In Maximum Entropy and Bayesian Methods, Ed. W. T. Grandy. Kluwer, 1991

MAXIMUM ENTROPY IN CONDENSED MATTER THEORY

David Drabold and Gerald Jones
Physics Department
University of Notre Dame
Notre Dame, Indiana 46556 USA

ABSTRACT. The principle of maximum entropy has recently been applied to several problems of condensed matter theory. In this paper we discuss some technical aspects of the maxent approach to these problems, and show some general properties of the applications of the method. In particular, we show that maxent can be thought of as a convenient way to close hierarchies, and to extrapolate perturbation series for quantities of physical interest. An illustration of this viewpoint is provided by an examination of the dynamics of a quantum mechanical spin system. We discuss a general maxent method for the extrapolation of power series, and apply the method both to problems of condensed matter (a virial equation of state and spin resonance problems), and to a classic example of a difficult series to handle: the anharmonic quantum oscillator with octic perturbation. We show that the inclusion of information beside Taylor coefficients is critical to obtaining a satisfactory extrapolation for the divergent perturbation series. A general maxent criterion is proposed for optimal series extrapolation.

I. Introduction

In providing a theoretical description of a physical system, the investigator commonly encounters a situation in which the minute details of the microscopic interactions are understood, but the experimentally measurable properties of real interest are not practically obtainable from the details. The most elementary example of this is the computation of the thermodynamics of a classical gas of hard spheres using only microscopic interactions. It is apparent that a complete solution of the equations of motion (and a complete specification of initial conditions!) would be capable of yielding macroscopic predictions of any kind desired. However, the dimensionality of the phase space of the system obviously precludes any implementation of a direct solution. Historically, the principle of maximum entropy has been used by researchers to make inferences about macroscopic properties from microscopic information. In most current applications of the maxent method to condensed matter theory, the philosophy is similar, though the details of the procedure differ from problem to problem. An example that we will work out in some detail is the problem of extracting the *salient* features of the dynamics of an interacting spin system using the spin-spin interactions (the Hamiltonian) and maxent to make macroscopic predictions. Analogous calculations can be performed for extracting the electronic density of states from a tight-binding model of the interactions between electrons, or for calculating the vibrational spectrum of a solid.

For each of these examples, the judicious use of maxent in conjunction with *some* information about the Hamiltonian avoids an impossibly complicated diagonalization, and helps to make the most of the information that is readily extracted. For a partial bibliography, see Drabold, Carlsson and Fedders, 1989.

II. Example: Nuclear Magnetic Resonance

In this section we explain in some detail how the paradigm discussed above is implemented for a particular class of solid state physics problems. Again, we remind the reader that the ideas presented for the spin system are transferable to other problems. Here, we describe the classic N.M.R. experiment. (A complete discussion of the background theory of N.M.R. is given in Abragam, 1961). Consider a solid in a large external and homogeneous magnetic field H_0 (in the direction $\mathbf{e}_Z$). Each atomic nucleus is assumed to posess spin 1/2. A weak (compared to H_0) linearly polarized R.F. field H_1 is applied along the $\mathbf{e}_X$ direction: $H_1(t) = H_1 cos(\omega t)$. Finally we assume that the experiment is not done at extremely low temperatures (nuclear spin energies $\ll \ kT$). Given only the spin-spin interactions, we wish to predict the steady state transverse macroscopic nuclear magnetization (along $\mathbf{e}_X$). Within the assumption of linear response theory (roughly that H_1 does not magnetically saturate the solid), the magnetization $\mathcal{M}_X$ is given by:

$$\mathcal{M}_X(t) \propto H_1 \omega \beta \sum_{i,j} \int_0^\infty d\tau cos\omega(t-\tau) G_{ij}^{XX}(\tau), \tag{1}$$

where $\beta = 1/kT$ is the inverse temperature, and

$$G_{ij}^{XX}(\tau) =< I_X^i(\tau) I_X^j(0) >, \tag{2}$$

with $I_X^i(\tau)$ a nuclear spin operator in the Heisenberg representation for time τ, and where X indicates the spatial type of spin operator and i is a site index. $<>$ denotes a thermal average on the canonical ensemble (for the assumed "high temperatures", this becomes just a trace over spin states). G is the transverse *autocorrelation function* of the spin system. From Eq. 1, if we know G, we may predict the transverse magnetization. G provides the link between theory and experiment.

The most straightforward conceivable approach to computing $G_{ij}^{XX}(\tau)$ is to compute the eigenstates of the bilinear (truncated dipolar) spin Hamiltonian $\mathcal{H}$ of the system:

$$\mathcal{H}|n>= \epsilon_n |n>, \tag{3}$$

and then use the completeness of the eigenfunctions $1 = \sum_n |n><n|$ and the definition (Eq. 2) to obtain G, and therefore $\mathcal{M}_X$ through Eq. 1. The dimension of the Hamiltonian matrix is $dim\ \mathcal{H} \ = \ 2^N$, N being the number of interacting spins. This is numerically tractable only for $N \lesssim 12$. Since we are interested in the behavior of macroscopic collections of spins, this is not very useful for most problems.

