Maximum entropy in the problem of moments

Lawrence R. Mead and N. Papanicolaou

Related Articles

Synchronization between integer-order chaotic systems and a class of fractional-order chaotic systems via sliding mode control
Chaos 22, 023130 (2012)

Optimal signal amplification in weighted scale-free networks
Chaos 22, 023128 (2012)

Auxiliary ECR heating system for the gas dynamic trap
Phys. Plasmas 19, 052503 (2012)

Resonant microwave-to-spin-wave transducer
Appl. Phys. Lett. 100, 182404 (2012)

Communication: Density functional theory overcomes the failure of predicting intermolecular interaction energies
J. Chem. Phys. 136, 161102 (2012)
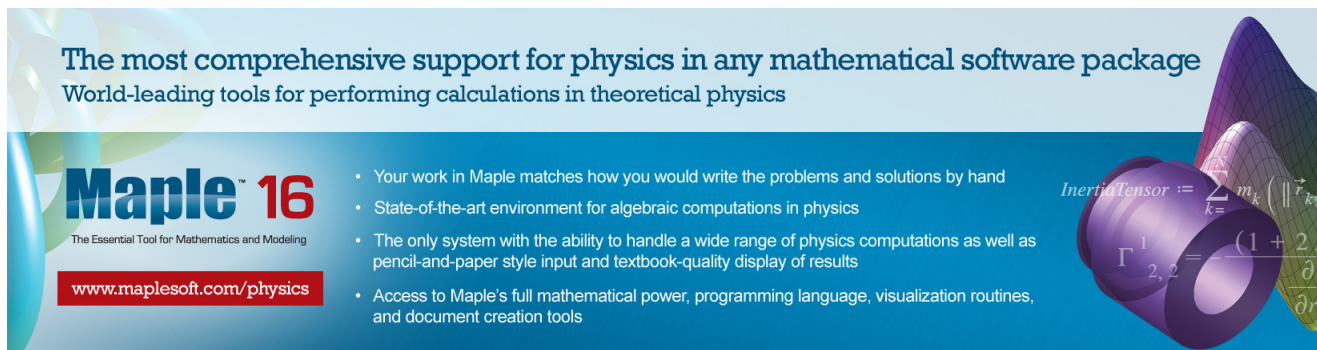
Additional information on J. Math. Phys.

Maximum entropy in the problem of moments

Lawrence R. Mead and N. Papanicolaou

Department of Physics, Washington University, St. Louis, Missouri 63130

The maximum-entropy approach to the solution of underdetermined inverse problems is studied in detail in the context of the classical moment problem. In important special cases, such as the Hausdorff moment problem, we establish necessary and sufficient conditions for the existence of a maximum-entropy solution and examine the convergence of the resulting sequence of approximations. A number of explicit illustrations are presented. In addition to some elementary examples, we analyze the maximum-entropy reconstruction of the density of states in harmonic solids and of dynamic correlation functions in quantum spin systems. We also briefly indicate possible applications to the Lee–Yang theory of Ising models, to the summation of divergent series, and so on. The general conclusion is that maximum entropy provides a valuable approximation scheme, a serious competitor of traditional Padé-like procedures.

PACS numbers: 02.60. + y, 75.10.Jm

I. INTRODUCTION

The maximum-entropy approach to the solution of underdetermined inverse problems was introduced some time ago.[1,2] Following the original contributions, there has been a long debate concerning the conceptual foundations of maximum entropy for problems outside the traditional domain of thermodynamics. The debate is currently more meaningful than ever in view of the augmenting list of successful practical applications[3] which have become possible because of the steadily increasing computing power available today. While our aim is not to engage in further conceptual ramifications of the rationale of maximum entropy,[4] we shall attempt to sharpen its mathematical foundations and to extend its applicability to various concrete problems encountered in quantum physics.

We consider the classical moment problem where a positive density $P(x)$ is sought from knowledge of its power moments

$$\int_a^b x^n P(x)dx = \mu_n, \quad n = 0,1,2,\ldots . \tag{1.1}$$

The extent to which the density $P(x)$ may be determined from its moments has been extensively discussed in the mathematical literature. In practice, only a finite number of moments, say $N + 1$, is usually available. Clearly then there exists an infinite variety of functions whose first $N + 1$ moments coincide and a unique reconstruction of $P(x)$ is impossible. Nevertheless, various approximation procedures exist which aim at constructing specific sequences of functions $P_N(x)$, such that

$$\int_a^b x^n P_N(x)dx = \mu_n, \quad n = 0,1,\ldots,N, \tag{1.2}$$

which eventually converge to the true distribution $P(x)$ as N approaches infinity. It is often mathematically expedient, and physically useful, to abandon the requirement of pointwise convergence and, instead, require weaker convergence for averages of the form

$$\langle F \rangle = \int_a^b F(x)P(x)dx = \lim_{N\to\infty} \int_a^b F(x)P_N(x)dx, \tag{1.3}$$

where $F(x)$ is some known function of physical interest.

The maximum-entropy approach offers a definite procedure for the construction of a sequence of approximations. The positive density $P(x)$ is interpreted as a probability density and the corresponding entropy is maximized under the condition that the first $N + 1$ moments be equal to the true moments $\mu_n, n = 0,1,\ldots, N$. Introducing appropriate Lagrange multipliers, one seeks maximization of the entropy functional $S = S(P)$ defined from

$$S = -\int_a^b [P(x)\ln P(x) - P(x)]dx$$
$$+ \sum_{n=0}^N \lambda_n \left(\int_a^b x^n P(x)dx - \mu_n \right). \tag{1.4}$$

Notice that we have incorporated in the definition of the entropy a term linear in $P(x)$, mostly for notational convenience. The linear term may be absorbed by a trivial redefinition of the Lagrange multiplier λ_0 in Eq. (1.4). Returning to the main point, the maxima $P = P_N(x)$ of (1.4) for $N = 1,2,\ldots$ will be taken as a sequence of approximations for the true density $P(x)$. It is customary to say that the maximum-entropy sequence $P_N(x)$ is the least-biased sequence of approximations.

The mathematical problem posed in the preceding paragraph was already considered in standard works on maximum entropy and concrete applications were also worked out in certain cases.[3,5,6] Nonetheless, recent reviews of a wealth of moment problems in quantum physics[7] do not even acknowledge possible use of the maximum-entropy approach. This situation is understandable because the more popular methods, such as polynomial expansions, Padé approximants, and the like, have had the advantage of extensive mathematical scrutiny over a period of a century or so. It is clear that a similar status for maximum entropy could be achieved only by an equally thorough study of its mathematial basis, by widening the scope of concrete applications, and by explicit comparison with the best approximation procedures currently in use.

For comparison purposes, it seems appropriate to brief-

ly outline here some of the better known methods for approximate solutions of the moment problem. A simple possibility is to expand $P(x)$ in some set of orthogonal polynomials. The resulting series is truncated after $N + 1$ terms and the expansion coefficients are determined by requiring that the first $N + 1$ moments be correct. This entails the solution of a $(N + 1) \times (N + 1)$ system of linear equations. Judicious choices of weighted orthogonal polynomials could lead to rapidly convergent sequences. In practice, the choice of a suitable weight is usually difficult, so the resulting sequences often produce notoriously oscillating approximations to $P(x)$ which are further impaired by lack of positivity at each finite stage of iteration.

Alternative, usually more powerful, procedures have been developed over the years, most of which are classified under the generic name of Padé approximations.[8] For instance, one may attempt to approximate the positive function $P(x)$ by finite sums of δ-functions of the form

$$P_N(x) = \sum_{i=1}^{(N+1)/2} m_i \delta(x - x_i), \tag{1.5a}$$

when N is odd, and

$$P_N(x) = \sum_{i=0}^{N/2} m_i \delta(x - x_i), \quad x_0 \equiv a, \tag{1.5b}$$

when N is even. In a language preferred by mathematicians, one writes $P(x)dx = d\mu(x)$, where the nondecreasing measure $\mu(x)$ is approximated by multistep functions. The parameters m_i and x_i in (1.5) are again determined by the requirement that the first $N + 1$ moments be correct:

$$\sum_i m_i x_i^n = \mu_n, \quad n = 0,1,..., N. \tag{1.6}$$

These are nonlinear equations but their solution may be reduced to the diagonalization of a tridiagonal Jacobi matrix.[9,10] The corresponding numerical procedure is apparently very stable and is often quoted in the literature as the Lanczos algorithm.[11] While the preceding method does not directly address a pointwise construction of $P(x)$, it is well suited for the computation of averages of the form (1.3) for which approximations may be obtained from

$$\langle F \rangle_N = \sum_i m_i F(x_i). \tag{1.7}$$

For the special case where $F(x) = 1/(1 + zx)$, Eq. (1.7) is but the standard Padé approximant associated with Stieljes integrals of the form

$$\langle F \rangle = \int_a^b \frac{P(x)dx}{1+zx}, \quad \langle F \rangle_N = \sum_i \frac{m_i}{1+zx_i}. \tag{1.8}$$

The distinction between even and odd N implicit in Eq. (1.5) results in two independent sequences of approximation which are the familiar diagonal and off-diagonal Padé sequences.

A number of questions raised in the preceding general introduction will be addressed in the following to varying degrees of completeness. In Sec. II, we briefly review well-known facts about maximum entropy and present some new mathematical results. In important special cases, we are able to derive necessary and sufficient conditions for the existence of a maximum-entropy solution and to some extent study the convergence of the resulting sequence. The numerical procedure and some elementary examples are also discussed in Sec. II. More interesting applications are described in the remainder of the paper. A detailed calculation of the density of states for a harmonic face-centered-cubic (fcc) crystal is presented in Sec. III and the results are compared with the earlier work of Gordon and Wheeler[10] using the Padé-like procedure outlined above; Sec. IV presents a similar calculation for dynamic correlation functions in some typical quantum spin systems. Further applications are contemplated in Sec. V and are illustrated by some simple exercises in the context of the Lee–Yang theory for Ising models. The same section contains a number of concluding remarks and some suggestions for possible generalizations.

