Maximum entropy and the problem of moments: A stable algorithm

K. Bandyopadhyay and A. K. Bhattacharya*
Department of Physics, University of Burdwan, Burdwan, West Bengal 713104, India

Parthapratim Biswas† and D. A. Drabold‡
Department of Physics and Astronomy, Ohio University, Athens, Ohio 45701, USA
(Received 27 December 2004; published 12 May 2005)

We present a technique for entropy optimization to calculate a distribution from its moments. The technique is based upon maximizing a discretized form of the Shannon entropy functional by mapping the problem onto a dual space where an optimal solution can be constructed iteratively. We demonstrate the performance and stability of our algorithm with several tests on numerically difficult functions. We then consider an electronic structure application, the electronic density of states of amorphous silica, and study the convergence of the Fermi level with increasing number of moments.

DOI: 10.1103/PhysRevE.71.057701 PACS number(s): 05.10.−a, 71.10.−w, 71.20.−b

One of the fixed themes of physics is the solution of inverse problems. A ubiquitous example in theoretical physics is the "classical moment problem" (CMP), in which only a finite set of power moments of a non-negative distribution function p is known, and the full distribution is needed [1]. It is obvious that the solution for p is *not unique* for a finite set of moments. This nonuniqueness suggests the need for a "best guess" for p, based upon the available information. With its ultimate roots in 19th century statistical mechanics and a subsequent strong justification based upon probability theory, the "maximum entropy" (maxent) method has provided an extremely successful variational principle to address this type of inverse problem [2]. Collins and Wragg used the maxent method to solve the CMP for a modest number of moments [3]. In a comprehensive paper with seminal applications, Mead and Papanicolaou [4] solved the CMP with maximum entropy techniques and proposed the first practical numerical scheme to solve the moment problem for up to 15 moments. In a host of subsequent papers, the utility of the method as an unbiased and surprisingly efficient (rapidly convergent) solution of the CMP has been established. The principle has been used extensively in a number of diverse applications ranging from image construction to spectral analysis [5], large-scale electronic structure problems [6,7], series extrapolation and analytic continuation [8–10], quantum electronic transport [11], ligand-binding distribution in polymers [12], and transport planning [13].

There exist a number of maximum entropy algorithms [4,7,14–16] that have been developed over the last two decades. Many of the algorithms (but not all) are constrained by the number of moments that they can deal with and become unreliable when the number of constraints exceeds a problem-dependent upper limit. As the number of moments increases, the calculation of moments (particularly the power moments) becomes more sensitive to machine precision and the optimization problem becomes ill conditioned. It has been observed that implementation of a maxent algorithm with more than 20 power moments is notoriously difficult even with extended precision arithmetic and it rarely gives any further information on the nature of the distribution. The use of orthogonal polynomials as basis set significantly improves the accuracy and remedies most of the problems that one encounters with power moments.

In this paper we present an iterative approach to construct the maxent solution of CMP, which is based upon discretization of the Shannon entropy functional [17]. The essential idea is to discretize Shannon entropy and map the problem from the primal space onto the dual space where an optimal solution can be constructed iteratively without the need of matrix inversion. We discuss theoretical ideas and develop algorithms that can be used with both power and Chebyshev moments. The stability and the accuracy of the method are discussed with reference to two numerically nontrivial examples—a uniform distribution and a double δ function. We illustrate the usefulness of our technique by computing the electronic density of states (EDOS) of amorphous silica with particular emphasis on convergence of the Fermi level as a function of number of moments.

The starting point of our approach is to use a discretized form of the Shannon entropy functional [17] $S[p]$ using a quadrature formula,

$$S = -\int p(x)\ln p(x)dx \approx -\sum_{j=1}^{n} w_j p_j \ln p_j. \tag{1}$$

Here w_j and x_j are the weights and abscissas of any accurate quadrature formula, say the Gauss-Legendre, and without any loss of generality we restrict ourselves to $x \in [0,1]$. We want to maximize S subject to the discretized moment constraints

$$\sum_{j=1}^{n} w_j x_j^i p_j = \sum_{j=1}^{n} a_{ij}\tilde{p}_j = \mu_i, \quad i = 1,2,\ldots,m, \tag{2}$$

where we define $\tilde{p}_j = w_j p_j$ and $a_{ij} = x_j^i$. The entropy optimization program (EOP) can now be stated so as to optimize the Lagrangian function,

