Maximum-entropy approach to series extrapolation and analytic continuation

1991 J. Phys. A: Math. Gen. 24 4705

(http://iopscience.iop.org/0305-4470/24/19/029)

J. Phys. A: Math. Gen. 24 (1991) 4705-4714. Printed in the UK

Maximum-entropy approach to series extrapolation and analytic continuation

D A Drabold† and G L Jones

Department of Physics, University of Notre Dame, Notre Dame, IN 46556, USA

Received 19 April 1991

Abstract. We use information theory (the principle of maximum entropy) to develop an approach to the problem of extrapolating power series. We suggest a well-defined way to map the extrapolation problem onto a moment problem, and show that the use of additional information about the function being extrapolated (such as asymptotic behaviour for large arguments) is important to obtaining accurate extrapolations. We apply the method to the virial expansion for the classical hard sphere equation of state, the quantum harmonic oscillator with octic perturbation and the symmetric Anderson model of relevance to magnetic impurities in metals. In each case the method yields excellent pointwise estimates of the function being extrapolated.

1. Introduction

The problem of extrapolation or analytic continuation of power series is important in virtually all fields of theoretical physics. In condensed matter physics, extrapolation is intimately related to problems in critical phenomena and can be related to the closure of hierarchies in field theories. In other areas, perturbation theory often produces a power series in some coupling parameter which is of primary interest for large values of the coupling parameter. It is therefore of considerable interest to develop new attacks on the extrapolation problem. This paper represents a step towards a unified approach to apply the methods of information theory (the method of maximum entropy (maxent)) to the problem. A preliminary report of some of this work is given in Drabold and Jones (1991).

Other workers have noted the possibility of using maxent as an extrapolation technique. The idea was first mentioned by Mead and Papanicolaou (1984), and first implemented by Bender *et al* (1987) for the case of a harmonic quantum oscillator with octic perturbation. The basic idea was to map the extrapolation problem onto a moment problem, then the moment problem was solved using maximum entropy. While this approach is reasonable, it makes no attempt to address the question of *how* to choose an integral representation (mapping). This is important, since the extrapolation depends upon this choice. Here, we use information about asymptotic behaviour to choose the kernel. We show that such information can greatly improve maxent extrapolations.

The rest of this paper is organized as follows. Section 2 outlines our formulation of the extrapolation problem as one of missing information. Section 3 provides three

† Present address: Department of Materials Science and Engineering, University of Illinois, Urbana, IL 61801, USA.

0305-4470/91/194705+10$03.50

examples of the technique from diverse areas of physics. We present a maximum-entropy equation of state for hard spheres from the virial expansion, an improved version of the octic oscillator problem, and an extrapolation of the spin susceptibility of the symmetric Anderson model (Anderson 1961) of magnetic impurities in metals.

2. Maximum-entropy and power series extrapolation

The method of maximum entropy provides a variational principle which yields an explicit recipe for inferring a 'best guess' for a positive, additive distribution (Skilling 1989) (essentially a probability distribution, either discrete or continuous) when we possess incomplete testable information about the distribution (Jaynes 1983). In the case of a discrete distribution p_i, this formal procedure consists of maximizing the entropy functional:

$$H_d = -\sum_i p_i \log(p_i) \tag{1}$$

subject to whatever information is available in the problem. In many instances, this constrained optimization problem is implemented using Lagrange multipliers in the expected way. The continuous analogue of equation (1) is

$$H_c = -\int dx\, p(x) \log[p(x)/\nu(x)] \tag{2}$$

where $\nu(x)$ is a measure (Jaynes 1968, 1983). Shannon and Weaver (1962) have proved that the form (1) is unique given some highly plausible requirements on H_d. In some contexts (statistical mechanics for example), entropy can also be thought of as a measure of the degeneracy of solutions, and the maximum-entropy solution is that solution which has the highest multiplicity consistent with the constraints (incomplete information) available.

At present maxent is being used with success in image reconstruction (for example Gull and Daniell 1979), analysis of noisy experimental data (Gull 1989), and analysis of computer simulated data (Silver *et al* 1990). It is well beyond the scope of this paper to further discuss maxent or its applications. We direct the reader to the burgeoning literature on the subject.

We now consider the application of maxent to extrapolation. Consider the following rather general problem.

Suppose that $f(x)$ is analytic in a neighbourhood of $x = 0$. Given Taylor coefficients of f: $\{a_n\}_{n=0}^{N}$, and additional information (the asymptotic behaviour of a_l as $l \to \infty$, or of $f(x)$, $x \to x_0$, etc), find an accurate approximate representation for $f(x)$, for all x in the domain of f.