II. FORMULATION AND ELEMENTARY EXAMPLES

The starting point for our discussion is the entropy defined by Eq. (1.4) for which we seek a maximum. Functional variation with respect to the unknown density $P(x)$ yields

$$\frac{\delta S}{\delta P(x)} = 0 \Rightarrow P = P_N(x) = \exp\left(-\lambda_0 - \sum_{n=1}^{N} \lambda_n x^n \right), \tag{2.1}$$

to be supplemented by the condition that the first $N + 1$ moments be given by μ_n:

$$\int_a^b x^n P_N(x)dx = \mu_n, \quad n = 0,1,..., N. \tag{2.2}$$

Equations (2.2) should be viewed as a (nonlinear) system of $N + 1$ equations for the $N + 1$ unknown Lagrange multipliers $\lambda_0, \lambda_1,..., \lambda_N$. Without loss of generality, we may assume in the following that the density $P(x)$ is normalized such that $\mu_0 = 1$. The first equation $(n = 0)$ in (2.2) then reads

$$\int_a^b P_N(x)dx = \int_a^b dx \exp\left(-\lambda_0 - \sum_{n=1}^{N} \lambda_n x^n \right) = 1, \tag{2.3}$$

and may be used to express λ_0 in terms of the remaining Lagrange multipliers:

$$e^{\lambda_0} = \int_a^b dx \exp\left(-\sum_{n=1}^{N} \lambda_n x^n \right) \equiv Z. \tag{2.4}$$

The system of equations (2.2) reduces to

$$\langle x^n \rangle = \mu_n, \quad n = 1,2,..., N,$$
$$\langle x^n \rangle \equiv \frac{\int_a^b dx \, x^n \exp\left(-\Sigma_{n=1}^{N} \lambda_n x^n \right)}{\int_a^b dx \, \exp\left(-\Sigma_{n=1}^{N} \lambda_n x^n \right)}. \tag{2.5}$$

An analytical solution of (2.5) is, of course, impossible except for the simple case $N = 1$. For numerical as well as theoretical purposes, one introduces a potential $\Gamma = \Gamma(\lambda_1, \lambda_2,..., \lambda_N)$ through the Legendre transformation[5]

$$\Gamma = \ln Z + \sum_{n=1}^{N} \mu_n \lambda_n, \tag{2.6}$$

there the μ_n's are the actual numerical values of the known moments. Stationary points of the potential Γ are solutions to the equations

$$\frac{\partial \Gamma}{\partial \lambda_n} = 0 \Rightarrow \langle x^n \rangle = \mu_n, \quad n = 1,2,..., N, \tag{2.7}$$

L. R. Mead and N. Papanicolaou   2405

which are precisely Eqs. (2.5). The first important property of Γ is summarized in the following lemma.

Lemma 1: The potential $\Gamma = \Gamma(\lambda_1,\lambda_2,..., \lambda_N)$ is everywhere convex. The proof of this result is already given in the literature[5] and proceeds by explicit construction of the Hessian

$$H_{nm} = \frac{\partial^2 \Gamma}{\partial \lambda_n \partial \lambda_m} = \langle x^{n+m} \rangle - \langle x^n \rangle \langle x^m \rangle, \qquad (2.8)$$

which may be proven to be positive definite for any generic set of λ's, not necessarily satisfying Eqs. (2.5). A more direct demonstration obtains by treating all Lagrange multipliers, including λ_0, on a common basis. Thus we introduce the potential $\Delta = \Delta(\lambda_0,\lambda_1,..., \lambda_N)$ from

$$\Delta = \int_a^b \left[\exp\left(-\lambda_0 - \sum_{n=1}^N \lambda_n x^n \right) - 1 \right] dx + \sum_{n=0}^N \mu_n \lambda_n, (2.9)$$

whose stationary points are given by

$$\frac{\partial \Delta}{\partial \lambda_n} = 0 \Rightarrow \langle x^n \rangle = \mu_n, \quad n = 0,1,..., N, \qquad (2.10)$$

which recombine Eqs. (2.5) with (2.4). Had we eliminated λ_0 using (2.4), the first term in (2.9) would vanish and the remaining terms would give (with $\mu_0 = 1$)

$$\Delta = \sum_{n=0}^N \mu_n \lambda_n = \mu_0 \lambda_0 + \sum_{n=1}^N \mu_n \lambda_n = \ln Z + \sum_{n=1}^N \mu_n \lambda_n, \qquad (2.11)$$

which is the potential Γ introduced earlier. However, one may directly work with $\Delta = \Delta(\lambda_0,\lambda_1,..., \lambda_N)$ which also satisfies Lemma 1. The corresponding Hessian reads

$$\Theta_{nm} = \frac{\partial^2 \Delta}{\partial \lambda_n \partial \lambda_m} = \int_a^b dx\, x^{n+m} \exp\left(-\sum_{n=0}^N \lambda_n x^n \right)$$

$$\equiv \langle x^{n+m} \rangle, \quad n,m = 0,1,.., \qquad (2.12)$$

and its positive definiteness is trivially established noting that

$$\int_a^b dx(u_0 + u_1 x + \cdots + u_k x^k)^2 \exp\left(-\sum_{n=0}^N \lambda_n x^n \right) \geqslant 0, \qquad (2.13)$$

for any nonnegative integer k and for any real $u_0, u_1,..., u_k$. Equation (2.13) may be rewritten as

$$\sum_{n,m=0}^k \langle x^{n+m} \rangle u_n u_m = \sum_{n,m=0}^k \Theta_{nm} u_n u_m \geqslant 0, \qquad (2.14)$$

which establishes that the Hessian Θ_{nm} is positive definite.

In practice, the potentials Γ or Δ may be used with comparable efficiency. We shall therefore proceed using the potential Γ. However, the potential Δ will prove more flexible for some generalizations discussed in Sec. V.

The convexity of Γ guarantees that if a stationary point is found for some finite values of $\lambda_1,\lambda_2,..., \lambda_N$, it must be a unique absolute minimum. Notice, however, that convexity alone does not imply that such a minimum should exist. This is not surprising because the convexity of Γ was established without any reference to the specific properties of the actual moments μ_n. A simple illustration may be given taking $N = 1$ and $[a,b] = [0,1]$, so that

$$Z = \int_0^1 dx\, e^{-\lambda_1 x} = \frac{1 - e^{-\lambda_1}}{\lambda_1},$$

$$\Gamma = \ln[(1 - e^{-\lambda_1})/\lambda_1] + \mu_1 \lambda_1. \qquad (2.15)$$

It is not difficult to see that the convex function $\Gamma = \Gamma(\lambda_1)$ possesses a minimum at some finite λ_1 only if $\mu_1 < 1 \ (= \mu_0)$. It is clear that this is the first of a set of conditions that the actual moments must satisfy in order to guarantee a minimum for $\Gamma = \Gamma(\lambda_1,...,\lambda_N)$. What are those conditions?

In order to answer the above question, it is pertinent at this point to review the general conditions under which the full moment problem shall have a solution, independently of the method of approximation. We restrict our discussion to the moment problem defined over a finite interval, say [0,1], which is the so-called Hausdorff moment problem.[12] Let $P(x)$ be a nonnegative density integrable in [0,1] and let $\{\mu_n, n = 0,1,2,...\}$ be the associated moment sequence. Noting that

$$\int_0^1 x^n(1 - x)^k P(x) dx > 0, \qquad (2.16)$$

and working out the integrand using the binomial expansion, the left side of (2.16) may be expressed in terms of the moments μ_n:

$$\Delta^k \mu_n \equiv \sum_{m=0}^k (-1)^m \binom{k}{m} \mu_{n+m} > 0, \quad n,k = 0,1,2,.... \qquad (2.17)$$

It is evident that the set of inequalities (2.17) is a set of necessary conditions for the moment sequence. Such a sequence is called completely monotonic. More importantly, Hausdorff established the sufficiency of the above conditions. Namely, given a completely monotonic moment sequence, there exists a nonnegative density $P(x)$ integrable in [0,1] whose moments coincide with μ_n.

Applying (2.17) for $k = 1$ and $n = 0$ one finds that $\mu_1 < \mu_0$, which is the condition we found earlier so that the potential $\Gamma = \Gamma(\lambda_1)$ of Eq. (2.15) will have a minimum. It is tempting to presume that the general potential $\Gamma = \Gamma(\lambda_1,\lambda_2,...,\lambda_N)$ will have a minimum if and only if the full set of conditions (2.17) is satisfied. That this is indeed so is guaranteed by the following theorem.

Theorem 1: A necessary and sufficient condition that the potential Γ should have a unique absolute minimum at some finite $\lambda_1,\lambda_2,...,\lambda_N$ for any N is that the moment sequence $\{\mu_n, n = 0,1,2,...\}$ should be completely monotonic.

We first establish sufficiency which is obviously the most relevant aspect of Theorem 1 for practical applications. In view of the convexity, it is clear that the essence of the proof should evolve around the asymptotic behavior of Γ at large λ. Hence the Lagrange multipliers are written as

$$\lambda_n = \lambda \alpha_n, \quad \sum_{n=1}^N \alpha_n^2 = 1, \qquad (2.18)$$

where λ is positive and the α_n's are the familiar direction cosines. One then obtains

$$\Gamma = \ln Z + \sum_{n=1}^N \mu_n \lambda_n$$

$$= \ln \left[\int_0^1 dx \exp\left(-\lambda \sum_{n=1}^N \alpha_n x^n \right) \right] + \lambda \sum_{n=1}^N \mu_n \alpha_n. \qquad (2.19)$$

L. R. Mead and N. Papanicolaou   2406

Our aim is to study the behavior of Γ as $\lambda \to \infty$ for an arbitrary choice of the direction cosines. It is convenient to combine both terms in (2.19) and write

$$\Gamma = \ln J, \quad J = \int_0^1 dx \, e^{\lambda \Pi_N(x)},$$

$$\Pi_N(x) = \sum_{n=1}^N \alpha_n (\mu_n - x^n), \tag{2.20}$$

so our task reduces to the study of the asymptotic behavior of the Laplace integral $J = J(\lambda)$ at $\lambda \to \infty$. The general procedure is explained in standard textbooks[13] and the result depends on the behavior of the N th degree polynomial $\Pi_N(x)$ in [0,1]. The relevant property of $\Pi_N(x)$ for our current purposes is summarized in the following lemma.

