EFFICIENT MAXIMUM ENTROPY ALGORITHMS FOR ELECTRONIC STRUCTURE

Author(s): R. N. Silver, H. Roeder, A. F. Voter, J. D. Kress

Submitted to: SCS High Performance Computing Mtg., New Orleans, LA
April 13-17, 1996
DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.
EFFICIENT MAXIMUM ENTROPY ALGORITHMS
FOR ELECTRONIC STRUCTURE *

R. N. Silver, H. Roeder, A. F. Voter, J. D. Kress
Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545
e-mail: rns@loke.lanl.gov

Keywords: Electronic Structure, Materials, Maximum Entropy, Statistical Analysis

Abstract

Two Chebyshev recursion methods are presented for calculations with very large sparse Hamiltonians, the kernel polynomial method (KPM) and the maximum entropy method (MEM). If limited statistical accuracy and energy resolution are acceptable, they provide linear scaling methods for the calculation of physical properties involving large numbers of eigenstates such as densities of states, spectral functions, thermodynamics, total energies for Monte Carlo simulations and forces for molecular dynamics. KPM provides a uniform approximation to a DOS, with resolution inversely proportional to the number of Chebyshev moments, while MEM can achieve significantly higher, but non-uniform, resolution at the risk of possible artifacts. This paper emphasizes efficient algorithms.

1 Introduction

Many computational physics problems involve calculations with very large sparse Hamiltonian matrices. Finding all eigenvectors and eigenvalues requires cpu time scaling as \(O(N^3)\) and memory scaling as \(O(N^2)\), which is impractical. For ground or isolated eigenstates the preferred method is Lanczos diagonalization, which uses only matrix-vector-multiply (MVM) operations and requires cpu and memory scaling as \(O(N)\). But new \(O(N)\) methods are needed for properties involving many eigenstates such as the density of states (DOS) and spectral functions, and for quantities that can be derived from DOS such as thermodynamics, total energies for electronic structure and forces for molecular dynamics and Monte Carlo simulations. Limited energy resolution and statistical accuracy are often acceptable provided the uncertainties can be quantified. Maximum entropy (MEM) (Mead 1984; Drabold 1993) has been a popular approach, usually fitting power moments of a DOS or spectral function. However, the non-linear convex optimization algorithms required to find MEM solutions may be difficult to implement for large numbers of power moments and for singular structures in DOS.

This paper considers methods which use Chebyshev moments rather than power moments of a DOS or spectral function. Chebyshev moments are advantageous for several reasons: their calculation is much less sensitive to the limitations of machine precision; the Hessian of the MEM convex optimization problem is much better conditioned; and the isomorphism to Fourier series enables use of advanced methods such as FFT’s and Gibbs damping. Moreover, a linear Chebyshev approximation to a DOS constructed from such moments, termed the Kernel Polynomial Method (KPM) (Silver & Roeder 1994; Wang 1994), provides a computationally simple alternative to MEM. This paper discusses the generation of Chebyshev moment data, describes the KPM; and then discusses an efficient MEM algorithm. We necessarily repeat discussion of KPM presented at the HPC ’95 (Silver, et al. 1995), as needed for the MEM algorithm that follows.

2 Data Generation

Consider the DOS as representative of the properties of interest. The first step is to scale the Hamiltonian, \(H = aX + b\) such that all eigenvalues \(x_n\) of \(X\) lie between \(-1\) and \(+1\). The DOS is then

\[
D(x) = \frac{1}{N} \sum_{n=1}^{N} \delta(x - x_n) .
\]  

(1)
The data about $D(x)$ consists of Chebyshev moments,
\[
\mu_m = \text{Tr}\{T_m(X)\} = \int_{-1}^{1} T_m(x) D(x)dx .
\]
Calculation of moments uses Chebyshev recursion,
\[
T_{m+1}(X) = 2XT_m(X) - T_{m-1}(X) ,
\]
requiring the same optimized matrix-vector-multiply (MVM) algorithm in Lanczos methods. Unlike Lanczos recursion, Chebyshev recursions are numerically stable to arbitrarily large numbers of moments without any need for expensive reorthogonalization. One can use the rules for multiplying Chebyshev polynomials, e.g. $T_2 = 2T_1T_1 - 1$, such that $M/2$ MVM’s are required to generate $M$ moments. Exact evaluation of $M$ moments using cpu time $\propto O(N^2M/2)$ generates $T_m(X)|i >$ on each basis state $| i >$, and then
\[
\mu_{2m} = \frac{2}{N} \sum_{i=1}^{N} < i|T_m(X)T_m(X)|i > - 1 .
\]
Similarly for $\mu_{2m-1}$.

