have used the notation $W_K = \sum_{p=1}^M W_K^{\eta_p}$, and the identity $C^{\eta_p} = 1$. The value of n is determined by the variant of SQ employed. For an interpolatory SQ with the quadrature nodes fixed apriori, we have $n = 1$. An example is the Clenshaw-Curtis/Fejér SQ, which in principle is equivalent to the FOE method in terms of Chebyshev polynomials. However, when all the nodes are treated as unknowns, we obtain Gauss SQ with the maximum value of $n = 2$. We note that the bounds derived above are particularly tight when the integrands $(g(\hat{\lambda}), u(\hat{\lambda})$ and $s(\hat{\lambda}))$ are analytic in only a small