Lemma 2: If $\{\mu_n, n = 0,1,2,...\}$ is a completely monotonic moment sequence, the N th-degree polynomial $\Pi_N(x) = \Sigma_{n=1}^N \alpha_n (\mu_n - x^n)$ is strictly positive in a nontrivial subinterval of [0,1], for arbitrary real $\alpha_1, \alpha_2,...,\alpha_N$ not all of which vanish.

The proof of the lemma proceeds by contradiction. Let us assume that $\Pi_N(x)$ is not positive anywhere in [0,1], i.e.,

$$\Pi_N(x) \leqslant 0, \quad x \in [0,1]. \tag{2.21}$$

The polynomial $\Pi_N(x)$ may not be identically equal to zero because not all of the coefficients $\alpha_1, \alpha_2,...,\alpha_N$ vanish. It is therefore evident that the polynomial $\Pi_N(x)$ may not vanish but at a finite number of points not exceeding its degree N. Furthermore, the general theory of Hausdorff guarantees that given a completely monotonic moment sequence there exists a nonnegative density $P(x)$ whose moments are $\mu_0 = 1$, $\mu_1, \mu_2,...$. While $P(x)$ may vanish over nontrivial subintervals of [0,1], it must also be strictly positive over some nontrivial regions in [0,1]. Hence, in view of (2.21) and the ensuing remarks, the product $\Pi_N(x)P(x) \leqslant 0$ may vanish over nontrivial regions but its integral over [0,1] is strictly negative:

$$\int_0^1 \Pi_N(x)P(x)dx < 0. \tag{2.22}$$

Some implicit smoothness assumptions about $P(x)$ are inherent in the above argument; $P(x)$ cannot be a δ-function, for instance. On introducing the explicit expression for the polynomial $\Pi_N(x)$ in (2.22), one finds that

$$\sum_{n=1}^N \alpha_n \int_0^1 (\mu_n - x^n)P(x)dx < 0. \tag{2.23}$$

Recalling that $\int_0^1 P(x)dx = \mu_0 = 1$ and $\int_0^1 x^n P(x)dx = \mu_n$ for $n = 1,2,...$, the left side of (2.23) vanishes. We have thus reached a contradiction implying that (2.21) cannot be true over the entire interval [0,1]. Hence there exist nontrivial regions in [0,1] where $\Pi_N(x)$ is strictly positive, establishing the validity of Lemma 2.

The conditions of Lemma 2 are valid for the polynomial $\Pi_N(x)$ defined in Eq. (2.20) because not all of the direction cosines $\alpha_1, \alpha_2,...,\alpha_N$ may vanish simultaneously in view of the normalization constraint (2.18). Let x_0 be the point where $\Pi_N(x)$ achieves its maximum which is positive:

$$\max_{x \in [0,1]} \Pi_N(x) = \Pi_N(x_0) > 0. \tag{2.24}$$

The behavior of $J(\lambda)$ at $\lambda \to \infty$ is governed by the behavior of

$\Pi_N(x)$ in the neighborhood of x_0. There are several cases to consider depending on whether x_0 lies at one of the endpoints or not, and whether the corresponding derivatives $\Pi_N'(x_0), \Pi_N''(x_0),...$ vanish or not. A complete analysis of the various cases may be found in Ref. 13, which we will not repeat here. The general result is that $J(\lambda)$ grows exponentially as $\lambda \to \infty$ for an arbitrary choice of the direction cosines. Hence the potential $\Gamma = \ln J$ grows linearly with λ in all directions. A convex function with the above asymptotic behavior must possess a unique absolute minimum at some finite $\lambda_1, \lambda_2,...,\lambda_N$, which establishes sufficiency in Theorem 1.

The necessity of the conditions stated in Theorem 1 does not have a direct bearing on the practical aspects of this problem, but we briefly sketch the proof of its validity for the sake of completeness. Let us assume that the sequence of real numbers $\{\mu_n, n = 0,1,2,...\}$ is such that the potential Γ possess a minimum at some finite $\lambda_1, \lambda_2,...,\lambda_N$. We are to prove that the sequence $\{\mu_n\}$ must be completely monotonic. Recall that Γ is convex everywhere for arbitrary values of μ_n. Since Γ possesses a minimum, by our hypothesis, the minimum is unique and absolute. Therefore, moving away from the minimum in any direction should result in monotonically increasing values for Γ. This behavior is compatible only with a polynomial $\Pi_N(x)$ in Eq. (2.20) that achieves a positive maximum at some point x_0 in [0,1] for any value of the direction cosines. We write

$$\phi_N(x) \equiv \lambda \Pi_N(x) = \sum_{s=1}^N \lambda \alpha_s (\mu_s - x^s), \quad \phi_N(x_0) > 0, \tag{2.25}$$

for any real $\lambda \alpha_1, \lambda \alpha_2,...,\lambda \alpha_N$ not all of which vanish. In particular, choose

$$\lambda \alpha_s = \begin{cases} 0, & 1 \leqslant s < n, \\ (-1)^{s-n} \binom{k}{s-n}, & n \leqslant s \leqslant n+k, \\ 0, & n+k < s \leqslant N, \end{cases} \tag{2.26}$$

so that, using the notation of Eq. (2.17),

$$\phi_N(x) = \Delta^k \mu_n - x^n (1-x)^k. \tag{2.27}$$

The only stationary point of $\phi_N(x)$ in [0,1] is a local minimum at the interior point $x = n/(n+k)$. Therefore the maximum of $\phi_N(x)$ occurs at one of the endpoints $(x_0 = 0$ or 1) where the second term in (2.27) vanishes, and $\phi_N(x_0) > 0$ implies that $\Delta^k \mu_n > 0$. The sequence $\{\mu_n, n = 0,1,2,...\}$ is thus completely monotonic.

To summarize, it is gratifying that the conditions for the existence of a maximum-entropy solution are identical to Hausdorff's conditions addressing the full moment problem. Given a completely monotonic moment sequence, Theorem 1 guarantees the existence of a maximum-entropy solution $P_N(x)$ for any N, however large. The solution $P_N(x)$ is nonnegative and integrable (in fact, absolutely continuous) in [0,1] and satisfies the moment conditions

$$\int_0^1 x^n P_N(x)dx = \mu_n, \quad n = 0,1,...,N. \tag{2.28}$$

A sequence of functions $P_N(x)$, $N = 1,2,...$ with the above general properties converges in the sense of the following theorem.

Theorem 2: Let $P(x)$ be a nonnegative function integrable in $[0,1]$ whose moments are $\mu_0, \mu_1, \ldots$, and let $P_N(x)$, $N = 1,2,\ldots$ be the maximum-entropy sequence associated with the same moments. If $F(x)$ is some continuous function in $[0,1]$ then

$$\lim_{N \to \infty} \int_0^1 F(x) P_N(x) dx = \int_0^1 F(x) P(x) dx. \qquad (2.29)$$

The proof of the above theorem can be obtained by putting together some standard results of analysis which may be found in the book of Widder[12] and are freely used in the following. Consider the sequence of functions

$$\psi_N(x) = \int_0^x [P(t) - P_N(t)] dt, \quad N = 1,2,\ldots, \qquad (2.30)$$

each member of which is a function of bounded variation because both $P(x)$ and $P_N(x)$ are nonnegative. The variation of $\psi_N(x)$ is given by

$$V[\psi_N(x)]_0^x = \int_0^x [P(t) + P_N(t)] dt$$

$$\leqslant \int_0^1 [P(t) + P_N(t)] dt = 2. \qquad (2.31)$$

Had we maintained arbitrary normalization for $P(x)$ and $P_N(x)$, the right side of (2.31) would read $2\mu_0$. In all cases, the right side of (2.31) is N-independent. Therefore the sequence (2.30) is of uniform bounded variation. Hence there exists a subsequence $\{\psi_{N_j}(x)\}$, and a function of bounded variation $\psi(x)$, such that

$$\lim_{j \to \infty} \psi_{N_j}(x) = \psi(x). \qquad (2.32)$$

It follows from (2.28) and (2.30) that

$$\int_0^1 x^n d\psi_N(x) = 0, \quad n = 0,1,\ldots,N, \qquad (2.33)$$

which combines with (2.32) to yield

$$\int_0^1 x^n d\psi(x) = 0, \quad n = 0,1,\ldots,\infty. \qquad (2.34)$$

This and the uniqueness theorem for moment sequences associated with functions of bounded variation gives

$$\psi(x) = 0, \qquad (2.35)$$

for every x in $[0,1]$. Furthermore, every subsequence of (2.30) has in it a subsequence that converges to a function of bounded variation, and all convergent subsequences may be shown to converge to the same limit $\psi(x) = 0$ by iterating the uniqueness theorem. Therefore the original sequence (2.30) also converges and

$$\lim_{N \to \infty} \psi_N(x) = \lim_{N \to \infty} \int_0^x [P(t) - P_N(t)] dt = 0, \qquad (2.36)$$

for every x in $[0,1]$. This result implies that Eq. (2.29) holds for every continuous function $F(x)$.

The weak convergence established by Theorem 2 was obtained using only general properties of the maximum-entropy sequence, notably, the positivity of each approximant $P_N(x)$. In principle, it may prove possible to establish stronger forms of convergence by incorporating the finer details of the actual construction and by imposing further constraints

on the original density $P(x)$. Rigorous results on the latter subject are not available at this point, so our subsequent discussion will be based on extensive empirical evidence. Similarly, the maximum-entropy solution for moment problems on an infinite interval has not been understood to any degree of mathematical rigor. To indicate some of the complications, we briefly consider here the two-moment solution for a semi-infinite interval, which is obtained by minimizing the potential

$$\Gamma = \ln Z + \mu_1 \lambda_1 + \mu_2 \lambda_2,$$

$$Z = \int_0^\infty dx\, e^{-\lambda_1 x - \lambda_2 x^2}. \qquad (2.37)$$

The integral in (2.37) is expressed in terms of the usual error function and its asymptotic behavior may be studied explicitly. It turns out that a minimum exists only if

$$\mu_1^2 < \mu_2 < 2\mu_1^2, \quad \mu_0 = 1. \qquad (2.38)$$

In practice, violation of the second inequality in (2.38) is the rule rather than the exception. Nevertheless numerical studies indicate that a solution always exists if the moments are incorporated in odd numbers. Explicit examples will be discussed later in the text. We would like to add here that conditions analogous to (2.38) were also discussed in a recent preprint[14] which was brought to our attention during the preparation of this work.