Another way to proceed is to form a *hierarchy* of correlation functions. The strategy is to develop a recurrence relation for G which we can solve, at least approximately. Exact solutions are possible for some rather simple model Hamiltonians. The hierarchy is obtained by repeated temporal differentiation as follows:

$$M_0(\tau) \equiv G_{ij}^{XX}(\tau) =< I_X^i(\tau) I_X^j(0) >, \tag{4a}$$

$$M_1(\tau) \equiv d/d\tau G_{ij}^{XX}(\tau) = -i < [I_X^i(\tau), \mathcal{H}] I_X^j(0) >, \tag{4b}$$

and, in general,

$$M_k(\tau) \equiv d^k/d\tau^k G_{ij}^{XX}(\tau) = (-i)^k < [\cdots [I_X^i(\tau), \mathcal{H}], \mathcal{H}] \cdots, \mathcal{H}] I_X^j(0) >, \tag{4c}$$

$$(k \ commutators)$$

where we have repeatedly used the Heisenberg equation of motion,

$$idX(t)/dt = [X(t), \mathcal{H}], \tag{5}$$

for any Heisenberg operator $X(t)$ ($\hbar = 1$ in this paper). The relevant feature of these equations is that the time derivative of the correlation function on the left hand side is related to the more complicated time dependence of the product of operators on the right. This can be phrased entirely in terms of multi-point (space point) correlation functions:

$$d/d\tau(\ n-point\ correlation\ function) = \sum \{(n+1)-point\ correlation\ functions\} \tag{6}$$

or more crudely:

$$d/d\tau(complicated) = (more\ complicated\ still). \tag{7}$$

In other words, the recurrence relation goes the wrong way: In trying to compute the two-point function $G_{ij}^{XX}(\tau)$, we have to estimate the more complex three-point function defined by the right hand side of Eq. 4b. This type of closure problem occurs frequently in both classical and quantum field theory. Note that the hierarchy has led us back to where we started from with the notion of diagonalizing $\mathcal{H}$: the information we want is in $G_{ij}^{XX}(\tau)$; there is much more than we need (or can handle) in the three-point or higher correlation functions that the hierarchy compels us to consider. We face a "missing information proliferation" with the increasingly complex members of the hierarchy. At this juncture we could proceed in one of two ways: (1) We could introduce an approximation which represents the n-point function as a functional of m-point functions with $m \leq n$. This is the conventional "decoupling approximation" (Kubo et. al., 1985). (2) We can treat the calculation of $G_{ij}^{XX}(\tau)$ as an exercise in power series extrapolation. Of these alternatives, (1) is the more dangerous, since it is "uncontrolled" (e.g., we don't know very clearly how the approximation will cause the approximate solution to depart from the exact solution we seek). This is a standard procedure in all cases where hierarchies are encountered.

To see that the spin correlation function hierarchy can be handled as an extrapolation procedure, one only needs to consider Eqs. 4, and evaluate both sides for $\tau = 0$. The right hand side may now be evaluated by simple (but tedious) algebraic manipulations. Obviously, these calculations yield the first few Taylor (MacLaurin) coefficients M_k for G:

$$G_{ij}^{XX}(\tau) = \sum_{k=0}^{\infty} M_k \tau^k/k! \tag{8}$$

Here, we use the notation $M_k \equiv M_k(\tau = 0)$. For our spin problem, four non-trivial coefficients are known. It would be a significant undertaking to compute more. Simply truncating the Taylor expansion would lead to intolerable errors for large time, and for

sufficiently large times the truncated representation becomes altogether meaningless. If we can obtain an accurate extrapolation of G, we have solved our N.M.R. problem.

In the field of spin dynamics it is well known that the cosine transform of the "total G":

$$G(\omega) = \int_0^\infty d\tau cos(\omega\tau) \sum_{ij} G_{ij}^{XX}(\tau) \tag{9}$$

has the interpretation of being the (positive, additive) spectral density of excitation for the spin system. Here, expansion coefficients M_{2k} are the power moments of the spectral density. This is just the usual relation between the Taylor coefficients in one space and power moments in the Fourier (or Laplace) transform space. The first application of maxent to this problem (Fedders and Carlsson, 1985) was entirely implemented in $\omega-$space, where the calculation is reduced to solving the Stieltjes moment problem for $G(\omega)$ (for high temperatures the odd moments vanish):

$$\int_0^\infty d\omega\omega^{2k} G(\omega) = M_{2k} \qquad (0 \leq k \leq 4). \tag{10}$$

The application of maxent to approximating $G(\omega)$ by solving the moment problem (Eq. 10) is justified because $G(\omega)$ is a spectral density. The solution of the moment problem may be viewed as a solution to the extrapolation problem: the cosine transform of the maxent $G(\omega)$ provides an excellent extrapolation of $G(\tau)$. The details of the solution will not be given here: we refer the interested reader to the literature. It was also found that information about the behavior of $G(\tau)$ in the complex plane was useful in producing improved extrapolations. The results of the maxent extrapolation led to unprecedented agreement between theory and a very precise experiment performed on the fluorine nuclear spins in a single crystal of CaF_2; the difference between theory and experiment being *uniformly* less than 2% in the spectral function in the measurable part of the line. We see that maxent has helped us to escape from *ad-hoc* attempts to close the hierarchy, and has given us a means of using the *relevant* information obtained from the Hamiltonian in a way free of unwanted artifacts that might be encountered in an *ad-hoc* solution of the moment problem. In the next section we show in considerable generality how to turn the series extrapolation problem into a moment problem, even when there is no obvious positive function such as the spectral density $G(\omega)$ of this example.