In this section we will map the extrapolation problem on to 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, and stress the importance of using prior information in forming an extrapolation. 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 begin by assuming that f may be expressed in the form of an integral representation with multiplicative kernel K:

$$f(x) = \int_{\mathscr{D}} \mathrm{d}\zeta\, \rho_K(\zeta) K(x\zeta) \tag{3}$$

such that ρ_K is a positive, additive density: conditions for which this is the case are discussed later. $\mathscr{D}$ is some interval on the real line. In order to incorporate the information we know (the Taylor coefficients), we Taylor expand the LHS of equation (3), and the kernel on the RHS. One easily obtains

$$a_l = \mu_l k_l \qquad l = 0, 1, 2, \ldots, N \tag{4}$$

where k_l are the Taylor coefficients of the kernel K and

$$\mu_l = \int_{\mathscr{D}} \mathrm{d}\zeta\, \zeta^l \rho_K(\zeta). \tag{5}$$

This is an $(N+1)$-constraint moment problem for the function ρ_K, something which maxent is very well suited to (Collins and Wragg 1977, Mead and Papanicolaou 1984). It is a simple exercise to see that the maxent solution to the moment problem takes the form

$$\rho_K(\zeta) = \exp\left(-\sum_{l=0}^{N} \lambda_l \zeta^l\right) \tag{6}$$

where λ_l are chosen to satisfy equations (4) and (5). For $N > 2$, the λ_l are obtained numerically.

An important point for the implementation of this method, and indeed for any maxent calculations involving the finite-interval (Hausdorff) moment problem is that practically useful necessary *and* sufficient conditions on the moments are known for the existence of a positive density ρ_K (Akheizer 1965). To determine whether a moment problem is well posed (meaning here that $\rho > 0$), we may apply the conditions (Akheizer 1965, Drabold and Jones 1991) to the given moments and interval. These conditions form one criterion in the selection of appropriate integral representations. In this paper the indicated intervals of integration are semi-infinite. The finite interval conditions are applicable, however, because in each case the moment problem could be formulated with no error with a finite cutoff because the integrand becomes exponentially small.

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 to somewhat different extrapolations. The most natural procedure is to adopt an integral representation which 'builds in' all the information we have (asymptotic behaviour for this paper, though other kinds of information may be considered). The choice of the kernel is thus determined by the problem we are considering, and a general recipe for selecting the kernel cannot be given without information in addition to a finite set of Taylor coefficients. This cannot usually *uniquely* specify K, of course, but provides important guidance in selecting appropriate kernels. A possible criterion for selecting a *unique* kernel within a restricted family is discussed in Drabold and Jones (1991) and briefly described in section 3.3

When we have some expectations about the function f we wish to extrapolate, one chooses the kernel $K(u)$ which has the qualitative features expected of $f(u)$: the kernel can be chosen to be an intelligent ad-hoc guess for f. The rationale for this is: if we

guessed the kernel to be the *exact f*, we should obviously find that the positive, additive density would just be

$$\rho_K(\zeta) = \delta(\zeta - 1) \tag{7}$$

from equation (3). The Taylor coefficients of f 'correct' our preliminary guess for f. If our kernel has Taylor coefficients close to those of f, we find that $\rho(\zeta)$ is sharply localized near $\zeta = 1$.

Having fixed upon a kernel K, our extrapolation for $f(x)$ takes the form

$$f^{\text{extrap}}(x) = \int_{\mathcal{D}} d\zeta\, \rho_K(\zeta) K(x\zeta). \tag{8}$$

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, 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. An additional point is that there is nothing restricting the choice of kernel to the multiplicative form given in equation (3): kernels which are not functions of the product form $K(x\zeta)$ 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 $K(x-\zeta)$ 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). For the integral representation above equation (3) 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 (appendix 1 of Drabold and Jones 1991) 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 kernels for which the moment problem may be addressed. If the spectrum of the moment-test matrices is non-negative, we proceed to solve a discrete approximation to the linear maxent problem using the method of Bretthorst (1987), and then polish the Lagrange multipliers for the continuous problem using an improved version of the original Newton minimization method (Mead and Papanicolaou 1984). The power series is then easily extrapolated with a numerical quadrature of equation (8). This method is quite stable for up to about 12 coefficients, at which point the continuous polish can sometimes become unstable. Turek (1988) has recently suggested an improved numerical procedure using orthogonal polynomials instead of raw powers which alleviates this instability.