A. Numerical procedure

The method for numerical computations is clearly suggested by the convexity of the potential Γ. One may use the familiar Newton minimization procedure. We briefly review the procedure in order to indicate possible differences between our numerical work and earlier calculations.[5] Starting with some initial choices for $\lambda_1, \lambda_2, \ldots, \lambda_N$, updated λ's are defined from

$$\lambda'_n = \lambda_n - \sum_{m=1}^N (H^{-1})_{nm} \frac{\partial \Gamma}{\partial \lambda_m}$$

$$= \lambda_n - \sum_{m=1}^N (H^{-1})_{nm} [\mu_m - \langle x^m \rangle], \qquad (2.39)$$

where $H = (H_{nm})$ is the Hessian matrix defined in Eq. (2.8) and H^{-1} is its inverse. All quantities on the right side of (2.39) are calculated at the input values of λ. This entails the evaluation of $2N$ expected values of the form $\langle x^k \rangle$, $k = 1,2,\ldots, 2N$, and the solution of a $N \times N$ system of linear equations. Write $\lambda'_n = \lambda_n - a_n$, so that the Newton shift a_n satisfies the linear system

$$\sum_{m=1}^N H_{nm} a_m = \mu_n - \langle x^n \rangle, \quad n = 1,2,\ldots,N. \qquad (2.40)$$

Because H is positive definite, the linear system (2.40) was solved with the standard IMSL routine LEQT2P (used here with double-precision arithmetic).

The demanded accuracy for the evaluation of the integrals involved in $\langle x^k \rangle$, see Eq. (2.5), was better than 12 significant figures. Two independent routines were used for the evaluation of the integrals thus obtaining an important

check of consistency. We first describe the calculation for the finite interval [0,1]. We found that four double-precision 24-point Gaussian quadratures evenly distributed over [0,1] very efficiently produce the demanded accuracy for $\langle x^k \rangle$. An independent check was performed with an adaptive Newton–Cotes integration algorithm. While the results of the Gaussian quadrature were always confirmed, the Newton–Cotes algorithm proved slower. However, there were two instances in which the Newton–Cotes algorithm proved indispensable. First, while the Gaussian quadrature is very accurate for the evaluation of integrals involving polynomials as well as exponentials of polynomials, its accuracy is impaired somewhat in the calculation of averages of the form $\langle F \rangle$, where $F = F(x)$ is some function containing square roots, integrable singularities, etc. Second, Gaussian quadratures proved inadequate for calculations on an infinite interval in view of the high degree of precision requested in our calculations. On the contrary, the adaptive Newton–Cotes algorithm could be adjusted to handle such complications, as is described in more detail in Sec. IV.

For finite-interval calculations, the updating procedure (2.39) may be initialized setting all λ's equal to zero, and using the previously computed λ's as input for all higher iterations. Two criteria were used to stop the procedure. First, the updated moments $\langle x^n \rangle, n = 1,2,...,N$ were requested to agree with the actual moments to an accuracy of one part in 10^{12} or better. Second, the updated values for $\lambda_1, \lambda_2,...,\lambda_N$ were stabilized to at least one part in 10^8. It is an empirical fact that averages such as $\langle x^n \rangle$ were consistently reproduced with higher relative accuracy than the corresponding accuracy for individual λ's. All calculations were performed interactively on an IBM 360/70 system. Five to six Newton iterations were typically sufficient. Under the precision standards set above, we were able to handle $10 \sim 12$ moments without difficulty.

Incorporating higher moments introduces instabilities in the algorithm reflected in failures of the LEQT2P routine, apparently due to accumulation of roundoff error. For most practical calculations described in the following we did not find it necessary to go beyond 12 moments. We include, however, some comments concerning possible procedures to remedy the numerical difficulties with higher moments. An obvious possibility is to increase the precision requirements at the cost of considerably slowing down the algorithm. Other remedies are (i) initiate the algorithm with a simple gradient method, or with a gradient and/or Newton algorithm with step adjustment, and (ii) introduce a line-search routine of the type discussed in Ref. 5. We have in fact built in our procedure all of the above options. As far as calculations involving up to $10 \sim 12$ moments are concerned, those options are a hindrance rather than an improvement. Nevertheless, they may prove valuable for higher-moment calculations and for calculations over an infinite interval.

To conclude this discussion, we mention that the preceding algorithm may be trivially adjusted to calculate the minimum of the potential Δ introduced in Eq. (2.9). The Hessian H_{nm} in Eq. (2.39) is replaced by $\Theta_{nm} = \langle x^{n+m} \rangle$ and the summation is extended to include $n,m = 0$. Our numerical results were thus reproduced with equal efficiency.

B. Elementary examples

Consider the elementary density

$$P(x) = \alpha + 2(1 - \alpha)x, \qquad (2.41)$$

where the parameter α varies in $0 \leqslant \alpha \leqslant 1$, so that $P(x)$ is positive in [0,1]. The corresponding moments are given by

$$\mu_n = \alpha/(n + 1) + 2(1 - \alpha)/(n + 2). \qquad (2.42)$$

It is not clear whether or not the maximum-entropy sequence will produce a good approximation for (2.41), because one attempts to approximate a simple monomial by an exponential of a polynomial. For $\alpha = 1$, $P(x)$ reduces to a constant and the correct result is, of course, obtained after a single iteration. The results of a pointwise comparison with the true density for $\alpha = \frac{1}{2}$ and $\alpha = 0$ are summarized in Table I using eight moments as input. (The value of N, instead of $N + 1$, will hereafter be referred to as the number of input moments.)

It is evident that the pointwise fit is excellent for $\alpha = \frac{1}{2}$, while the accuracy diminishes for $\alpha = 0$ apparently because $P = 2x$ contains a zero. Much more stable is, however, the convergence of averages of the form (2.29) even for $P = 2x$. Choosing

$$F(x) = \frac{1}{2}\sqrt{x}, \qquad (2.43)$$

the corresponding average $\langle F \rangle = U_0$ will be called the zero-temperature internal energy for reasons explained in Sec. III. The sequence of approximations to U_0 is given in Table II for $P = 2x$ and is shown to accurately reproduce the exact answer. Such behavior is in fact typical for densities $P(x)$ with sophisticated structure.

We have further studied the behavior of individual λ's. Typically, they alternate in sign while their absolute value increases substantially with increasing N. We found it instructive to rearrange the polynomial

$$- \ln P_N(x) = \sum_{n=0}^{N} \lambda_n(N)x^n \qquad (2.44)$$

as a linear superposition of the Legendre polynomials defined in [0,1],

$$L_n(x) = \frac{(-1)^n}{n!} \frac{d^n}{dx^n}[x^n(1 - x)^n], \qquad (2.45)$$

and write

$$\sum_{n=0}^{N} \lambda_n(N)x^n = \sum_{n=0}^{N} \nu_n(N)L_n(x). \qquad (2.46)$$

It is not difficult to see that

$$\nu_n(N)$$

$$= (2n + 1) \sum_{m=n}^{N} \frac{\Gamma^2(m + 1)}{\Gamma(m - n + 1)\Gamma(m + n + 2)} \lambda_m(N). \qquad (2.47)$$

We found through numerical experimentation that the $\nu_n(N)$'s approach the ordinary Legendre coefficients of $- \ln P(x)$ when N increases. It is, of course, incorrect to presume that Eq. (2.46) for finite N is directly related to the Legendre expansion of $- \ln P(x)$.

The preceding comparison relies on the assumption

that both $P(x)$ and $-\ln P(x)$ are integrable in $[0,1]$. We were thus prompted to consider the density

$$P(x) = \mu_0^{-1} e^{-1/x},$$

$$\mu_0 = \int_0^1 dt\, e^{-1/t} = 0.148\,495\,51, \tag{2.48}$$

whose logarithm is not integrable. Nevertheless, the pointwise fit of the corresponding maximum-entropy solution presented in Table I, as well as the average $\langle \sqrt{x}/2 \rangle$ given in Table II, is in good agreement with the exact answer. This may not be surprising because functions with nonintegrable singularities, such as $-\ln P(x) \sim 1/x$, may also be approximated by polynomials closely related to the Bernstein polynomials.[15] The preceding remarks should, however, make one cautious in comparing the maximum-entropy sequence with standard polynomial expansions.

We conclude this section by mentioning an amusing calculation. Setting $\alpha = -1$ in Eqs. (2.41) and (2.42), the resulting distribution is not positive everywhere in $[0,1]$. The moments

$$\mu_n = -1/(n+1) + 4/(n+2) \tag{2.49}$$

are still positive and monotonically decreasing. However, the moment sequence (2.49) is not completely monotonic. For instance, the Hausdorff inequality (2.17) is violated for $n = 0$ and $k = 2$. According to Theorem 1, the potential Γ cannot have a finite minimum in this case, a fact that was easily demonstrated numerically feeding (2.49) into our algorithm.

III. DENSITY OF STATES IN HARMONIC SOLIDS

We now turn our attention to some more sophisticated moment problems that arise in physical applications. The first example will be the discussion of the density of states in a harmonic solid. Upon the assumption that anharmonic forces are relatively insignificant, the problem of calculating various quantities of physical interest reduces to the diagonalization of a matrix. In most cases, however, an analytical diagonalization is impossible and, even today, direct numerical procedures for 3-D lattices are difficult to implement with sufficient accuracy. On the contrary, indirect moment methods have long been known to provide very valuable substitutes for exact solutions.