III. A Maxent Variational Principle for Extrapolating Series

The previous section was presented to show how decoupling approximations for a hierarchy could be avoided by a reformulation of the problem into one of missing information, and to give a typical example of a condensed matter problem for which a series extrapolation/moment problem is encountered. Similar extrapolation problems occur in electronic structure calculations, the computation of vibrational spectra, and critical phenomena, where phase transitions are treated through a study of analytic continuations of power series and extrapolations of high temperature expansions. Of course there are also many applications outside condensed matter theory. In this section we consider the rather general mathematical problem:

Suppose that $f(x)$ is analytic in a neighborhood of $x = 0$. Given Taylor coefficients of f: $\{a_n\}_{n=0}^{N}$, and possibly additional information (the asymptotic behavior of a_l as $l \to \infty$, or of $f(x), x \to x_0$, etc.), What is the "best" approximate representation for $f(x)$, for all x in the domain of f?

In this section we will map the extrapolation problem onto a certain type of moment problem, discuss the necessary and sufficient conditions for the existence of solutions to the moment problem on the finite interval, propose a general maxent criterion for optimal extrapolation, and work some examples of interest. We are motivated to work with an integral representation of the extrapolated series because it is necessary to translate the purely *local* information given by the Taylor coefficients into global constraints on maxent: this is somewhat like introducing pixel-pixel correlations in the image reconstruction problem (Gull, 1989). Maxent does not directly provide useful answers for detailed information *concerning one point*. We note that there is certainly no unique definition of "best" extrapolation, but the following approach seems very reasonable to us. Our method is implemented in the following way:

We begin by assuming that f may be expressed in the form of an integral representation with multiplicative kernel K:

$$f(x) = \int_A^B d\xi \rho_K(\xi) K(x\xi), \tag{11}$$

such that ρ_K is a positive, additive density (Skilling, 1989): conditions for which this is the case are discussed later. In order to incorporate the information we know (the Taylor coefficients), we Taylor expand the LHS of Eq. 11, and the kernel on the RHS. One easily obtains:

$$a_l = \mu_l k_l, \qquad l = 0, 1, 2...N, \tag{12}$$

where k_l are the Taylor coefficients of the kernel K and

$$\mu_l = \int_A^B d\xi \xi^l \rho_K(\xi). \tag{13}$$

This is a (N+1) constraint moment problem for the function ρ_K, something which maxent is very well suited to (Collins and Wragg, 1977, Mead and Papanicolaou, 1984).

An important point for the implementation of this method, and indeed for any maxent calculations involving the finite-interval (Hausdorff) moment problem is that necessary *and* sufficient conditions are known for the existence of a positive density ρ_K. To determine whether a moment problem is well-posed (meaning here that $\rho > 0$), we may apply the conditions given in Appendix 1 to the given moments and interval.

The heart of the problem is choosing a *particular* kernel K: in practice there are continuous families of kernels which lead to soluble moment problems, and each of which leads to a somewhat different extrapolation. To solve this problem, we have found it useful to use the maxent principle to pick not only the Lagrange multipliers in solving the moment problem, but *also in picking the optimal kernel*. In particular, we take as the best estimate of $f(x)$ that integral representation (kernel) within the family under study, yielding a $\rho_K(\xi)$ with maximum entropy:

$$H(K) = -\int_A^B d\xi \ \rho_K(\xi) \ log\rho_K(\xi) \ = \ maximum, \tag{14}$$

while agreeing with the conditions (12). If we denote the maxent density as ρ_M and the associated kernel K_M, then our extrapolation for $f(x)$ takes the form:

$$f^{extrap}(x) = \int_A^B d\xi \rho_M(\xi) K_M(x\xi). \tag{15}$$

In practice, one chooses the family of kernels $K(u)$ which has the qualitative features expected of $f(u)$: the kernels can be chosen to be intelligent ad-hoc guesses for f. Maxent can then be used to select a best kernel out of the family. It is important to note that a given set of kernels yields a *unique* representation for the extrapolation of the power series, if there is a unique maximum of the entropy functional (4) (the usual case). The selection of the kernel gives us considerable flexibility in including prior information: for example, we can easily select a kernel so that the integral representation (Eq. 11) has known asymptotic behavior built in. If we are vague about our prior information, but feel that it must be included in some form, we may introduce a parameter into the kernel to incorporate that information and use maxent to fix the undetermined parameter. This criterion can therefore be regarded as a variational principle of series extrapolation quite analogous in philosophy to the Rayleigh-Ritz method (Courant and Hilbert, 1953) of calculating eigenvalues from trial eigenfunctions with variationally determined free parameters. And of course like the Rayleigh-Ritz method, a sufficiently poor choice of kernel will lead to unsatisfactory results! Indeed, a completely inappropriate kernel will lead to a moment problem for which there is no positive, additive weight function ρ_K.

This is a rather easy program to implement (at least after the numerical analysis for solving the indicated moment problems is complete). We have found that rather simple choices of the kernel lead to well posed maxent moment problems (equations A2 implying the existence of a nonnegative solution ρ_K), and that the extrapolated estimates for $f(x)$ are often quite weakly dependent upon K: the differences between different kernels being largely absorbed into the positive weight function ρ_K. The entropy H however, tends to depend substantially on K for the simple kernels which we have investigated so far. We note that there is nothing restricting the choice of kernel to the type given in Eq. 11: kernels which are not functions of the product form $K(x\xi)$ just lead to a "generalized moment problem" in which the constraint equations do not involve simple powers, but more complicated functions. For some problems the form of the expansion would naturally lead to the generalized case. For example, in physics kernels of the form $K(x-\xi)$ are very common, and lead to a generalized moment problem. Examples of this type are currently under study.