*Electronic address: physakb@yahoo.com
†Electronic address: biswas@phy.ohiou.edu
‡Electronic address: drabold@ohio.edu

$$L(\tilde{\mathbf{p}},\eta) \equiv \sum_{j=1}^{n} \tilde{p}_j \ln\left(\frac{\tilde{p}_j}{w_j}\right) - \sum_{i=1}^{m} \tilde{\eta}_i \left(\sum_{j=1}^{n} a_{ij}\tilde{p}_j - \mu_i\right), \quad (3)$$

and the solution can be written as

$$\tilde{p}_j = w_j \exp\left(\sum_{i=1}^{m} a_{ij}\tilde{\eta}_i - 1\right), \quad j = 1,2,\ldots,n. \quad (4)$$

Since $\mathbf{w} \geqslant 0$, Eq. (4) implies that $\tilde{\mathbf{p}} \geqslant 0$. Furthermore, the conditions in Eqs. (2) and (4) can be lumped together,

$$h_i(\tilde{\eta}) \equiv \sum_{j=1}^{n} a_{ij} w_j \exp\left(\sum_{k=1}^{m} a_{kj}\tilde{\eta}_k - 1\right) - \mu_i = 0, \quad \forall\ i. \quad (5)$$

We now see from Eq. (5) that the original constrained optimization program is now reduced to an *unconstrained convex optimization program* involving the dual variables

$$\min_{\tilde{\eta}\in R^m} d(\tilde{\eta}) \equiv \sum_{j=1}^{n} w_j \exp\left(\sum_{i=1}^{m} a_{ij}\tilde{\eta}_i - 1\right) - \sum_{i=1}^{m} \mu_i \tilde{\eta}_i. \quad (6)$$

If the dual optimization program stated above has an optimal solution $\tilde{\eta}^*$, the solution $\tilde{p}_j(\eta^*)$ can be obtained from Eq. (4). Bergman has proposed an iterative method to minimize the dual objective function $d(\tilde{\eta})$ taking *only one* dual variable at a time [18]. The method starts with an arbitrarily chosen $\tilde{\eta}^{\mathbf{0}} \in R^m$, and then cyclically updates all the dual variables as follows.

Step 1. Start with any $\tilde{\eta}^{\mathbf{0}} \in R^m$ and a sufficiently small tolerance level $\epsilon > 0$. Set $k=0$ and obtain $\tilde{p}_j^0$.

Step 2. Let $i = (k \bmod m) + 1$. Solve the equation

$$\phi_i^k(\lambda^k) = \sum_{j=1}^{n} a_{ij}\tilde{p}_j^k \exp(a_{ij}\lambda^k) - \mu_i = 0. \quad (7)$$

Step 3. Update each component of $\tilde{\eta}$,

$$\tilde{\eta}_l^{k+1} = \begin{cases} \tilde{\eta}_l^k + \lambda^k & \text{if } l = i, \\ \tilde{\eta}_l^k & \text{if } l \neq i. \end{cases} \quad (8)$$

Step 4. If Eq. (5) is satisfied within the preset level of tolerance, then stop with $\eta^* = \tilde{\eta}^k$, and obtain the primal solution from Eq. (4). Otherwise, calculate

$$\tilde{p}_j^{k+1} = w_j \exp\left(\sum_{i=1}^{m} a_{ij}\tilde{\eta}_i^{k+1} - 1\right), \quad j = 1,2,\ldots,n \quad (9)$$

and go to step 2.

From a computational point of view, the most problematic part of the above algorithm is the solution of the set of Eq. (7) in step 2. In a variant of the above scheme known as the multiplicative algebraic reconstruction technique [19,20], one uses the following closed-form expression to approximate the correction term λ^k:

$$\lambda_i^k = \ln\left(\frac{\mu_i}{\sum_{j=1}^{n} a_{ij}\tilde{p}_j^k}\right). \quad (10)$$

Step 3 of the algorithm is now modified by substituting the expression above for λ_i^k in Eq. (8). A convergence theorem for the modified algorithm can be found in Lent [21]. It is, however, quite easy to see that the algorithm will fail unless for every $i=1,2,\ldots,m$, either

$$\mu_i > 0 \quad \text{and} \quad 0 \leqslant a_{ij} \leqslant 1, \quad j = 1,2,\ldots,n \quad (11)$$

or

$$\mu_i < 0 \quad \text{and} \quad 0 \geqslant a_{ij} \geqslant -1, \quad j = 1,2,\ldots,n. \quad (12)$$

We note that in this case we are assured of convergence of the solution of our discretized EOP because the condition (11) holds.