TABLE I. Maximum-entropy solution using eight moments and comparison with the exact density for $P = x + \frac{1}{2}$, $P = 2x$, and $P = \mu_0^{-1} e^{-1/x}$.

| | $P = x + \frac{1}{2}$ | | $P = 2x$ | | $P = \mu_0^{-1} e^{-1/x}$ | |
x	Maxent	Exact	Maxent	Exact	Maxent	Exact
0.0	0.500 003 5	0.5	0.0172	0.0	10^{-9}	0.0000
0.1	0.600 000 6	0.6	0.2032	0.2	0.0004	0.0003
0.2	0.699 999 9	0.7	0.3998	0.4	0.0450	0.0454
0.3	0.799 999 3	0.8	0.5946	0.6	0.2409	0.2402
0.4	0.900 000 9	0.9	0.8062	0.8	0.5513	0.5528
0.5	1.000 000 0	1.0	1.0005	1.0	0.9131	0.9114
0.6	1.099 999 1	1.1	1.1934	1.2	1.2711	1.2719
0.7	1.200 000 7	1.2	1.4047	1.4	1.6130	1.6139
0.8	1.300 000 2	1.3	1.6015	1.6	1.9314	1.9294
0.9	1.399 999 3	1.4	1.7949	1.8	2.2143	2.2168
1.0	1.499 996 3	1.5	1.9736	2.0	2.4689	2.4774

TABLE II. Maximum-entropy sequence of approximations of the zero-temperature internal energy, for the model densities $P = 2x$ and $P = \mu_0^{-1} e^{-1/x}$, and for a harmonic fcc crystal.

| Number of moments | Internal energy U_0 | | |
	$P = 2x$	$P = \mu_0^{-1} e^{-1/x}$	fcc crystal
1	0.396 913 3	0.422 396 9	0.333 333 3
2	0.399 500 9	0.425 637 5	0.339 133 9
3	0.399 864 8	0.425 870 2	0.340 615 7
4	0.399 951 8	0.425 894 1	0.341 081 2
5	0.399 979 4	0.425 897 2	0.340 775 9
6	0.399 990 0	0.425 897 7	0.340 867 6
7	0.399 994 7	0.425 897 8	0.340 879 6
8	0.399 997 0	0.425 897 8	0.340 865 3
9	0.399 998 2		0.340 884 5
10	0.399 998 8		0.340 886 4
Accurate	0.400 000 0	0.425 897 8	0.340 887 2

The quantity of prime interest is the (positive) density of states $g = g(\omega)$, where ω is the frequency varying in a finite interval $[0, \omega_0]$. The maximum frequency ω_0 is usually known and may be scaled out so that the interval becomes $[0,1]$. It is generally possible to explicitly determine a number of power moments of the form

$$\mu_n = \int_0^1 \omega^{2n} g(\omega) d\omega, \tag{3.1}$$

which involve computations of traces rather than eigenvalues of a given matrix. Only even moments appear, but the moment problem may be reduced to its standard form by introducing a new density $P = P(x)$ from

$$x = \omega^2, \quad g(\omega)d\omega = P(x)dx, \tag{3.2}$$

so that $g(\omega) = 2\omega P(\omega^2)$ and Eqs. (3.1) read

$$\mu_n = \int_0^1 x^n P(x)dx, \quad n = 0,1,2,\dots. \tag{3.3}$$

If the density of states were known, various thermodynamic quantities could be computed as suitable averages with respect to $P(x)$. For instance, the internal energy and specific heat are given by

$$U = \tau \int_0^1 \left[\frac{\sqrt{x}}{2\tau} \coth\left(\frac{\sqrt{x}}{2\tau} \right) \right] P(x)dx \tag{3.4}$$

and

$$C = \int_0^1 \left(\frac{\sqrt{x}/2\tau}{\sinh(\sqrt{x}/2\tau)} \right)^2 P(x)dx. \tag{3.5}$$

All units have been scaled out of (3.4) and (3.5) and τ stands for a suitably normalized temperature. It will prove useful to consider, in particular, a special case of (3.4), the zero-temperature limit of the internal energy:

$$U_0 = \frac{1}{2} \int_0^1 \sqrt{x} P(x)dx. \tag{3.6}$$

Notice that (3.4)–(3.6) are concrete examples of averages of the form considered in Eq. (2.29).

The specific problem we study here is that of a harmonic fcc crystal with nearest-neighbor interactions. This problem has attracted interest over a long period of time,[16] espe-

TABLE III. The Z and λ coefficients for the density of states in a harmonic fcc crystal (using ten moments), and for the Fourier transform of a two-point correlation function in the 1-D isotropic Heisenberg and XY models (using five moments).

	fcc crystal	Heisenberg	XY model
Z	$1.471\ 329\ 4 + 1$	$1.915\ 731\ 738 + 0$	$2.121\ 488\ 856 + 0$
λ_1	$-6.470\ 649\ 9 + 1$	$5.610\ 630\ 136 - 1$	$7.624\ 279\ 265 - 1$
λ_2	$9.945\ 635\ 3 + 2$	$-8.991\ 552\ 268 - 2$	$-2.871\ 904\ 750 - 1$
λ_3	$-8.642\ 201\ 9 + 3$	$6.070\ 910\ 692 - 3$	$4.830\ 147\ 309 - 2$
λ_4	$4.235\ 568\ 8 + 4$	$-1.619\ 301\ 504 - 4$	$-3.552\ 353\ 395 - 3$
λ_5	$-1.247\ 859\ 5 + 5$	$1.507\ 717\ 407 - 6$	$9.382\ 829\ 849 - 5$
λ_6	$2.287\ 607\ 1 + 5$		
λ_7	$-2.606\ 395\ 1 + 5$		
λ_8	$1.772\ 624\ 8 + 5$		
λ_9	$-6.474\ 692\ 4 + 4$		
λ_{10}	$9.504\ 657\ 1 + 3$		

cially after Isenberg[17] was able to explicitly compute a large number of moments (35 moments). A thorough analysis of the associated moment problem was later given by Wheeler and Gordon[10] using the Padé-like procedure outlined in the Introduction. We thought it appropriate to calculate the density of states and some thermodynamic averages by the maximum-entropy approach, so stringent comparisons can be made with independent powerful methods. The explicit values of the moments used here are known to high precision and may be found summarized in Ref. 10. Our usual convention $\mu_0 = 1$ is implicit in the following. Hence we consider the N-moment maximum-entropy approximation of the form

$$P_N(x) = \frac{1}{Z} \exp\left(-\sum_{n=1}^{N} \lambda_n x^n\right), \qquad (3.7)$$

for various values of N. The calculated values for $Z, \lambda_1, \lambda_2, ...$ with $N = 10$ are given in Table III. It is then straightforward to plot the corresponding approximation to the density of states, as is done in Fig. 1. The general features of the emerging curve are consistent with the known van-Hove critical points [where $P(x)$ has infinite derivative] which are located at

$$x = \tfrac{1}{4}, \tfrac{1}{2} \quad \text{and} \quad \tfrac{1}{4}(2 + \sqrt{2}) = 0.853... \,. \qquad (3.8)$$

The obtained curve is also consistent with the general form of $P(x)$ anticipated by root-sampling techniques.[10]

How good is the approximation furnished by Fig. 1? Since the exact solution is unknown, an indirect assessment can be made by examining the sequence of successive approximations for $N = 1$–10. We found that the gross features
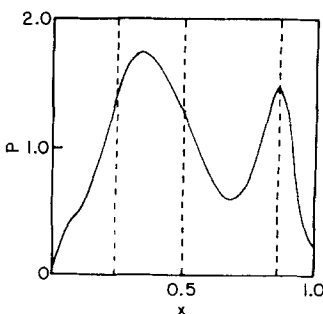
FIG. 1. Maximum-entropy calculation of the density of states for a harmonic fcc crystal using ten moments ($N = 10$). The vertical dashed lines indicate the location of the van-Hove critical points.

of the curve given in Fig. 1 have stabilized, but changes are still underway in the neighborhood of the critical points (especially at $x = 0.853$) and in the neighborhood of the endpoints. It should be mentioned here that direct comparison with the calculation of Wheeler and Gordon cannot be made at this level because they approximate the density by a sum of δ-functions. However, such comparison is possible for thermodynamic averages. Table II contains the maximum entropy sequence of approximations for the internal energy at zero temperature obtained by substituting the approximation (3.7) in Eq. (3.6). The result obtained here using ten moments is consistent with the upper and lower bounds derived by Wheeler and Gordon[10] using 30 moments:

$$U_0 = 0.340\ 88^{83}_{07}. \qquad (3.9)$$

This is an encouraging fact especially because the authors of that reference note that using 12 moments in their procedure yields a result accurate only to one part in 10^3. Some caution should be exercised, however, because it is not presently known whether incorporating higher moments in the maximum-entropy calculation will rapidly lead to significant improvements. We noted earlier that our numerical procedure is not yet capable of handling a large number of moments. From the practical point of view, it is interesting that we obtain good results for averages using a relatively small number of moments.

Further important tests are obtained by calculating the specific heat from Eq. (3.5) for various temperatures. It is evident that the coefficients of the high-temperature expansion of the specific heat may be directly expressed in terms of the moments of $P(x)$. On the other hand, the high-temperature expansion is known to diverge for $\tau < \tau_c = 1/2\pi = 0.159$. Hence the maximum-entropy sequence may be thought of as a method of summation of the high-temperature series outside its radius of convergence. It is to be expected that the same procedure will also yield a faster sequence of approximations in the region $\tau > \tau_c$. Our results are summarized in Table IV. A few iterations suffice to yield very accurate results for $\tau > \tau_c$. The sequence is also convergent in the region $\tau < \tau_c$, albeit at a slower rate. For very low temperatures the rate of convergence becomes poor because the integrand in Eq. (3.6) is concentrated over a small region around $x \simeq 0$. This situation could be remedied by incorporating the information from a short-frequency expansion of $P(x)$.[10]

To summarize, the maximum-entropy approach gives good results for thermodynamic averages even when a relatively small number of moments is included. An equally satisfactory pointwise fit of the density of states generally requires a larger number of moments. However, wild oscillations typical of polynomial expansions are less likely to occur in a maximum-entropy calculation. Some more demanding examples where gaps appear in the spectrum will be discussed in later sections.

IV. DYNAMIC CORRELATION FUNCTIONS IN QUANTUM SPIN SYSTEMS

Dynamic correlation functions provide the most direct tool for comparisons between theory and experiment in the study of quantum spin systems.[18] Nevertheless, a variety of

TABLE IV. Results for the specific heat in a harmonic fcc crystal at various temperatures.