One aspect of this method needs to be treated with care. Because we are working with continuous densities, we must be aware of the choice of measure (Jaynes, 1968, Jaynes, 1978). For the integral representation above (Eq. 11) the measure is fixed by our choice of representation: this is an example of the use of our prior expectations about the series we are extrapolating.

In outline, the numerical implementation of the method is the following. For a given kernel and interval, apply the moment conditions (Eqs. A2) to see whether the problem is well posed or not. These simple conditions, which involve the diagonalization of a pair of small matrices, are of great guidance in the choice of kernels, and finding ranges of parameters in the kernel over which the moment problem may be addressed. If the spectrum of the moment-test matrices is nonnegative, we proceed to solve a discrete approximation to

the linear maxent problem (Eqs. 13,14) using the method of Bretthorst (Bretthorst, 1987), and then polish the Lagrange multipliers for the continuous problem using an improved version of the original Newton minimization method given by Mead and Papanicolaou. (It is important to note that, in general, the discrete Lagrange multipliers are different from the continuous multipliers). The power series is then easily extrapolated with a numerical quadrature of Eq. 15. This method is quite stable for up to about 12 coefficients, at which point the continuous polish can sometimes become unstable.

A simple illustrative example.

To illustrate the practical use of the method outlined above, we begin with an elementary test case. We consider the algebraic function

$$f(x) = (1 + \pi x + \gamma x^2)^{-\pi}, \tag{16}$$

where π and γ are the usual mathematical constants. We observe that f has the Taylor expansion:

$$f(x) \sim 1 - 9.869x + 62.394x^2 - 322.117x^3 + 1479.182x^4 - ... \quad x \to 0. \tag{17}$$

To illustrate the method, we suppose that our only knowledge about f is (i) some Taylor coefficients and (ii) f decays according to *some* power-law for large argument x. As we suggested above, the kernel $K(u)$ may be chosen to be an intelligent *ad-hoc* guess for $f(u)$. The information (ii) leads us to try an integral representation for f of the form:

$$f(x) = \int_0^\infty d\xi \ \rho_\alpha(\xi) \ (1 + x\xi)^\alpha, \tag{18}$$

with $\alpha < -1$, $x > 0$. We will use the maxent principle to determine α. We study extrapolations based upon four and six Taylor coefficients. Although ρ_α of Eq. 18 is defined on an infinite interval, we find the finite interval moment test conditions of Appendix 1 useful because ρ is essentially zero for $\xi > 5$: so for numerical purposes we cut the integral off outside this range and the conditions provide meaningful information concerning the existence of a positive density on the (finite) interval we integrate over. By following the method described above, we find that the choice $\alpha \approx -5.75$ leads to a maximum in the entropy functional (Eq. 14) for both four and six coefficients. We reproduce the results of the extrapolation in Fig. 1 along with results for the [2/2] and [3/3] Padé approximants (Bender and Orszag, 1978). Note that the results for the four term maxent extrapolation are much better than the Padé extrapolation. Indeed, the four term maxent extrapolation is better than the [3/3] Padé, even though the latter has two more Taylor coefficients of information. The six term maxent extrapolation is seen to be in nearly perfect agreement with the exact result. It is also interesting to note that our procedure produces a leading order asymptotic behavior close to the true behavior:

$$f^{extrap}(x) \sim c_1 x^{-5.75} \quad x \to \infty \tag{19}$$

(c_1 a real constant) to be compared to

$$f(x) \sim x^{-2\pi}/\gamma^\pi \quad x \to \infty. \tag{20}$$

If one *requires* $\alpha = 2\pi$ (corresponding to knowledge of the true asymptotic behavior), the resulting extrapolation is very accurate (much better than the Padé sums), but not quite as accurate as the maxent choice for α. This can be easily understood: the simple kernel of integral representation Eq. 18 does not have the flexibility to include obtain both the power of the decay and the prefactor. Maxent is better able to represent $f(x)$ by compensating for this by slightly reducing the power of the decay law.

Virial Equation of State for Hard Spheres.

It is well known that physical systems are almost always most easily described in various *limits*. In particular, it is much easier to treat the thermodynamics of a very dilute gas or a close packed solid than the difficult liquid state. In either limit there are important simplifications which allow progress on the problem. An important undertaking is the extension of the results from a limiting regime to the more difficult (intermediate) case.