3. Examples

3.1. 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

$$\mathcal{H} = p^2/2 + x^2/2 + gx^8. \tag{9}$$

This problem has been studied by Bender *et al* (1987) using maxent and a particular integral representation of $E_0(g)$. These authors have used the Rayleigh-Schrödinger perturbation expansion for $E_0(g)$, which yields a divergent power series in the coupling constant:

$$E_0(g) \sim \tfrac{1}{2} + \sum_{n=1}^{\infty} (-1)^{n+1} A_n g^n \qquad g \to 0. \tag{10}$$

The coefficients grow exceedingly fast ($A_n \sim (3n)!$). Using five 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 behaviour of $E_0(g)$, namely that

$$E_0(g) \sim g^{1/5} \qquad g \to \infty \tag{11}$$

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

$$F_0(g) = [E_0(g) - \tfrac{1}{2}]/g \tag{12}$$

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

$$F_0(g) = \int_0^{\infty} d\zeta\, \rho_\alpha(\zeta)(1 + g\zeta)^{\alpha} \tag{13}$$

with $\alpha = -\frac{4}{5}$. Bender *et al* used the representation (equation (13)) with $\alpha = -1$. Following the procedure indicated, we obtain the results presented in figure 1 along with the result of Bender and coworkers and the exact result (Hioe *et al* 1976) The Lagrange multipliers are given in appendix 1. It is clear tha the choice $\alpha = -\frac{4}{5}$ produces a much

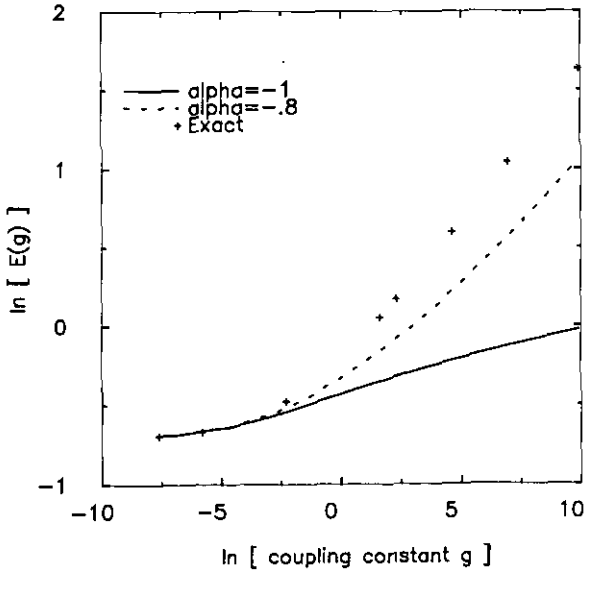

Figure 1. Octic oscillator ground state eigenvalue versus coupling constant g.

better fit than that of Bender *et al.* It is also apparent that the *prefactor* of the asymptotic behaviour (equation (11)) is different for the exact answer and our representation. It is not possible with the simple representation (equation (13)) to obtain the asymptotic behaviour exactly. We are presently investigating a broader class of kernels so that the prefactor can be exactly obtained. Note that for this badly divergent series, the inclusion of information concerning large-g behaviour improves the extrapolation, even for rather small g.

3.2. Spin susceptibility of the symmetric Anderson model

As another example of this method, we use the method outlined above to extrapolate the reduced spin susceptibility $\chi_s(u)$ of the symmetric Anderson model of magnetic impurities in metals. Here, $u = U/\pi\Delta$ and $U \ll (\gg)\Delta$ is the weak (strong) correlation regime. Here, we take the expansion of Yamada (1975) and Zlatic and Horvatic (1983):

$$\chi_s(u) = \sum_{n=0}^{\infty} C_n u^n. \tag{14}$$

The idea is to use the maxent method to estimate χ_s for the strong correlation regime from the first few perturbation coefficients obtained from the $u \to 0$ limit. To some extent this exercise is academic, since the exact solution to the problem is known from Bethe-ansatz techniques, and Zlatic and Horvatic have been able to develop recurrence relations on the expansion coefficients C_n which allow these coefficients to be computed to any order. Still the problem is of interest as an illustration of the approach and as useful input into the important asymmetric Anderson model (Horvatic and Zlatic 1985) for which no exact solution is known, but where an analogous approach to the one presented here is possible. In table 1 we present the first several C_n, obtained from the recurrence relation of Zlatic and Horvatic (1983).