Generation of moments using cpu time $\propto O(NM)$ can be accomplished either by stochastic methods (Silver & Roeder 1994) or by local truncation (Voter, Kress and Silver 1996). The stochastic method with cpu scaling as $O(NMN_r)$ uses estimators
\[
\tilde{\mu}_m \approx \frac{1}{N_r} \sum_{r} < r|T_m(X)|r > ,
\]
where $| r >$ are $N_r$ Gaussian random vectors. Such data have statistical variance proportional to $(N/MN_r)^{-1}$ which may be expressed directly in terms of moments. More sophisticated choices of random vector appear to reduce statistical variance (Drabold and Sankey 1993), but they introduce statistical bias.

The local truncation method may be applicable if the density matrix has only local off-diagonal elements, as in tight-binding Hamiltonians for insulators. Chebyshev moments are estimated using a locally truncated Hamiltonian $H_i$,
\[
\hat{\mu}_m \approx \sum_{i} < i|T_m(X_i)|i > ,
\]

generating data with a systematic error determined by the truncation range. ‘Logical’ truncation (Voter, Kress and Silver 1996) appears to converge more rapidly and smoothly than ‘physical’ truncation schemes (Goedecker and Colombo 1994). Cpu scales as $O(NMJ)$, where $J$ is the average number of states in the truncation range. Exact moment derivatives (related to forces) can also be calculated from a Chebyshev derivative formula.

\section{The Kernel Polynomial Method}

Cpu time and memory limit the number of moments $M$ and their statistical and systematic errors. Given such limited data, KPM and MEM are two ways to estimate DOS. KPM provides a linear Chebyshev approximation to a DOS with a uniform resolution in $\phi = \cos^{-1}(x)$. It is based on an exact moment expansion,
\[
D(x) = \frac{1}{\sqrt{1-x^2}} \left[ \mu_0 + 2 \sum_{m=1}^{\infty} \mu_m T_m(x) \right] .
\]
The KPM truncates this expansion at $M$ moments and introduces a factor $g_m$ to damp the Gibbs phenomenon,
\[
D_K(x) = \frac{1}{\sqrt{1-x^2}} \left[ \mu_0 + 2 \sum_{m=1}^{M} \mu_m g_m T_m(x) \right] .
\]
Let $D(\phi) \equiv \sin(\phi)D(X)$ and $T_m(x) = \cos(mx)$. Then $D_K(\phi)$ is both a simple convolution and a truncated Fourier series,
\[
\delta_K(\phi) = \frac{1}{2\pi} \left[ g_0 + 2 \sum_{m=1}^{M} g_m \cos(m\phi) \right] .
\]
The “kernel” $\delta_K(\phi)$ is a $2\pi$-periodic polynomial approximation to a Dirac delta function, analogous to the resolution function of a spectrometer. Resolution is uniform in $\phi$ with width $\Delta \phi \propto M^{-1}$. If $g_m = 1$, at large $| \phi |$ the kernel is oscillatory with period $\Delta \phi = \pi/M$ within an envelope function decreasing slowly as $1/\phi^2$. The result is the Gibbs phenomenon of a lack of uniform convergence at singular structures in DOS. An optimal $g_m$ can be determined variationally by requiring the kernel to be a polynomial of degree $M$, strictly positive, normalized and have minimal variance in $\phi$ (Silver, et al. 1996). Specifically, by the Fejer-Riesz theorem
\[
\delta_K(\phi) = \frac{1}{2\pi} \left| \sum_{\nu=0}^{M} a_{\nu} e^{i\nu\phi} \right|^2 ; \quad g_m = \sum_{\nu=0}^{M-m} a_{\nu} a_{\nu+m} .
\]
Minimize the variance,
\[
\Delta \phi^2 \equiv \int_{-\pi}^{\pi} \phi^2 \delta_K(\phi)d\phi
\]
\[
\simeq \int_{-\pi}^{\pi} (2 - 2\cos(\phi)) \delta_K(\phi)d\phi = 2g_0 - 2g_1 .
\]
subject to a normalization constraint, \( \int_{-\pi}^{\pi} \delta_K(\phi) d\phi = 1 \). The variational problem,

\[
Q = g_1 - \lambda g_0 = \sum_{\nu=0}^{M-1} a_{\nu} a_{\nu+1} - \lambda \sum_{\nu=0}^{M} a_{\nu} a_{\nu} , \tag{12}
\]

results in

\[
\frac{\delta Q}{\delta a_{\nu}} = 0 \Rightarrow a_{\nu+2} - 2\lambda a_{\nu+1} + a_{\nu} = 0 . \tag{13}
\]