The EOP algorithm above can only be used provided that the condition stated by the inequality (11) or (12) is satisfied. This constrains us to apply the algorithm for power moments but neither of these two is necessarily true for other polynomial moments. In order to work with Chebyshev polynomials, we first employ the averages of shifted Chebyshev polynomials [22] of the first kind $T_n^*(x) = T_n(2x-1)$ to recast the entropy optimization program given by statement (5). The only change needed for this purpose is to redefine a_{ij} by $a_{ij} = T_i^*(x_j)$.

Our next step is to find a transformation that will convert the EOP into an equivalent problem in which all the program parameters are non-negative. For finding the necessary transformation, we define for $i=1,2,3,\ldots,m$,

$$u_j = [\max_j(-a_{ij})] + 1. \quad (13)$$

Obviously, for $i=1,2,3,\ldots,m$ and $j=1,2,3,\ldots,n$,

$$(u_i + a_{ij}) > 0. \quad (14)$$

Let us now define for $i=1,2,3,\ldots,m$,

$$M_i \equiv \max_j(u_j + a_{ij}), \quad t_i \equiv \frac{1}{m(M_i+1)}. \quad (15)$$

It is easy to see that the following relations hold for $i=1,2,3,\ldots,m$:

$$M_i > 0, \quad t_j > 0,$$

$$(M_i+1)t_j = \frac{1}{m}, \quad t_i(u_i + a_{ij}) \leqslant t_i M_i < \frac{1}{m}.$$

For $i=1,2,3,\ldots,m$ and $j=1,2,3,\ldots,n$, let us define

$$a'_{ij} \equiv t_i(u_i + a_{ij}). \quad (16)$$

Apparently, for $i=1,2,3,\ldots,m$ and $j=1,2,3,\ldots,n$, we have

$$\frac{1}{m} > a'_{ij} > 0, \quad 0 < \sum_{i=1}^{m} a'_{ij} = \sum_{i=1}^{m} t_i(u_i + a_{ij}) < 1. \quad (17)$$

It is interesting to note that if $\tilde{\mathbf{p}}$ is a feasible solution to the EOP involving averages of $T_n^*(x)$, then for $i=1,2,3,\ldots,m$

$$\sum_{j=1}^{n} a'_{ij}\tilde{p}_j = \sum_{j=1}^{n} t_i(u_i + a_{ij})\tilde{p}_j = t_i(u_i + \mu_i). \quad (18)$$

Hence, if we define for $i=1,2,3,\ldots,m$,

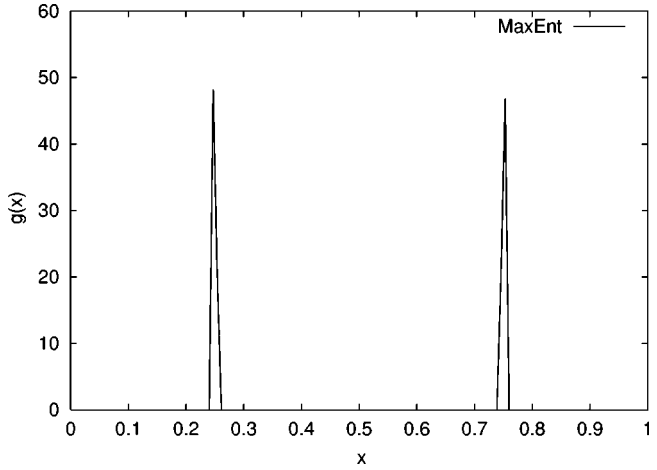

FIG. 1. Reconstruction of a double δ function $g(x)=\delta(x-1/4)+\delta(x-1/3)$ using the first 25 shifted Chebyshev moments.

$$\mu_i' \equiv t_i(u_i + \mu_i) = \sum_{j=1}^{n} a_{ij}' \tilde{p}_j. \tag{19}$$

It is easy to verify that $1/m > \mu_i'$ for $i=1,2,3,\ldots,m$. The transformed EOP has thus the same form as previously, except for the fact that we use Eq. (19) in place of Eq. (2). Since both a_{ij}' and μ_i' can take only positive values, a feasible solution to the original program can now be obtained by replacing a_{ij} and μ_{ij} in Eq. (5) by a_{ij}' and μ_{ij}' [23].

We consider two numerically difficult examples, a uniform distribution and a double δ function, to study the stability and accuracy of the algorithm. The Chebyshev moments of these two functions can be exactly calculated. Earlier efforts to reproduce these distributions have met with limited success because of the difficulty in matching a sufficient number of moments and because of the singular nature of the functions. It would be interesting to see how the algorithm performs in the case of (a) a uniform distribution $f(x)=1$, $x\in[0,1]$ and (b) a double δ function $g(x)=\delta(x-1/4)+\delta(x-3/4)$, $x\in[0,1]$.