Table 1. Expansion coefficients for the symmetric Anderson model.

n	C_n
0	1.0
1	1.0
2	0.532 598
3	0.195 593
4	5.501 868E − 2
5	1.256 072E − 2
6	2.414 885E − 3
7	4.011 442E − 4

As usual, our first task is to choose an appropriate kernel for our integral representation equation (3). We choose a simple exponential kernel

$$K(x\zeta) = \exp(x\zeta) \tag{15}$$

for the following reasons. (i) According to the notion that we should pick a kernel which is an educated guess for χ_s, we note that the Taylor expansion for χ_s has coefficients which resemble the expansion coefficients of e^u. (ii) On physical grounds it is clear that $\chi_s(u)$ is a rapidly increasing function of u. We may take the point of view that we are making a crude guess for χ_s, and we use the principle of maximum entropy to systematically improve our primitive guess (by using Taylor coefficients as

input) for χ_s. These considerations lead us to the integral representation:

$$\chi_s(u) = \int_0^\infty d\zeta\, \rho(\zeta)\, e^{u\zeta}. \tag{16}$$

We study this representation for three levels of input coefficients (five, seven and nine Taylor coefficients (C_n)), and present the results of the extrapolation in table 2. For comparison, the exact Bethe ansatz is also provided. We note that the integral (equation (16)) is convergent because ρ dies faster than $e^{u\zeta}$ diverges for $\zeta \to \infty$. The convergence to the exact result is remarkably rapid: the seven-coefficient maxent extrapolation is within 1% of the exact result for $u \leqslant 6$, the nine-coefficient extrapolation is valid to $u \approx 15$ to the same level of accuracy. Of course these results are *vastly* superior to a simple truncation of the series (equation (14)) for appreciable u. These results suggest that the method may be useful in the more difficult asymmetric case.

Table 2. Maxent extrapolations against Bethe ansatz solution for the symmetric Anderson model. Maxent extrapolations are from equation (16). The integer in parentheses is the number of Taylor coefficients used.

u	Bethe ansatz	Maxent (5)	Maxent (7)	Maxent (9)
1	2.798 65	2.798 67	2.798 65	2.798 65
3	24.83 45	24.8594	24.8348	24.8346
5	242.167	244.111	242.219	242.173
7	2 483.23	2 552.62	2 486.27	2 483.65
10	85 949	94 070	86 503	86 093

3.3. 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 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):

$$P(x) = \int_{-\infty}^{1} d\zeta\, \rho_\alpha(\zeta)(1 - x\zeta)^\alpha \tag{17}$$

where by the assumption that the pressure is singular near $x=1$, we have $\alpha<-1$. Because we expect the reduced pressure to have a simple pole, we would expect to formulate the problem with $\alpha=-2$. To our surprise, we find that the moment problem is ill-posed for $\alpha \gtrsim -10.5$ (the eigenvalues of the moment-test matrices are not all positive). We are still left with selecting an optimal kernel from the one-parameter (α) family of integral representations (equation (17)). Within some given category of kernels it is possible to consider a maxent criterion for choosing the *kernel as well as the density* ρ, by maximizing

$$-\int_{\mathscr{D}} \rho_K(\zeta) \log \rho_K(\zeta)\, d\zeta \tag{18}$$

over kernels K. While it is clear that this is not a substitute for using *a priori* information, it is a reasonable criterion for eliminating the ambiguity of selecting a kernel from a restricted category of kernels. We have explained this in detail in Drabold and Jones (1991). Here we merely repeat the result. Using the six known virial coefficients (Ree and Hoover 1967) we have found that the maximum entropy over the kernels K of equation (17) occurs at $\alpha \approx -17.5$, or

$$P^{\text{extrap}}(x) = \int_{-\infty}^{1} d\zeta\, \rho_{-17.5}(\zeta)(1-x\zeta)^{-17.5}. \tag{19}$$

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

The extrapolation resulting from equation (19) is in close agreement with numerical simulations (Ree and Hoover 1967) (figure 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 constructed from the six known virial coefficients. Those Padé extrapolations which most resemble Monte Carlo or molecular dynamics data are then compared with the simulation 'data'.

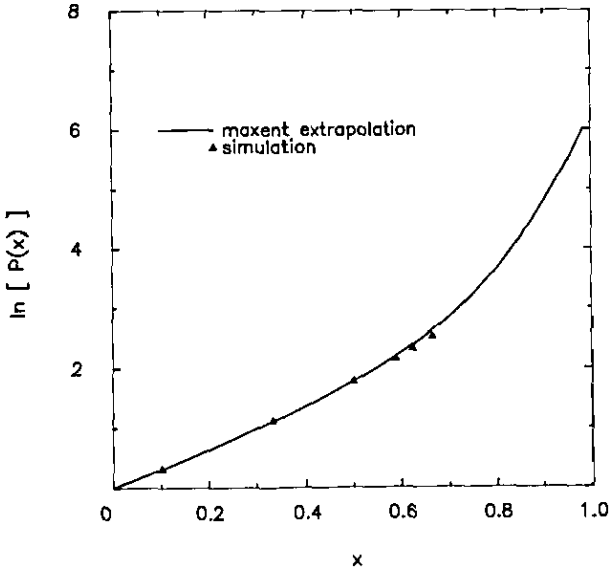

Figure 2. Maxent virial equation of state. The solid curve is the extrapolation from equation (10), the triangles are from computer experiments (Ree and Hoover 1967).