The solution to Eq. (13) is

\[
a_{\nu} = \frac{U_{\nu}(\lambda)}{\sqrt{\sum_{\nu=0}^{M} U_{\nu}^2(\lambda)}} \sin((\nu + 1)\phi_{\lambda}) ; \cos(\phi_{\lambda}) = \lambda , \tag{14}
\]

where the \( U_{\nu} \) are Chebyshev polynomials of the second kind. The same kernel is also obtained by minimizing the uniform norm (Jackson 1933). Its envelope function decreases exponentially at large \(|\phi|\).

Figure 1 illustrates the application of KPM to the electronic structure of a 216 atom Si supercell using a tight binding Hamiltonian (Silver et al. 1996) based on the parameterization of Goodwin, et al. (Goodwin, Pettifor & Skinner 1989). This system is small enough to be exactly diagonalized. Vertical lines are at the energies of the exact eigenstates and their height is proportional to their degeneracy. The solid line is the KPM approximation to the DOS obtained for 200 Chebyshev moments. A Fermi energy \( E_F \) is the energy at which the cumulative DOS \( C_K(E) \) equals the number of electrons. The total band energy \( E_B \) is then the cumulative energy \( E_K(E) \) at \( E_F \). For band energies KPM converges \( \propto M^{-2} \) reaching \( 10^{-5} \) relative accuracy at about \( M \approx 150 \). Subtle details about removing systematic bias and achieving a faster convergence rate are discussed in (Voter, Kress & Silver 1996).

KPM can be applied to other properties such as spectral functions (Silver et al. 1996),

\[
A(\omega) = \lim_{\eta \to 0+} \frac{1}{\pi} \text{Im} \left\{ \langle \Psi_0 | O^\dagger \left( \frac{1}{\omega - i\eta} - \frac{1}{\omega + i\eta} \right) O | \Psi_0 \rangle \right\} , \tag{15}
\]

where \( O \) is an operator. KPM approximations use moments \( \mu_m^O = \langle \Psi_0 | O^\dagger T_m(X) O | \Psi_0 \rangle \).
4 The Maximum Entropy Method

MEM uses the same Chebyshev moment data as KPM. The entropy,

\[ S \equiv \int_0^\pi \left[ D(\phi) - D_\circ(\phi) + D(\phi) \ln \left( \frac{D(\phi)}{D_\circ(\phi)} \right) \right] d\phi \quad . \tag{16} \]

Here \( D_\circ(\phi) \) is a default model for the DOS in the absence of data. Consider the case where the data are subject to Gaussian independent statistical noise,

\[ \mu_m = \mu_m + \eta_m \quad ; \quad \mu_\eta = 0 \quad ; \quad \sigma_m^2 \delta_{mm'} \quad . \tag{17} \]

(\( \mu \) denotes the statistical expectation value of the random variable that follows it.) If the data are exact, \( \mu \) represents the numerical precision required of the MEM fit to the data. The primal optimization problem uses moments as variables to maximize entropy as a function of \( \phi \) constrained by the known moments.

Maximize

\[ Q_p \equiv S - \frac{\chi^2}{2\alpha} = \sum_{m=0}^M \left( \frac{\mu_m - \mu_m}{\sigma_m^2} \right)^2 \quad . \tag{18} \]

The statistical regularization parameter \( \alpha \) sets a balance between the fit, measured by \( \chi^2 \), and an information measure, \(-S\), of distance between the inferred \( \phi \) and the default model \( \phi_\circ \). (Alternatively, \( 1/\alpha \) is a Lagrange multiplier.) The moment required to constrain normalization, \( \mu_0 = 1 \). Taking the limit \( \alpha \to 0 \) strictly enforces normalization.

Our MEM algorithm consists of three nested loops: iterations in \( \alpha \), until a stopping criterion is reached; at each \( \alpha \), Newton-Raphson iterations of a dual optimization problem defined using Lagrange multipliers as variables to solve for the MEM \( \phi \); at each \( \alpha \) and MEM \( \phi \) conjugate gradient iterations to apply the Hessian onto a vector.