The algorithm produces the uniform distribution correctly up to five decimal places. We found that the first 25 shifted Chebyshev moments are sufficient for this purpose. The fact that the end points have been produced so accurately without any spurious oscillations is a definitive strength of this approach and reflects the stability and accuracy of our algorithm. In Fig. 1, we have plotted the result for the double δ function. The result is equally convincing and certainly establishes the usefulness of this method over the other existing ones in the literature. In addition to these examples, we have also tested our algorithm to reconstruct a Tent map, a semicircular distribution, a square-root distribution, and a distribution with a gap in the spectrum. In all these cases, the algorithm correctly produces all the features of the distributions without failing. Throughout the work we have used double precision arithmetic to compute the moments and the distribution functions. The results clearly demonstrate that the algorithm is very stable and accurate and is capable of reproducing some very uncommon distributions (such as the double δ function) without any difficulty.

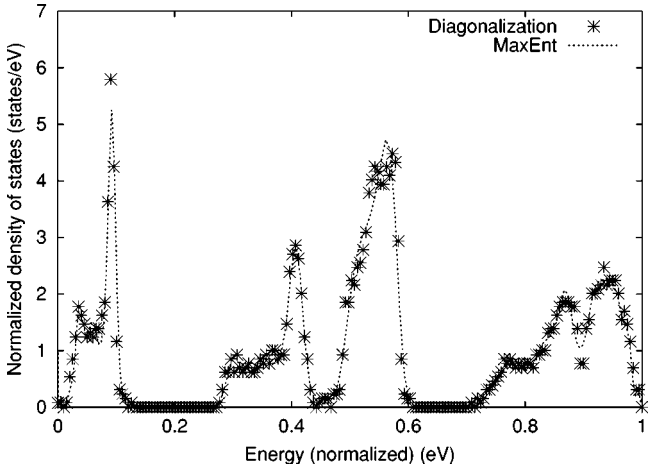

FIG. 2. Normalized electronic density of states (dotted line) of amorphous silica using the first 60 shifted Chebyshev moments. The distribution of energy eigenvalues (points) from direct diagonalization of the Hamiltonian matrix is also plotted in the figure. Normalized Fermi level is at 0.595 eV, which corresponds to −5.4655 eV in Fig. 3.

We now consider a practical case where analytical expressions for the moments are not available but can be computed numerically. An archetypal example is the calculation of an electronic density of states from its moments. In the context of solid state physics, the maxent method has been used profitably to calculate the density of electronic (vibrational) states from a knowledge of the moments of the Hamiltonian (dynamical) matrix. The computation of moments itself is an interesting problem in this field, and there are methods available in the literature that specifically address this issue [6,14]. Here one is interested in determining physical quantities such as the Fermi level and band energy of large systems (e.g., clusters, biological macromolecules, etc.) without diagonalizing the Hamiltonian matrix. For disordered systems, this is particularly suitable because of disordered scattering (of electrons) that washes out the van Hove singulari-

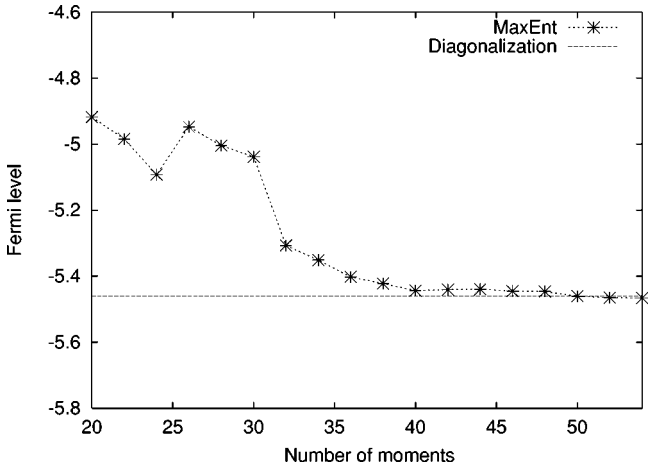

FIG. 3. Fermi level of amorphous silica as a function of number of the shifted Chebyshev moments. The value obtained from direct diagonalization of the Hamiltonian matrix is −5.465 eV and is plotted as a horizontal line in the figure.

ties in the electronic spectrum. A stable and accurate maxent algorithm, therefore, would be very useful in calculating electronic properties of amorphous semiconductors. The two examples discussed above suggest that we should be able to produce a complex electronic spectrum with a gap (or gaps) to a high degree of precision and hence the Fermi level and band energy. As for metallic systems, the determination of Fermi energy is a nontrivial problem for $O(n)$ methods. A primary requirement for a maxent algorithm in this case is that (1) it must produce the distribution accurately and (2) it must do so in a stable way using a sufficient number of moments to correctly produce the singularities of the spectrum. It is very pleasing to note that our algorithm does satisfy this requirement, and therefore may offer an alternative approach to computing the Fermi energy of metallic systems.