Number of moments	Specific Heat			
	$\tau = 0.05$	0.1	0.5	1.0
3	0.011 899 47	0.100 879 49	0.852 041 352 2	0.959 599 573 6
4	0.007 687 81	0.097 249 15	0.852 040 548 4	0.959 599 569 6
5	0.011 754 13	0.098 997 75	0.852 040 575 8	0.959 599 569 6
6	0.010 289 76	0.098 685 65	0.852 040 575 5	0.959 599 569 6
7	0.010 086 38	0.098 661 53	0.852 040 575 5	0.959 599 569 6
8	0.010 317 67	0.098 676 25	0.852 040 575 5	0.959 599 569 6
9	0.010 036 05	0.098 666 40	0.852 040 575 5	0.959 599 569 6
10	0.010 011 97	0.098 665 93	0.852 040 575 5	0.959 599 569 6

theoretical methods that are suitable for the calculation of a wide range of static properties (ground state, spectrum of elementary and collective excitations, thermodynamic averages) often prove inadequate for detailed predictions of dynamical properties. Even for 1-D Heisenberg models where powerful Bethe-ansatz techniques apply and yield exact results for static properties, the computation of time-dependent correlation functions has proved difficult. The best known exception is the XY model for which the two-point longitudinal function is known for all temperatures.[19,20] Some extensions to more complicated cases have also become possible through the continuing effort of a number of authors.[19–25]

Needless to say, the preceding remarks apply also to various semiclassical methods whose limitations for calculations of dynamical properties were often emphasized in the literature.[26] On the other hand, it is generally agreed that indirect moment methods can provide a very valuable tool for the calculation of dynamic correlations. To illustrate the ideas, let us consider a simple 1-D anisotropic Heisenberg model described by the Hamiltonian

$$H = -\sum_{n,\delta} [J_\perp S_n^- S_{n+\delta}^+ + J_\parallel S_n^z S_{n+\delta}^z], \qquad (4.1)$$

where standard notation has been used. Setting $J_\parallel = 0$ in (4.1) yields the XY model, while the isotropic Heisenberg model corresponds to $J_\perp = J_\parallel \equiv J$. A typical correlation function is given by

$$\tfrac{1}{4} G_{n-m}(t) = \langle S_n^z(t) S_m^z(0) \rangle, \qquad (4.2)$$

where $\langle \cdots \rangle$ denotes the usual thermodynamical average, which reduces to the vacuum-expectation value at zero temperature. Translation invariance is reflected in the notation of Eq. (4.2) and the factor of $\tfrac{1}{4}$ was introduced for future convenience.

The difficulties in the explicit computation of the average of Eq. (4.2) become apparent on writing

$$S_n^z(t) = e^{iHt} S_n^z e^{-iHt}, \qquad (4.3)$$

where $S_n^z = S_n^z(0)$. However, the computation of the coefficients of a short-time expansion is feasible. An exact calculation of a modestly large number of coefficients is possible at infinite temperature, where the averages $\langle \cdots \rangle$ reduce to simple traces, and at zero temperature when the exact ground state is known.[27] For our illustrations, we will con-

sider the infinite-temperature limit for which $6(= 5 + 1)$ moments have been computed[28] in some important cases [for the spin-$\tfrac{1}{2}$ model described by the Hamiltonian (4.1), for instance]. The correlation function (4.2) is then symmetric under time reversal $(t \rightarrow -t)$, so the short-time expansion contains only even terms:

$$G_{n-m}(t) = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!} \mu_{n-m}^{(2k)} t^{2k}, \qquad (4.4)$$

where the moments $\mu_{n-m}^{(2k)}$ are static averages (traces) of the form

$$\tfrac{1}{4} \mu_{n-m}^{(0)} = \langle S_n^z S_m^z \rangle,$$
$$\tfrac{1}{4} \mu_{n-m}^{(2)} = \langle [H,[H,S_n^z]] S_m^z \rangle, \qquad (4.5)$$

and so on.

A further restriction in our calculation will be to consider the correlation function for zero-space separation $(n - m = 0)$. The Fourier transform defined from

$$G_0(t) = \int_0^\infty \cos(\omega t) P(\omega) d\omega \qquad (4.6)$$

yields a nonnegative density $P = P(\omega)$ whose power moments

$$\int_0^\infty \omega^{2n} P(\omega) d\omega = \mu_0^{(2n)} \equiv \mu_n, \qquad n = 0,1,2,... \qquad (4.7)$$

are equal to the perturbation coefficients appearing in (4.4) with $n - m = 0$. Explicit values for the first few moments in the spin-$\tfrac{1}{2}$ XY model $(J_\perp = 0, J_\parallel \equiv 1)$, and in the isotropic Heisenberg model $(J_\parallel = J_\perp \equiv 1)$, may be extracted from the work of Morita[28] and are summarized in Table V of our paper. While $G_0(t)$ for the Heisenberg model is not known, an exact solution for the XY model is available[20]:

$$G_0(t) = [J_0(2t)]^2, \qquad (4.8)$$

where J_0 is the familiar Bessel function. The Fourier transform reads

$$P(\omega) = \begin{cases} (1/\pi^2) K'(\omega/4), & \text{for } \omega < 4, \\ 0, & \text{for } \omega > 4, \end{cases} \qquad (4.9)$$

where $K'(k) = K[(1 - k^2)^{1/2}]$ and $K(k)$ is the complete elliptic integral of the first kind.

The preceding exact result for the XY model will provide important tests for our approximate calculations. Our task is to reconstruct the density $P(\omega)$ by a maximum-en-

TABLE V. Moments of a two-point function for the 1-D spin-$\frac{1}{2}$ XY and isotropic Heisenberg models at infinite temperature.

	XY model	Heisenberg
μ_0	1	1
μ_1	4	4
μ_2	36	44
μ_3	400	652
μ_4	4 900	11 636
μ_5	63 504	242 816

tropy calculation using as input the moments of Table V. The calculation will be carried out for both the XY and the isotropic Heisenberg model. We will further compute $G_0(t)$ from Eq. (4.6), which is an average of the general form considered earlier in Eq. (2.29). Hence the maximum-entropy calculation may be thought of as a resummation procedure for the perturbative series (4.4).

We would like to stress from the outset that the current examples furnish demanding tests of the efficiency of maximum-entropy techniques. As is suggested by the exact solution in the XY model, Eq. (4.9), the Fourier transform $P(\omega)$ typically vanishes outside a certain region whose extent is not known. We are thus forced to perform a semi-infinite-interval calculation and eventually predict the support along with other details of $P(\omega)$. Furthermore, $P(\omega)$ often develops (integrable) singularities in the infrared region $(\omega \simeq 0)$ whose precise nature is unknown.

By analogy with the calculation of Sec. III, one could reduce the moment problem (4.7) to a more standard form by defining a new density $Q = Q(x)$ from

$$P(\omega) = 2\omega Q(\omega^2), \quad \omega^2 = x. \tag{4.10}$$

Equations (4.7) then read

$$\int_0^\infty x^n Q(x)dx = \mu_n, \quad n = 0,1,2,..., \tag{4.11}$$

and the maximum-entropy solution is given by

$$Q_N(x) = \frac{1}{Z} \exp\left(-\sum_{n=1}^{N} \lambda_n x^n\right),$$
$$P_N(\omega) = \frac{2\omega}{Z} \exp\left(-\sum_{n=1}^{N} \lambda_n \omega^{2n}\right). \tag{4.12}$$

However, we take this opportunity to point out that the procedure is not unique. For instance, the maximum-entropy sequence obtained directly from (4.7) would read

$$P_N(\omega) = \frac{1}{Z} \exp\left(-\sum_{n=1}^{N} \lambda_n \omega^{2n}\right), \tag{4.13}$$

where $Z, \lambda_1, \lambda_2, ...$ are, of course, different from those appearing in (4.12). This fact is obviated by the appearance of the factor ω in Eq. (4.12). Had we taken into account the general expectation that $P(\omega)$ diverge rather than vanish at $\omega = 0$, the option (4.13) would be the least-biased candidate. Nevertheless, additional information about the detailed behavior of $P(\omega)$ is often unavailable in practice. How serious then is the nonuniqueness of the maximum entropy approximation?

To obtain some empirical evidence, we studied the current problem using both (4.12) and (4.13). While the details of

the corresponding pointwise approximations to $P(\omega)$ differ, the computation of Fourier averages of the form (4.6) proved much less sensitive. Such behavior is in fact typical with many examples we considered, and with more general types of nonuniqueness hinted at in Sec. V. Hence judicious choices of weights and other a priori information could lead to significant improvements in the pointwise approximation of the density, but are less likely to affect the behavior of averages.

We present here the explicit results obtained through Eq. (4.13) using five moments ($N = 5$) as input. The values for Z and $\lambda_1,...,\lambda_5$ are given in Table II for both the XY and the isotropic Heisenberg model. The corresponding approximations to the density $P(\omega)$ are plotted in Fig. 2 together with the exact result for the XY model given by Eq. (4.9). It is observed that significant oscillations are still present which could be attributed to the low number of input moments and to the combined effect of the singularity at $\omega = 0$ and the appearance of a gap in the spectrum for $\omega > \omega_0$. However, the appearance of a gap at $\omega_0 \sim 4$ in the XY model is indicated clearly in Fig. 2. Similarly, a maximum frequency in the region $5 \leqslant \omega_0 \leqslant 6$ is also indicated for the Heisenberg model in agreement with the earlier semiclassical estimates and finite-lattice calculations.[29]

Despite the limited success of the pointwise approximation of $P(\omega)$, a decisive improvement is obtained for the Fourier transform (4.9) over the earlier Gaussian fit of Morita.[28] Thus the maximum-entropy approximation (4.13) with $N = 5$ is inserted in Eq. (4.9) to produce the results for $G_0(t)$ depicted in Fig. 3. For the XY model, we plot the exact solution (4.8) by a solid line and superimpose a number of dots calculated by maximum entropy. The agreement with the exact answer, even within the third period of oscillation of
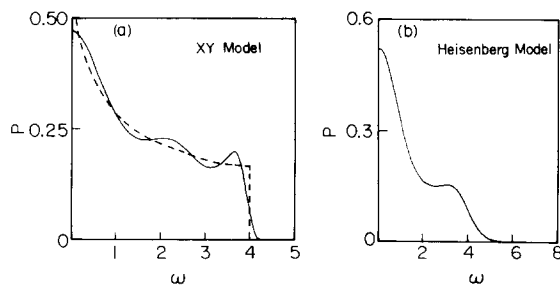

FIG. 2. A five-moment calculation ($N = 5$) of the Fourier transform of a two-point correlation function. (a) Results for the XY model and comparison with the exact answer (dashed line). (b) Corresponding results for the isotropic Heisenberg model.
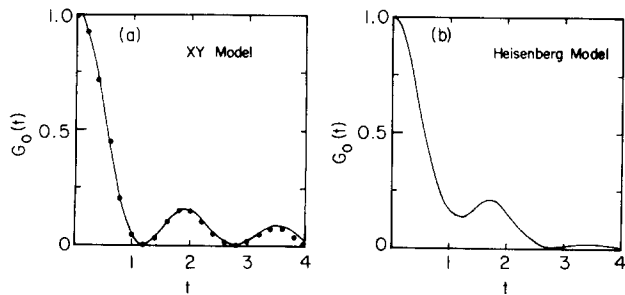
FIG. 3. A five-moment calculation ($N = 5$) of a two-point correlation function. (a) The exact result for the XY model is depicted by a solid line while the dots correspond to the results from maximum entropy. (b) Maximum-entropy prediction for the isotropic Heisenberg model.

the Bessel function, is impressive. The actual numbers are summarized in Table VI together with the calculated numbers for the isotropic Heisenberg model for which an exact solution is not known. If such a solution is found in the future, we have no doubt that the result will be graphically indistinguishable from the plot given in Fig. 3(b). These results are in concert with our earlier remarks that maximum entropy consistently gives good results for averages, even when the pointwise approximation of the actual density is relatively poor.