Using the methods of classical statistical mechanics it is possible to express the thermodynamic pressure as an expansion in powers of the density of the gas (Grandy, 1988). It is reasonably straightforward to obtain the first few of these coefficients from analytical computation or simulation for a given pair potential. Here, we will consider the case of a gas of classical hard spheres, and use maxent to sum the virial series for the pressure to obtain an information theoretic equation of state for the system. If we set $x = \eta/\eta_0$, where η_0 is the close-packing density of the spheres, then we expect that the pressure p should have a singularity near $x = 1$, and (presumably) monotonically increase from $x = 0$. Using the method outlined above, our prior expectations about the reduced pressure $P(x)$ $(= pV/NkT)$ lead us to choose the one parameter family of integral representations (indexed by α):

$$P(x) = \int_{-\infty}^{1} d\xi \ \rho_\alpha(\xi) \ (1 - x\xi)^\alpha, \tag{21}$$

where by the assumption that the pressure is singular near $x = 1$, we have $\alpha < -1$. Using the six known virial coefficients (Ree and Hoover, 1967) we have found that the maximum entropy over the kernels K of Eq. 21 occurs at $\alpha \approx -17.5$, or

$$P^{extrap}(x) = \int_{-\infty}^{1} d\xi \ \rho_{-17.5}(\xi) \ (1 - x\xi)^{-17.5}. \tag{22}$$

The Lagrange multipliers for the ρ_α with maximum entropy are given in Appendix 2.

The extrapolation resulting from Eq. 22 is in excellent agreement with numerical simulations (Ree and Hoover, 1967) (Fig. 2). It is worth contrasting the present work with Padé approximant continuations of the virial expansion. In the most comprehensive Padé treatment (Baker et. al., 1984), 27 different Padé continuations were been constructed from the six known virial coefficients. Those Padé extrapolations which most resemble Monte Carlo or molecular dynamics data are then compared to the simulation "data". The Bernal random close packing density of $x \approx 0.87$ is claimed to be extracted from two of the virial extrapolations. This is a somewhat questionable claim, since the authors also claim to see spinodal effects from some of the Padés: presumably, however, there is only one branch of the hard sphere phase diagram which is represented by the virial coefficients. The most disturbing feature of the Padé analytic continuation is the large number of completely different answers obtained for the equation of state. We regard the *uniqueness* of our extrapolation as being a significant point in its favor.

Anharmonic Oscillator: Using Prior Information.

In this section we briefly discuss an improved extrapolation of the ground state eigenvalue $E_0(g)$ of the quantum harmonic oscillator with octic perturbation. The Hamiltonian is:

$$\mathcal{H} = p^2/2 + x^2/2 + gx^8. \tag{23}$$

This problem has been studied by Bender, Mead and Papanicolaou (Bender et. al., 1987) using maxent and a particular integral representation of $E_0(g)$. These authors have shown that Rayleigh-Schrödinger perturbation theory for $E_0(g)$ leads to a divergent power series in the coupling constant:

$$E_0(g) \sim 1/2 + \sum_{n=1}^{\infty} (-)^{n+1} A_n g^n \quad g \to 0. \tag{24}$$

The coefficients grow exceedingly fast ($A_n \sim (3n)!$). Using 5 expansion coefficients, they found results *much* improved over Padé extrapolations. Here, we show that using an additional easily obtained (Hioe et. al., 1976) piece of information concerning the asymptotic behavior of $E_0(g)$: namely that

$$E_0(g) \sim g^{1/5} \quad g \to \infty, \tag{25}$$

greatly improves the earlier extrapolation; better than splitting the difference between the exact (numerically obtained) result and the earlier maxent extrapolation. Following Bender et. al. 1987, we reconstruct the function

$$F_0(g) = [E_0(g) - 1/2]/g, \tag{26}$$

rather than E_0 directly. The known asymptotic behavior of E_0 implies that $F_0(g) \sim g^{-4/5}$, leading us to choose the integral representation:

$$F_0(g) = \int_0^{\infty} d\xi \ \rho_\alpha(\xi) \ (1 + g\xi)^{\alpha}, \tag{27}$$

with $\alpha = -4/5$. Bender et. al. used the representation (Eq. 27) with $\alpha = -1$. Following the procedure indicated at the beginning of the section, we obtain the results presented in Fig. 3. The Lagrange multipliers are given in Appendix 3. It is clear that the choice $\alpha = -4/5$ produces a much better fit than that of Bender et. al. It is also apparent that the *prefactor* of the asymptotic behavior (Eq. 25) is different for the exact answer and our representation. It is not possible with the simple representation (Eq. 27) to obtain the asymptotic behavior exactly. We are presently investigating a broader class of kernels so that the prefactor can be exactly obtained.

For this example we did not find it useful to maximize the entropy for selecting the kernel (of course we used maxent to solve the moment problem). We considered the integral representation (Eq. 27) for general α, and find that the entropy is a monotonically decreasing function of α over the range of α for which the moment problem was well posed, and that we could numerically investigate. We therefore choose the simpler form (Eq. 27) with $\alpha = -4/5$, which properly employs prior information which is important to obtaining

an improved extrapolation. The extrapolated energy eigenvalue is recovered from Eq. 26. We speculate that the lack of a well defined maximum in the entropy may be related to the remarkable divergence of the A_n. Although we do not present the results here, we have empirically shown that the choice $\alpha \approx 0.72$ yields a *very* accurate extrapolation of $E_0(g)$ for $0 < g < 20000$: this is consistent with the first example we gave, where a slight sacrifice in asymptotic behavior led to an improved extrapolation. Note that the inclusion of information concerning large-g behavior improves the extrapolation, even for rather small g.

Conclusions concerning maxent extrapolation.