Popular stopping criteria for \( \alpha \) are \( \chi^2 = M \) and \( \chi^2 - 2\alpha S = M \), although many other criteria are discussed in the literature. However, the algorithm for finding the MEM \( \phi \) tends to be unstable if initiated at such small \( \alpha \). Instead, start at large \( \alpha \approx \chi^2_0 \), and use \( \phi_\circ \) to initiate the optimization of \( \phi \). Progress down in \( \alpha \) such that \( \alpha^{k+1} = \alpha^k/2 \). If this is unstable, halve the step down in \( \alpha \) repeatedly until stability is reached. At each \( \alpha \), use \( D^{\phi}(\phi) \) as the starting point for the optimization of \( D^{k+1}(\phi) \). Once the stopping criterion is passed, perform a golden search for the optimal \( \alpha \).

In the case of exact moment data, set \( \sigma_m \) to the numerical precision required, which can be very small. In our applications to electronic structure, errors of one part in 10^5 or smaller were used. Iterate \( \alpha \to 0 \) until the entropy \( S \) saturates at an \( \alpha \)-independent value.

Given an \( \alpha \), a variety of algorithms have been developed to find MEM solutions (Turek 1988; Skilling 1993). The primal problem maximizes \( Q_p \) as a function of \( \phi \),

\[ \frac{\delta Q_p}{\delta \phi} = -\ln \left( \frac{D(\phi)}{D_\circ(\phi)} \right) + \sum_{m=0}^M \frac{\mu_m - \mu_m}{\alpha^2} \cos(m\phi) = 0 \quad , \tag{19} \]

which has a unique solution. Define parameters \( \bar{\mu} \) by

\[ \mu_m - \mu_m + \alpha^2 \lambda_m = 0 \quad . \tag{20} \]

Then the \( \phi \) satisfying Eq. (19) is

\[ D(\phi) = D_\circ(\exp \left( -\sum_{m=0}^M \lambda_m \cos(m\phi) \right) ) \quad . \tag{21} \]

This form is also obtained by maximizing entropy subject to Lagrange contrains on moments with Lagrange multipliers \( \bar{\lambda} \).

However, a dual optimization problem as a function of the \( M \) Lagrange multipliers (Auyeng & Mersereau 1991) solves the same problem, and it is more stable numerically than the primal problem. The \( \bar{\lambda} \) of the dual problem vary more slowly than the \( \phi \), and they are a finite rather than a continuous set of variables. The quantity,

\[ Q_d \equiv \ln \left( \int_0^\pi D(\phi) d\phi \right) + \sum_{m=0}^M \left[ \bar{\mu}_m \lambda_m + \frac{\alpha^2 \lambda_m^2}{} \right] \quad , \tag{22} \]

is maximized as a function of the \( \bar{\lambda} \) when Eq. (20) is satisfied. Away from the maximum, define

\[ \frac{\delta Q_d}{\delta \lambda_m} \equiv \xi_m \equiv \bar{\mu}_m - \mu_m + \alpha^2 \lambda_m \quad . \tag{23} \]

Then,

\[ Q_d = Q_p + \sum_{m=0}^M \frac{\xi_m^2}{2\alpha^2} \quad . \tag{24} \]

The Hessian of the dual problem is a positive definite \( M \times M \) matrix and a simple function of the moments,

\[ \frac{\partial^2 Q_d}{\partial \lambda_m \partial \lambda_m'} = \frac{\mu_m + \mu_m}{2} + \alpha^2 \delta_{mm'} \quad . \tag{25} \]

A solution to Eq. (20) may be found by Newton-Raphson iteration. Beginning with some starting \( \bar{\lambda}^0 \), the \( n + 1 \)th step is

\[ \bar{\lambda}^{n+1} = \bar{\lambda}^0 - H_n^{-1} \xi_n \quad . \tag{26} \]
The quantity, $H^{-1} \xi$, may be calculated, e.g., by conjugate gradients. In view of Eq. (24), converging bounds at the $n$'th iteration are $Q_d^n \geq Q^\infty \geq Q_p^n$ where $Q^\infty \equiv \lim_{n \to \infty} \{Q_d^n, Q_p^n\}$. This provides stopping criteria for the iteration.