We conclude this section with a few remarks. Generalizations of the calculation to the completely anisotropic Heisenberg model, to nonvanishing space separation $(n - m \neq 0)$, to zero-temperature calculations, and to higher-dimensional lattices are relatively straightforward. A modest number of moments necessary for those generalizations have already appeared in the literature.[26–28] Independent knowledge of the maximum frequency ω_0 would reduce the calculation to a moment problem over a finite interval and would improve the pointwise approximation for $P(\omega)$. Ex-

perimentation with the XY model incorporating the fact that $\omega \leqslant \omega_0 = 4$ has indeed shown that the oscillations of $P(\omega)$ are reduced significantly but the results for $G_0(t)$ are essentially unchanged. In general ω_0 is not known. However, assuming that the density does vanish for $\omega > \omega_0$, the limit

$$\lim_{n \to \infty} \mu_n^{1/n} = \Lambda \tag{4.14}$$

may be shown to be finite. The frequency ω_0 is then given by $\omega_0 = \sqrt{\Lambda}$. Equation (4.14) may be used as the starting point for a numerical determination of ω_0.

Our actual calculation was performed on a semi-infinite interval. It was thus necessary to use the adaptive Newton–Cotes integration algorithm mentioned in Sec. II setting an upper limit Ω which we could vary to as large a value as 50 000. For the problems discussed in this section, a modest value in the region $\Omega = 50 \sim 100$ was sufficient. Additional complications of the type indicated by Eq. (2.38) are absent if the moments are used in odd numbers.

V. FURTHER EXAMPLES AND DISCUSSION

The explicit examples studied so far are but a few representatives of a host of moment problems that arise in practical applications. In this last section, we sketch some potential applications of the maximum-entropy approach, discuss possible generalizations, and summarize our conclusions.

(1) The Lee–Yang theory[30] of ferromagnetic Ising models provides a very elegant setting for the description of the mechanism by which phase transitions may occur. The same authors indicated that their procedure could also prove useful for practical calculations. In the interim, however, most of the calculational effort has gone into the renormalization-group approach addressing directly the critical region. Nevertheless, several authors have considered the moment problem naturally associated with the Lee–Yang theory and

TABLE VI. Detailed numerical results for the correlation function $G_0 = G_0(t)$ in the 1-D isotropic Heisenberg and XY models (using five moments), and comparison with exact result $G_0 = [J_0(2t)]^2$ for the XY model.

	Correlation function $G_0(t)$		
t	Heisenberg	XY model	$[J_0(2t)]^2$
0.0	1.000 000 00	1.000 000 00	1.000 000 00
0.2	0.922 876 11	0.922 364 75	0.922 364 75
0.4	0.723 406 49	0.716 202 28	0.716 202 28
0.6	0.479 818 14	0.450 419 16	0.450 419 16
0.8	0.275 542 35	0.207 391 15	0.207 391 13
1.0	0.160 012 87	0.050 127 34	0.050 127 08
1.2	0.133 030	0.000 008	0.000 006
1.4	0.157 033	0.034 248	0.034 238
1.6	0.184 765	0.102 559	0.102 520
1.8	0.184 296	0.153 600	0.153 483
2.0	0.149 565	0.158 018	0.157 728
2.2	0.095 432	0.117 734	0.117 345
2.4	0.044 075	0.058 805	0.057 804
2.6	0.011 684	0.013 488	0.012 164
2.8	0.001 825	0.001 868	0.000 727
3.0	0.007 064	0.022 462	0.022 694
3.2	0.015 935	0.055 607	0.059 200
3.4	0.020 049	0.076 353	0.085 905
3.6	0.017 295	0.068 951	0.087 067
3.8	0.010 458	0.034 964	0.063 303
4.0	0.003 645	− 0.008 826	0.029 464

there is little doubt that considerable progress in that direction is to be expected supplementing the renormalization-group approach.

For a ferromagnetic system described by the familiar Ising Hamiltonian

$$H = -J \sum \sigma_i \sigma_j - B \sum \sigma_i, \quad J \geqslant 0, \qquad (5.1)$$

Lee and Yang showed that the free energy may be expressed as a dispersion integral:

$$F = -B - \frac{c}{2} J - \frac{1}{\beta} \int_0^\pi \ln(1 - 2z \cos\theta + z^2) g(\theta,x) d\theta, \qquad (5.2)$$

where c is the number of neighbors, β is the inverse temperature, $x = e^{-2\beta J}$, and $z = e^{-2\beta B}$ is the activity. For temperatures below a critical value T_c, the support of the positive density $g(\theta,x)$ extends over the entire interval $0 \leqslant \theta \leqslant \pi$. For $T > T_c$, $g(\theta,x)$ vanishes in a finite region $0 \leqslant \theta \leqslant \theta_0(x)$ whose precise extent is not known. Calculational progress can be made noting that several moments of the form

$$\int_0^\pi [\cos(\theta/2)]^{2n} g(\theta,x) d\theta = \mu_n(x), \quad n = 0,1,2,\dots \qquad (5.3)$$

may be computed explicitly for all temperatures. The results of the work of several authors are summarized in Ref. 31 where the moments $\mu_n = \mu_n(x)$ were shown to be polynomials in x with remarkable properties.

Our task is to approximately construct $g(\theta,x)$ from a finite number of moments. The maximum-entropy approximation associated with (5.3) reads

$$g_N(\theta,x) = \frac{1}{2Z} \exp\left\{ -\sum_{n=1}^N \lambda_n \left[\cos\left(\frac{\theta}{2}\right)\right]^{2n} \right\}. \qquad (5.4)$$

For purposes of illustration, we give in Fig. 4 the solution with $N = 7$ and $x = 0.5$ for the 1-D Ising model ($c = 2$) for which the moments (5.3) read

$$\mu_n(x) = \frac{1}{2^{2n+1}} \frac{\Gamma(2n+1)}{\Gamma^2(n+1)} (1 - x^2)^n. \qquad (5.5)$$

The result is compared with the known exact answer

$$2\pi g(\theta,x) = \frac{\sin(\theta/2)}{[\sin^2(\theta/2) - x^2]^{1/2}}, \quad \sin\left(\frac{\theta}{2}\right) \geqslant x, \qquad (5.6)$$

which exhibits a gap for $0 \leqslant \theta < 2 \arcsin(x)$ and a singularity at $\theta_0 = 2 \arcsin(x)$. While a gap is clearly indicated by the maximum-entropy solution, significant oscillations are still present. However, thermodynamical averages calculated
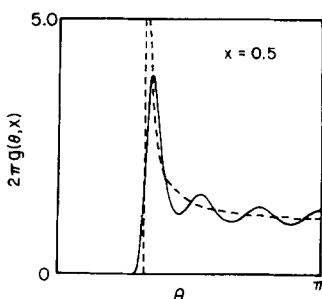
from (5.2) are obtained with much greater accuracy. As an example, consider the intensity of magnetization

$$I = -\frac{\partial F}{\partial B} = 2(1 - z^2) \int_0^\pi \frac{g(\theta,x) d\theta}{1 - 2z \cos\theta + z^2}, \qquad (5.7)$$

whose exact form for the 1-D model reads

$$I = (1 - z)/[1 - 2z(1 - 2x^2) + z^2]^{1/2}. \qquad (5.8)$$

Notice incidentally that the spontaneous magnetization $M = I(z \to 1^-)$ is given by $M = 2\pi g(0,x)$ in any dimension.

The exact result (5.8) is plotted in Fig. 5(a) for various temperatures together with results from the $N = 7$ maximum-entropy approximation. The agreement is excellent for essentially all values of the activity z and the temperature variable x. To push the calculation further, we examine the slope of the magnetization $I(z)$ at vanishing magnetic field ($z = 1^-$) for various temperatures:

$$\chi = \frac{\partial I}{\partial B}\bigg|_{B=0^+} = \beta \int_0^\pi \frac{g(\theta,x) - g(0,x)}{\sin^2(\theta/2)} d\theta. \qquad (5.9)$$

This is but the magnetic susceptibility whose exact form in the 1-D model reads $\chi = \beta/x$. It is evident from Fig. 5(b) that departures from the exact answer occur for small temperatures, that is, temperatures near the critical point of the 1-D model.

One may conclude from the preceding elementary calculation that application of the maximum-entropy approach to higher-dimensional Ising models will provide useful information for finite magnetic fields, and for vanishing magnetic fields but temperatures away from the critical point. The extent to which the critical region may be reached depends on the number of input moments and other technical details which we will not discuss further in this paper.