For each of the three examples presented above, the maxent approach we outlined led to results superior to conventional Padé methods. The method has three virtues: (1) it is easy to include important prior information about the function being extrapolated, (2) the method usually produces a *unique* extrapolant for the given information provided and family of kernels under investigation and (3) by employing the principle of maximum entropy, we make the safest, "most conservative" guess consistent with the given information. The usefulness of the approach is in no way limited to divergent perturbation series; it is useful for any extrapolation problem for which a limited number of expansion coefficients are available. We have shown that the numerical results of the procedure are very satisfactory for three very different types of series.

ACKNOWLEDGMENTS.

The authors have benefited from discussions with R. Kennedy, A. Carlsson and P. Fedders. We are grateful to the organizers of MaxEnt '90 for financial assistance.

Appendix 1: The finite-interval moment problem

In this Appendix, we give the necessary *and sufficient* conditions for the solution of the Hausdorff moment problem. These results have been copied from the mathematical literature (Akhiezer, 1965). We include them here because they are very useful in maxent calculations involving moment problems. Define:

$$\mu_k = \int_0^1 x^k \rho(x) dx \qquad k = 0, 1, 2, ..., n < \infty. \tag{A1}$$

If it is true that the following quadratic forms satisfy:

(for an **odd** number of moments: $n = 2m + 1$)

$$\sum_{i,k=0}^{m} \mu_{i+k+1} x_i x_k \quad > \; 0 \tag{A2O}$$

and

$$\sum_{i,k=0}^{m} (\mu_{i+k} - \mu_{i+k+1}) x_i x_k \quad > \; 0$$

(for an **even** number of moments: $n = 2m$)

$$\sum_{i,k=0}^{m} \mu_{i+k} x_i x_k \quad > \; 0 \qquad (A2E)$$

and

$$\sum_{i,k=0}^{m-1} (\mu_{i+k+1} - \mu_{i+k+2}) x_i x_k \quad > \; 0$$

then the moment problem Eq. A1 has a positive solution $\rho(x)$ for $0 \leq x \leq 1$. These conditions are easily generalized to the general finite-interval moment problem by a linear transformation. Practically speaking, these conditions mean that given a set of moments μ_n, on a given finite interval, we must transform the problem to the interval (0,1), then form the real-symmetric matrices defined by the quadratic forms above, and diagonalize them. If all the eigenvalues for both matrices are positive, we are guaranteed that a positive ρ satisfying the moment problem exists. Empirically we have found examples for which $\rho > 0$ exists, but such that we cannot find the solution for numerical reasons. This happens for kernels leading to quadratic forms with eigenvalue spectra positive, but with at least one eigenvalue very near zero.

Appendix 2: Lagrange Multipliers for Hard Sphere Problem: $\alpha = -17.5$.

The maxent density is $exp\ \{-\sum_{l=0}^{6} \lambda_l \xi^l\}$.

$\underline{l}$	$\underline{\lambda_l}$
0	-0.881413
1	3.788254
2	1257.488
3	-19566.264
4	100422.993
5	-207176.972
6	149166.371

Appendix 3: Lagrange Multipliers for Octic Oscillator: $\alpha = -4/5$.

The maxent density is $exp\ \{-\sum_{l=0}^{5} \lambda_l \xi^l\}$.

$\underline{l}$	$\underline{\lambda_l}$
0	3.67728
1	4.66059E-03
2	-1.34174E-06
3	1.78886E-10
4	-9.24848E-15
5	1.59402E-19

REFERENCES

Abragam, A.: 1961, *Principles of Nuclear Magnetism*, Oxford, Clarendon.

Akhiezer, N. I.: 1965, *The Classical Moment Problem*, Oliver and Boyd, Edinburgh p. 74.

Baker, G., Gutierrez, G. and de Llano, M.: 1984, *Ann. Phys.* **153** 283.

Bender, C. M., and Orszag, S.: 1978, *Advanced Mathematical Methods for Scientists and Engineers*, McGraw-Hill, New York.

Bender, C. M., Mead, L. R. and Papanicolaou, N.: 1987, *J. Math. Phys.* **28** 1016.

Bretthorst, G. L.: 1987 (unpublished).

Collins, R. and Wragg, A.: 1977, *J. Phys. A* **10** 1441.

Courant, R. and Hilbert, D.: 1953, *Methods of Mathematical Physics, Vol. 1*, Interscience, New York.

Drabold, D. A., Carlsson, A. E., Fedders, P. A.:1989, in J. Skilling (ed.), *Maximum Entropy and Bayesian Methods*, Kluwer, Dordrecht, Holland.

Fedders, P. A. and Carlsson, A. E.: 1985, *Phys. Rev. B* **32** 229.

Grandy, W. T.: 1988, *Foundations of Statistical Mechanics, I*, Reidel, Dordrecht, Holland. Chapter 7.

Gull, S.: 1989, in J. Skilling (ed.), *Maximum Entropy and Bayesian Methods*, Kluwer, Dordrecht, Holland.

Hioe, F. T., MacMillen, D. and Montroll, E. W.: 1976, *J. Math. Phys.* **17** 1320.

Jaynes, E. T.: 1957, *Phys. Rev.* **106**, 620.

Jaynes, E. T.: 1968, *IEEE Transactions on System Science and Cybernetics*, SSC-4, 227.

Jaynes, E. T.: 1978, in R. D. Levine and M. Tribus (eds.), *The Maximum Entropy Formalism*, MIT Press, Cambridge. The articles of E. T. Jaynes cited here may also be found in Jaynes, E. T.: 1983, *Papers on Probability, Statistics and Statistical Physics*, Reidel, Dordrecht, Holland.