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.

We have recently received a preprint of Wang *et al* (1991), who have independently considered the hard sphere and hard disk problems. These authors also found a satisfactory maxent virial equation of state. However, they bypassed the question of selecting the kernel, offering no real justification for the form they chose, and comment that their representation was very sensitive to the values of the higher virial coefficients and they were not able to use the sixth virial coefficient at all. Our extrapolation suffered from none of these difficulties. As we have started above, the pointwise estimates tend to be fairly insensitive to the choice of the kernel, and our two extrapolations are reasonably similar, except for very high densities. For a complete discussion of our earlier work, see Drabold and Jones (1991).

4. Conclusion

In conclusion we have seen that maxent offers a useful approach to the extrapolation and analytic continuation problem. We have shown that the inclusion of information beyond the mere Taylor coefficients of an unknown function is crucial to obtaining an accurate fit. The method is in no way limited to divergent series: it is useful for any extrapolation problem for which a limited number of expansion coefficients is available. Future developments of the method should include further work on the *selection criterion for the kernel* and the generalization of the method beyond the multiplicative form of the kernel. A particularly interesting possibility involves linking the present method to the Gammel-Baker (for example Baker 1967) generalization of Padé approximants. It will be apparent to readers familiar with these techniques that there is much similarity between our integral representation and the Gammel-Baker approach. We are presently studying this connection.

Acknowledgments

We would like to thank Dr R Silver for suggesting the symmetric Anderson model spin susceptibility series as an example, and Dr G Baker for pointing out a resemblance between this method and the Gammel-Baker approximants.

Appendix 1. Lagrange multipliers for octic oscillator: $\alpha = -\frac{4}{5}$

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

l	λ_l
0	3.677 28
1	4.660 59E − 03
2	−1.341 74E − 06
3	1.788 86E − 10
4	−9.248 48E − 15
5	1.594 02E − 19

Appendix 2. Lagrange multipliers for hard sphere problem: $\alpha = -17.5$

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

l	λ_l
0	−0.881 413
1	3.788 254
2	1257.488
3	−19 566.264
4	100 422.993
5	−207 176.972
6	149 166.371

References

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

Anderson P W 1961 *Phys. Rev.* **124** 41

Baker G A 1967 *Phys. Rev.* **161** 434

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

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: Math. Gen.* **10** 1441

Drabold D A and Jones G L 1991 *Maximum Entropy and Bayesian Methods* ed W T Grandy Jr and L Schick (Dordrecht: Kluwer) p 79

Grandy W T Jr 1988 *Foundations of Statistical Mechanics* (Dordrecht: Reidel) ch 7

Gull S F 1989 *Maximum Entropy and Bayesian Methods* ed J Skilling (Dordrecht: Kluwer)

Gull S F and Daniell G J 1979 *Nature* **272** 686

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

Horvatic B and Zlatic V 1985 *J. Physique* **46** 1459

Jaynes E T 1968 *IEEE Trans. on System Science and Cybernetics* **SSC-4** 227; reprinted in Jaynes (1983)

—— 1983 *Papers on Probability, Statistics and Statistical Physics* (Dordrecht: Reidel)

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

Ree F H and Hoover W G 1967 *J. Chem. Phys.* **46** 4181

Silver R N, Sivia D S and Gubernatis J E 1990 *Phys. Rev.* B **41** 2380

Shannon C and Weaver W 1962 *The Mathematical Theory of Communication* (Urbana, IL: University of Illinois Press)

Skilling J 1989 *Maximum Entropy and Bayesian Methods* ed J Skilling (Dordrecht: Kluwer)

Turek I 1988 *J. Phys. C: Solid State Phys.* **21** 3251

Wang D, Mead L R and de Llano M 1991 *J. Math. Phys.* in press

Yamada 1975 *Prog. Theor. Phys.* **53** 970

Zlatic V and Horvatic B 1983 *Phys. Rev.* B **28** 6904