In Fig. 2, we have plotted the EDOS of amorphous silica using the first 60 moments and compared it to the result obtained by direct diagonalization of the Hamiltonian matrix. It is clear from the figure that all the features of the EDOS are correctly produced by our maxent algorithm. Finally, in Fig. 3 we have plotted the variation of Fermi energy with the number of moments. The Fermi energy is computed by integrating the normalized density of states to obtain the correct number of total electrons. It is clear from Fig. 3 that the Fermi energy starts to converge after the first 30 moments and eventually converges after 40 moments. We emphasize at this point that physical quantities that can be expressed as an average over the distribution functions (such as band energy, susceptibility, and specific heat) converge more rapidly than the distribution functions themselves. Since the distribution functions converged excellently in the present work, one is automatically assured of convergence of averaged quantities.

In conclusion, we present an algorithm for maximum entropy construction of a distribution from its moments. The algorithm is very stable and accurate and can handle a large number of moments (up to 500). The usefulness of this algorithm is demonstrated by constructing some numerically difficult distributions and applying it to amorphous silica to compute the electronic density of states and the Fermi level.

We acknowledge the support of the National Science Foundation under Grants No. DMR-0205858 and No. DMR-0310933.

[1] J. A. Shohat and J. D. Tamarkin, *The Problem of Moments* (American Mathematical Society, Providence, RI, 1963).
[2] E. T. Jaynes, *Probability Theory: The Logic of Science* (Cambridge University Press, Cambridge, U.K., 2003)
[3] R. Collins and A. Wragg, J. Phys. A **10**, 1441 (1977).
[4] L. R. Mead and N. Papanicolaou, J. Math. Phys. **25**, 2404 (1984).
[5] N. Canosa, R. Rossignoli, and L. Diambra, Phys. Lett. A **185**, 133 (1994).
[6] D. A. Drabold and O. F. Sankey, Phys. Rev. Lett. **70**, 3631 (1993).
[7] R. N. Silver and H. Röder, Phys. Rev. E **56**, 4822 (1997).
[8] C. M. Bender *et al.*, J. Math. Phys. **28**, 1016 (1987).
[9] D. Wang *et al.*, J. Math. Phys. **32**, 2258 (1991).
[10] D. A. Drabold and G. L. Jones, J. Phys. A **24**, 4705 (1991).
[11] P. A. Mello and Jean-Louis Pichard, Phys. Rev. B **40**, 5276 (1986).
[12] D. Poland, J. Chem. Phys. **113**, 4774 (2000).
[13] W.-H. Steeb, F. Solms, and R. Stoop, J. Phys. A **27**, L399 (1994).
[14] J. Skilling, in *Maximum Entropy and Bayesian Methods*, edited by J. Skilling (Kluwer, Dordrecht, 1989).
[15] I. Turek, J. Phys. C **21**, 3251 (1988).
[16] G. L. Bretthorst (unpublished).
[17] C. Shannon, Bell Syst. Tech. J. **27**, 379 (1948).
[18] L. M. Bergman, USSR Comput. Math. Math. Phys. **7**, 200 (1967).
[19] S. C. Fang, J. R. Rajasekara, and H.-S. J. Tsao, *Entropy Optimization and Mathematical Programming* (Kluwer Academic Publishers, Dordrecht, 1997).
[20] R. Gordon, R. Bender, and G. T. Herman, J. Theor. Biol. **29**, 471 (1970).
[21] A. Lent, in *Image Analysis and Evaluation*, edited by R. Shaw (SPSE, Washington, D.C., 1953).
[22] *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (Dover Publications, New York, 1972).
[23] The transformed problem in terms of a'_{ij} and μ'_i has exactly the same solution as the original problem. If the original problem is infeasible (due to inaccurate values of higher power moments, etc.), this gets reflected by the lack of positive definiteness of a'_{ij} and μ'_i.