Kubo, R., Toda, M., and Hashitsume, N.: 1985 *Statistical Mechanics II: Nonequilibrium Statistical Mechanics*, Springer, Berlin.

Mead, L. R. and Papanicolaou, N.: 1984, *J. Math. Phys.* **25** 2404.

Ree, F. H. and Hoover, W. G.: 1967, *J. Chemical Phys.* **46** 4181.

Skilling, J.: 1989, in J. Skilling (ed.), *Maximum Entropy and Bayesian Methods*, Kluwer, Dordrecht, Holland.

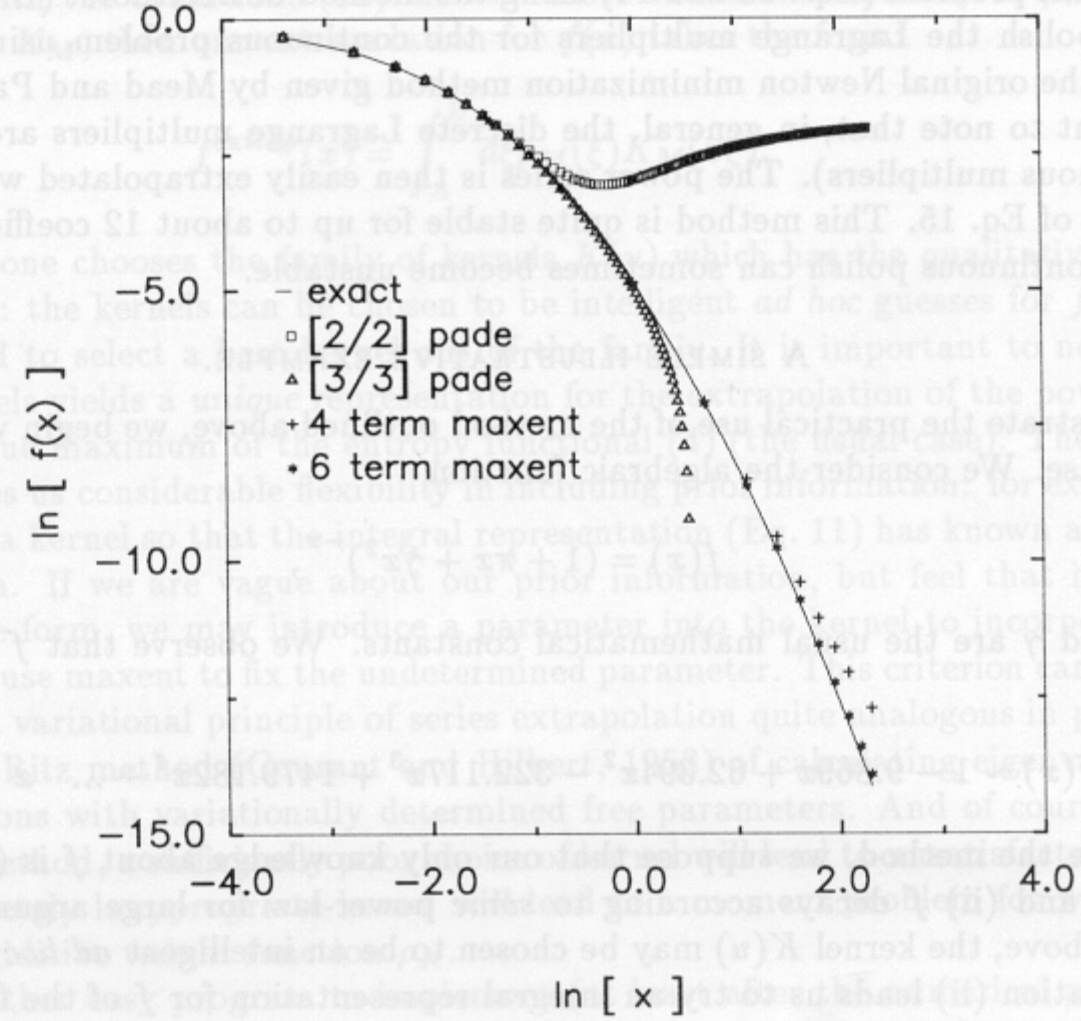

Fig. 1. Extrapolations of Equation 17.