For electronic structure applications, high numerical precision (e.g. $\approx 10^{-6}$) is needed for accurate energy derivatives. Careful attention to how the MEM algorithms are discretized then becomes very important. Practical fast Fourier transform (FFT) algorithms calculate the $\mu_m = \int_0^\pi \cos(m\phi)D(\phi)d\phi$ by sampling the domain $0 \leq \phi \leq \pi$ at a discrete set of $N_p$ equally spaced points. The Shannon sampling theorem says that such naive discretization corresponds to representing a DOS in a $N_p$-order truncated Chebyshev series. In effect, MEM is used to infer an $N_p$-order Chebyshev approximation from knowledge of $M$ true moments.

But typical DOS contain singular structures such as $\delta$-functions, van Hove singularities, band edges, etc. These structures are properly described by an infinite order Chebyshev expansion. As discussed previously for the KPM, abrupt truncation of a Chebyshev expansion at $N_p$ terms results in the Gibbs phenomenon; i.e. singular structures in the true DOS at $\theta_0$ induce oscillations in Chebyshev approximated DOS of period $\Delta \phi = \pi/N_p$ with an envelope function decreasing slowly as $1/(\phi - \phi_0)^2$. While the moment data satisfy the Hausdorff conditions for the existence of a positive solution (Mead & Papanicolaou 1984), the added requirement that the solution be an $N_p$-order Chebyshev series is stronger. For the exact moment problem, the $\alpha$-iteration may have difficulty forcing $\chi^2 \to 0$ and saturating the entropy $S$.

Fortunately, the kernel polynomial method (KPM) provides a solution to this discretization problem. In the MEM problem, replace the $M$ Chebyshev moment data $\hat{\mu}_m$ by modified moments $\hat{\mu}_m g_m^{N_p}$, where the $g_m^{N_p}$ are the Gibbs damping factors in Eq. (10). In other words, change the goal of the MEM algorithm to the inference of a $N_p$-order KPM approximation to the DOS. Iteration toward $\chi^2 \to 0$ and saturation of $S$ becomes easy.

By choosing $N_p \gg M$, MEM can achieve significant resolution enhancements over KPM from the same $M$ moments. In tests with tight binding Hamiltonians for the electronic structure of Si, band energies converge approximately 4 times faster with MEM than with KPM. For the example in Fig. 1, MEM reaches $10^{-5}$ accuracy at $M \approx 35$. Setting $N_p \geq 4M$ is adequate to achieve this gain. The cpu time required by MEM scales as $O(M^2)$, and it is negligible compared...
to the cpu time required to generate the moment data. Use of MEM cuts the overall cpu requirements by at least a factor of 4 over KPM. Isolated features in DOS, such as individual states and band edges, may converge even faster.

Figure 2 compares MEM and KPM for the DOS of a 1D polaron formation problem. The Hamiltonian consists of an electron placed into a 10,000 atom chain with a Peierls distortion, which is then allowed to relax resulting in the polaron state at \( E = 1.0 \). This example demonstrates that MEM achieves dramatically better energy resolution than KPM for isolated states and band edges, but it tends to "ring" (or oscillate) in smooth positive regions of a DOS when singular structures, such as Van Hove singularities, are nearby. For such regions of a spectrum, MEM may converge more slowly than KPM. A solution to the ubiquitous MEM ringing problem most likely will require a modification to the entropy functional to include local smoothness constraints (Silver & Martz 1994).

Energy derivatives needed for molecular dynamics and Monte Carlo can be derived for MEM using the same expressions for exact derivatives of moments. The statistical error for stochastic methods using Gaussian random vectors can easily be accommodated, because the covariance of the moments is proportional to the Hessian. Details of these extensions will be presented elsewhere.

Acknowledgements

Supported in part by the Office of Basic Energy Sciences of the U. S. Department of Energy.

References


Biography

Richard N. Silver is a staff member in the Condensed Matter and Statistical Physics group at the Los Alamos National Laboratory. He received his B.S. in physics in 1966 and his Ph. D. in 1971 in theoretical physics, both from the California Institute of Technology. He held postdoctoral positions at Brown University '71-'72 and CalTech '72-'74. He has been at Los Alamos since '74 in various capacities including 7 years as neutron scattering group leader. Topics of research publications include particle physics, laser spectroscopy, semiconductor physics, neutron scattering experiments, many-body physics, statistical data analysis and fluid turbulence.
DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.