(2) Implicit in the discussion of Secs. III and IV was the suggestion that the maximum-entropy approach may be used to sum divergent series. A typical example is a Stieljes integral of the form

$$F(z) = \int_0^\infty \frac{P(x) dx}{1 + zx}, \qquad (5.10)$$

whose formal expansion in z may be expressed in terms of the power moments of the density $P(x)$:

$$F(z) = \sum_{n=0}^\infty (-1)^n \mu_n z^n. \qquad (5.11)$$

FIG. 4. Maximum-entropy prediction of the Lee–Yang density in the 1-D Ising model using seven moments ($N = 7$) and comparison with the exact answer (dashed line).
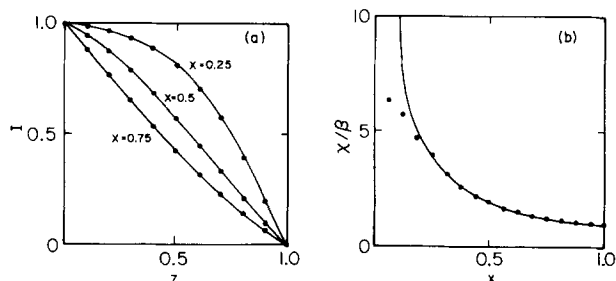

FIG. 5. Results from a $N = 7$ maximum-entropy calculation in the 1-D Ising model. (a) The solid lines correspond to the exact answer for the intensity of magnetization (I) as a function of the activity (z) at various temperatures (x). The dots are results from maximum entropy. (b) Similar results for the magnetic susceptibility.

In most practical applications the formal series (5.11) is divergent for all z. Nevertheless various methods exist for summing divergent asymptotic series, notably, Padé approximants.[8] Bender urged us to study the maximum-entropy sequence for series of the form (5.11) for which the asymptotic behavior of the moments for large n reads[32]

$$\mu_n \sim (\nu n)! \qquad (5.12)$$

Series with $\nu > 2$ are said to violate the Carleman criterion[8,13] which states that an essentially unique reconstruction of the measure $P(x)dx = d\mu(x)$ is possible from its moments if $\nu < 2$. While for $\nu > 2$ the moments do not uniquely determine $\mu(x)$, violation of the Carleman condition does not necessarily imply nonuniqueness for the average (5.10). It is also known that the diagonal and off-diagonal Padé sequences converge but their limits may not coincide with each other and with the true average $F(z)$. Bender further provided us with 20 moments for a nontrivial example with $\nu = 3$ (the ground-state energy of an octic oscillator), and with the associated Padé sequences which, indeed, stabilize away from the true answer, the latter being known from independent numerical calculation.[33]

In contrast, the first few iterations of the maximum-entropy approach showed substantial improvement over the Padé sequence, which indicates that the maximum-entropy sequence of approximations for averages of the form (5.10) might converge to the true answer even though $\nu > 2$. Because of the tremendous growth of successive moments in the current problem, we have not yet been able to incorporate a reasonably large number of moments in our numerical procedure. We thus postpone further discussion for a future occasion.

(3) We now return to the question of the inherent non-uniqueness of the maximum-entropy procedure mentioned in Sec. IV. A possible generalization may be obtained by defining a new density $Q(x)$ from

$$P(x) = w(x)Q(x), \qquad (5.13)$$

where $w(x)$ is a known positive weight whose specific form is dictated by some *a priori* knowledge of detailed properties of $P(x)$. The usual power moments of $P(x)$ are then interpreted as weighted moments of the density $Q(x)$:

$$\int_a^b x^n Q(x) w(x) dx = \mu_n. \qquad (5.14)$$

The entropy functional (1.4) is replaced by

$$S = -\int_a^b [Q(x)\ln Q(x) - Q(x)]dx$$
$$+ \sum_{n=0}^{N} \lambda_n \left(\int_a^b x^n Q(x) w(x) dx - \mu_n \right), \qquad (5.15)$$

whose extrema are of the form

$$Q(x) = \exp\left(-w(x) \sum_{n=0}^{N} \lambda_n x^n \right), \qquad (5.16)$$

and the original density reads

$$P(x) = w(x)\exp\left(-w(x) \sum_{n=0}^{N} \lambda_n x^n \right). \qquad (5.17)$$

In order to construct a potential whose minimum determines $\lambda_0, \lambda_1, ..., \lambda_N$, one should treat all Lagrange multipliers, including λ_0, on a common basis. We thus seek generalization of the potential Δ defined in Eq. (2.9) rather than the potential Γ of Eq. (2.6). The correct generalization reads

$$\Delta = \int_a^b \left[\exp\left(-w(x) \sum_{n=0}^{N} \lambda_n x^n \right) - 1 \right]dx + \sum_{n=0}^{N} \mu_n \lambda_n. \qquad (5.18)$$

Notice that the (positive-definite) Hessian of the above potential

$$\Theta_{nm} = \int_a^b dx\, x^{n+m} w^2(x)\exp\left(-w(x) \sum_{n=0}^{N} \lambda_n x^n \right) \qquad (5.19)$$

is expressed in terms of moments that are weighted by $w^2(x)$ rather than $w(x)$.

Judicious choices of the weight $w(x)$ may lead to improvements in the pointwise approximation for $P(x)$. For instance, knowledge of the asymptotic behavior of the moments as in Eq. (5.12) may be used to determine $w(x)$ so that $P(x) \sim w(x)$ at large distances. However, our experience shows that averages of the form (2.29) are substantially stable against variations of $w(x)$.

(4) We believe to have presented ample evidence for the potential as well as the limitations of the maximum-entropy approach. It is important to keep in mind that the maximum-entropy approach, just as any other approximation procedure, should not be looked upon as a panacea for the solution of all moment problems that may arise in practice. After all, a polynomial expansion would be ideal if the true density happened to be a finite polynomial, the Padé-like procedure outlined in the Introduction would be ideal if the density were a finite sum of δ-functions, maximum entropy would be ideal if the density were the exponential of a finite polynomial, and so on.

The merits of the current approach should be searched for in the larger context of the frequency of successful applications to a wide disparity of actual problems. While the limited ensemble of problems treated in this paper may not qualify for a genuine random sample, it nevertheless suggests the following appealing features for the maximum-entropy approach.

(i) Accurate averages are obtained, even when a low number of moments are available, and are substantially stable against variations in the specific mode of calculation. For instance, incorporation of suitable weights and/or other detailed information about the density (the appearance of singularities, the actual support, and so on) typically does not improve or diminish the accuracy of averages.

(ii) However, the method is flexible enough to incorporate such additional information which may result in significant improvements of the pointwise approximation of the density.

(iii) Maximum-entropy results compare favorably with corresponding results from independent powerful methods. It is fair to mention, however, that the Padé procedure often has the advantage of furnishing rigorous upper and lower bounds for the approximated averages. To the extent that the maximum-entropy technique has been developed, an accurate estimate of the error is not possible at this point.

ACKNOWLEDGMENTS

[1] C. Shannon, Bell. Syst. Tech. J. **27**, 379, 623 (1948).
[2] E. T. Jaynes, Phys. Rev. **106**, 620 (1957); **108**, 171 (1957).
[3] *The Maximum Entropy Formalism*, edited by R. D. Levine and M. Tribus (MIT, Cambridge, MA, 1979).
[4] E. T. Jaynes, Proc. IEEE **70**, 939 (1982).
[5] N. Agmon, Y. Alhassid, and R. D. Levine, J. Comp. Phys. **30**, 250 (1979).
[6] R. D. Levine, J. Phys. A: Math. Gen. **13**, 91 (1980).
[7] *Theory and Applications of Moment Problems in Many-Fermion Systems*, edited by B. J. Dalton, S. M. Grimes, J. P. Vary, and S. A. Williams (Plenum, New York, 1979).
[8] G. A. Baker and P. Graves-Morris, *Padé Approximants* (Addison–Wesley, Reading, MA, 1980).
[9] R. G. Gordon, J. Math. Phys. **9**, 655 (1968).
[10] J. G. Wheeler and R. G. Gordon, J. Chem. Phys. **51**, 5566 (1969).
[11] R. R. Whitehead, in Ref. 7, p. 235.
[12] D. V. Widder, *The Laplace Transform* (Princeton U. P., Princeton, 1946).
[13] C. M. Bender and S. A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers* (McGraw–Hill, New York, 1978).
[14] Z. Gburski, C. G. Gray, and D. E. Sullivan, "Information Theory of Line-Shape in Collision-Induced Absorption," University of Guelph, Ontario, preprint, 1983.
[15] G. G. Lorentz, *Bernstein Polynomials* (Toronto U. P., Toronto, 1953).
[16] R. B. Leighton, Rev. Mod. Phys. **20**, 165 (1948).
[17] C. Isenberg, Phys. Rev. **132**, 2427 (1963); **150**, 712 (1966).
[18] M. Steiner, J. Villain, and C. G. Windsor, Adv. Phys. **25**, 87 (1976).
[19] Th. Niemeijer, Physica (Utrecht) **36**, 377 (1967); **39**, 313 (1968).
[20] S. Katsura, T. Horiguchi, and M. Suzuki, Physica (Utrecht) **46**, 67 (1970).
[21] E. Barouch and B. M. McCoy, Phys. Rev. A **3**, 2137 (1971).
[22] H. W. Capel, E. J. Van Dongen, and Th. J. Sishens, Physica (Utrecht) **76**, 445 (1974).
[23] A. Sur, D. Jasnow, and I. J. Love, Phys. Rev. B **12**, 3845 (1975).
[24] B. M. McCoy, J. H. H. Perk, and R. E. Shrock, Nucl. Phys. B **220**, 269 (1983).
[25] G. Müller and R. E. Shrock, "Dynamic Correlation Functions in Quantum Spin Chains," Stony Brook preprint, 1983.
[26] J. H. Taylor and G. Müller, "On the Limitations of Spin-Wave Theory in $T = 0$ Spin Dynamics," Stony Brook preprint, 1983.
[27] G. Müller, Phys. Rev. B **26**, 1311 (1982).
[28] T. Morita, J. Math. Phys. **12**, 2062 (1971); **13**, 714 (1972).
[29] F. Carboni and P. M. Richards, Phys. Rev. **177**, 889 (1969).
[30] T. D. Lee and C. N. Yang, Phys. Rev. **87**, 410 (1952).
[31] J. D. Bessis, J. M. Drouffe, and P. Moussa, J. Phys. A: Math. Gen. **9**, 2105 (1976).
[32] C. M. Bender and T. T. Wu, Phys. Rev. Lett. **27**, 461 (1971).
[33] F. T. Hioe, D. MacMillan, and E. W. Montroll, J. Math. Phys. **17**, 1320 (1976).