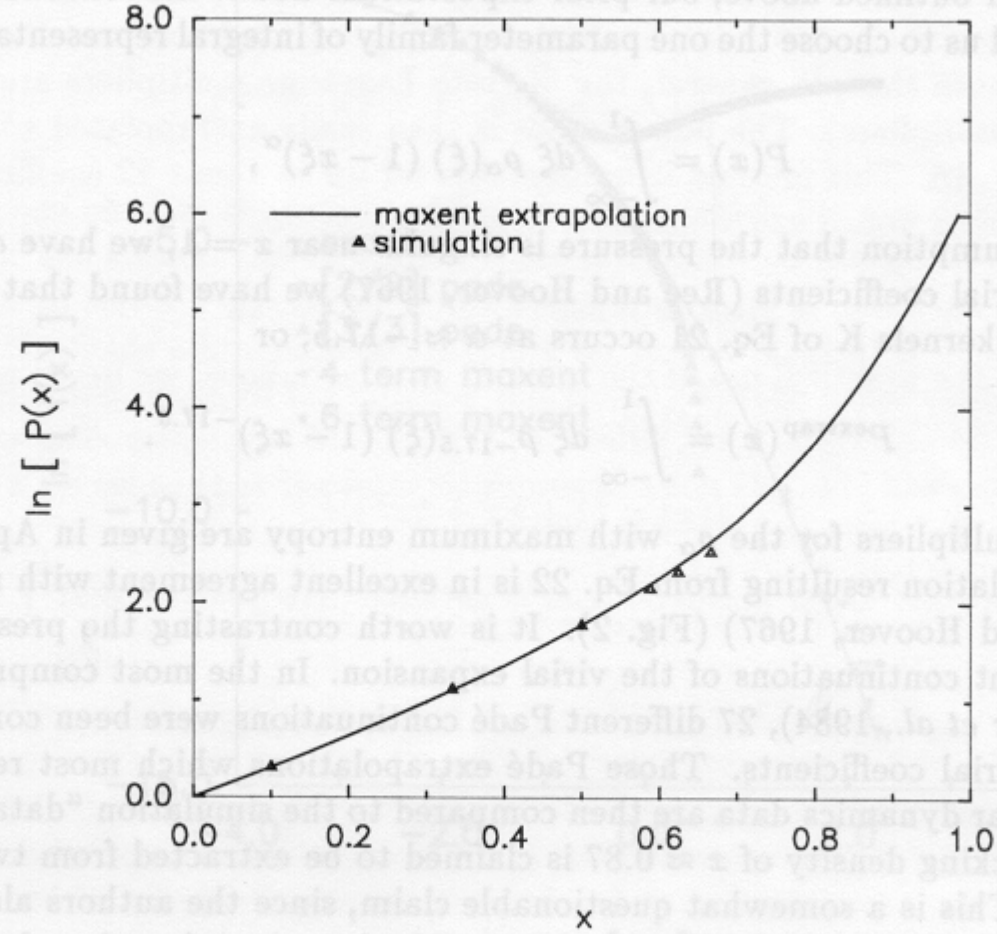

Fig. 2. Maxent Virial Equation of State. The solid curve is the extrapolation from Eq. 22, the triangles are from computer experiments.

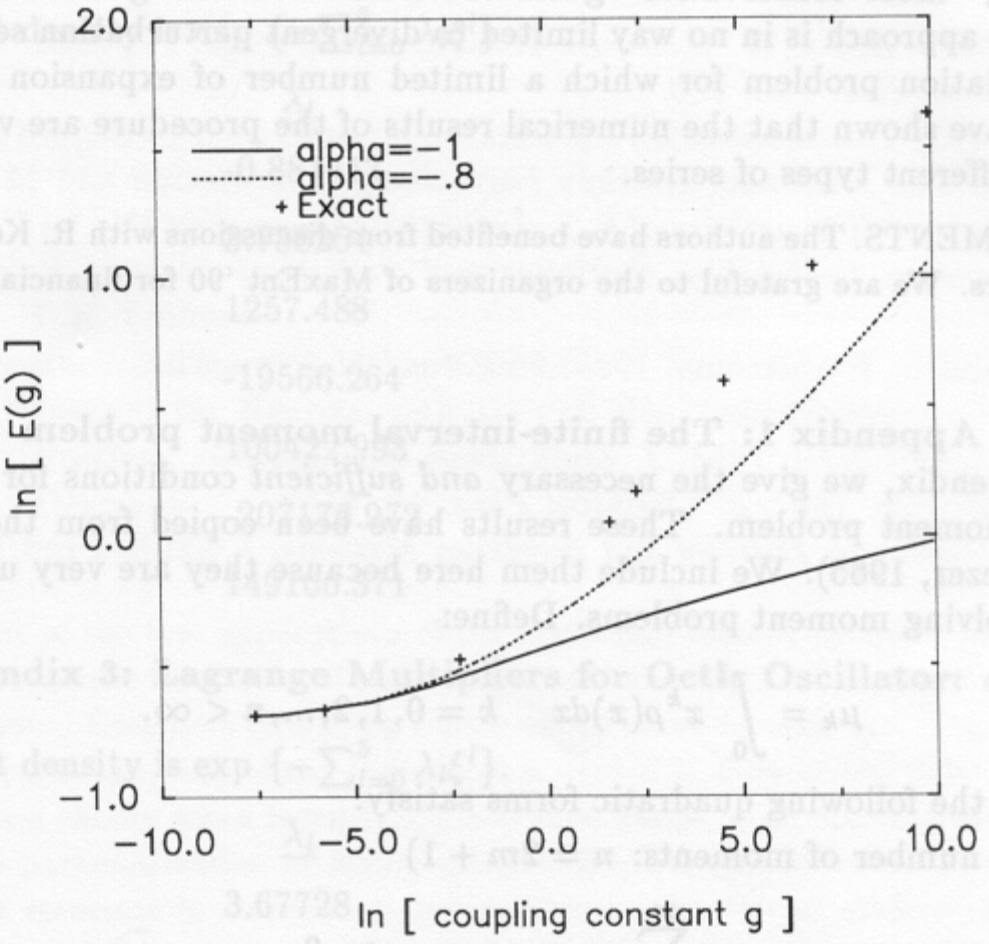

Fig. 3. Octic Oscillator ground state eigenvalue vs. coupling constant g.