Title
Quantum Statistics Basis, Thermodynamic Analogies and the Degree of Confidence for Maximum Entropy Restoration and Estimation

Journal
Progress of Theretical Physics (Japan) Suppliment, No. 115

Authors
Soffer, Bernard H
Kikuchi, Ryoichi

Publication Date
1994

Peer reviewed

Quantum Statistics Basis, Thermodynamic Analogies and the Degree of Confidence for Maximum Entropy Restoration and Estimation

Bernard H. SOFFER and Ryoichi KIKUCHI*

Hughes Research Laboratories, Malibu, CA 90265, U. S. A.
**University of California at Los Angeles, Los Angeles, CA 90024-1595, U. S. A.*

(Received April 3, 1992)

The Maximum Entropy method, using physical statistics, chooses the most probable estimate consistent with limited measurements. Thermodynamic analogies and the degree of confidence are discussed.

Maximum Entropy (ME) estimation has been applied in various forms with various names to a wide variety of problems ranging from the depths of seismic spectral analysis, sonar and radar beam forming and filter formation, to astronomical imaging and beyond to economics. The particular techniques and theoretical points of view differ greatly among the several disciplines. Our very general interpretation, which includes these others as special cases, is based on two considerations. Any image, measured as signal, pattern or spectrum, whatever it represents, is necessarily a degraded version of the true object because real measurement systems have limited spatial and temporal bandwidth. The samples are finite and perhaps undersampled. Furthermore, noise cannot be ignored. Therefore, many different possible object patterns can produce the same measured image pattern. One way to resolve this ambiguity is to apply the ME method. In our interpretation, a probability is assigned to every possible object pattern and the most probable pattern is chosen as the estimated or restored object. Patterns are assigned probabilities based on the physics and statistics of the immediate problem. The entropy is understood to mean the logarithm of the probability, following Boltzmann. So, to find a maximum of the entropy is to find a maximum of the probability, subject to the measured image data constraints and any *a priori* bias. No new "principle of ME" or appeal to information theory is needed to justify the method, though they may enrich our understanding. Sometimes misunderstandings have arisen in the use of the information theoretic entropy of Shannon, $-f \log f$, and it has been used inappropriately. These considerations have been developed at length,[1)~3)] so only a brief summary will be given here. We develop the idea of ME in an analogy to the well-known statistical mechanical principle of the minimization of free energy, and derive some useful benefits in the consideration of fluctuations or noise. An outline for the calculation of the degree of confidence in the ME estimate is presented.

The two entropy expressions commonly used by different groups in solving ME problems, $\log B$ and $-B \log B$, where B is the brightness, power, intensity, or their spectral counterparts, can be derived as limiting special cases of a more general entropy expression based on the underlying properties of the source *and the measuring processing*. For photon or electromagnetic signaling or imaging, the Bose-Einstein statistics, and for electrons, the Fermi-Dirac statistics, are employed. The n quantum mechanical particles comprising the intensity are distributed over z degrees of freedom as calculated by these statistics. The number of degrees of freedom can be understood as the ratio of the space-time size of the detection volume to the coherence volume of the particles. The entropy to be maximized is the logarithm of probability as given by the physical statistics of the problem, following the original meaning of entropy. The entropies $\log B$ and $-B \log B$ result in the limit $n \gg z \gg 1$ and $n \ll z$, respectively. When n is interpreted as an average over an ensemble, we find in addition, for $n \gg z=1$, the Burg form $\log B$ results. The distribution of intensity for this case is exponential, or expressed in another way, Gaussian in complex amplitude. Shannon's entropy expression is a special case appropriate for his special interest, the $z=1$ (sampled at the Shannon rate) Gaussian case of equal *a priori* probabilities.

We now will develop a thermodynamic analogy between the ME method and the principle of minimization of free energy. First, we describe the method itself. Let us consider an object pattern, and designate it by {Ob}. The reasoning presented in Kikuchi and Soffer (Ref. 1)) shows that there are a number of different photon states that can make the same pattern {Ob}. We write this number, the degeneracy, as $P(\{\mathrm{Ob}\})$. If we accept the postulate that each state for the pattern is equally probable, $P(\{\mathrm{Ob}\})$ is the relative probability that the pattern {Ob} appears. This postulate is analogous to a basic property of quantum mechanics. [However, in context of image processing, Frieden[4] recently proposed a generalization in which *a priori* probabilities are assigned to different patterns. His treatment can be interpreted as a modification of the discussions presented here.] Following Boltzmann, we will call the logarithm of $P(\{\mathrm{Ob}\})$ the entropy S:

$$S(\{\mathrm{Ob}\}) = \log P(\{\mathrm{Ob}\})\,. \tag{1}$$

Our goal is to find the most probable pattern {Ob} under the constraint that the image made from {Ob} is the observed image {I}. In the actual computation, we consider fluctuations of the calculated image from the given image:

$$E(\{\mathrm{Ob}\}, \{\mathrm{I}\}) = \sum_k \Big(I_k - \sum_i \mathrm{PSF}_{k,i} * \mathrm{Ob}_i\Big)^2\,, \tag{2}$$

where $\mathrm{PSF}_{k,i}$ is the point spread function that transforms object space into image space, $_k$ and $_i$ represent (possibly two dimensional) image and object coordinates, respectively, and * denotes convolution. We require that E in Eq. (2) takes a certain value E_0. Therefore, we maximize

$$S_c(\{\mathrm{Ob}\}) \equiv S_{\mathrm{constrained}}(\{\mathrm{Ob}\}) \equiv S(\{\mathrm{Ob}\}) - \beta \cdot E(\{\mathrm{Ob}\}, \{\mathrm{I}\})\,, \tag{3}$$

where β is a Lagrange multiplier. This maximization procedure can be interpreted as analogous to the method of finding the equilibrium state in the microcanonical ensemble treatment of statistical mechanics. In the microcanonical ensemble treatment, the system is isolated from the rest of the universe so that the energy is fixed.

The equilibrium state is the state of a maximum entropy constrained by a fixed given energy. Thus, the treatment given in Eq. (3) is exactly the same as the microcanonical treatment with E interpreted as the energy.

It is known that the microcanonical ensemble treatment is equivalent to the canonical ensemble treatment. In the latter, we place the system in contact with a heat bath of temperature T and allow slow exchange of energy between the system and the heat bath. The energy of the system fluctuates and is not a fixed quantity. The equilibrium state is calculated not as a maximum of the entropy, but as a minimum of a free energy. Since the canonical ensemble point of view is taken in most formulations of statistical mechanics, it is helpful to illustrate the relation between the microcanonical and canonical treatments. From the latter point of view, what we maximize is the following constrained probability function:

$$P_c(\{\mathrm{Ob}\}, \{\mathrm{I}\}) \equiv P_{\mathrm{constrained}}(\{\mathrm{Ob}\}, \{\mathrm{I}\}) = \exp[S_c(\{\mathrm{Ob}\}, \{\mathrm{I}\})]$$
$$= \exp[S(\{\mathrm{Ob}\})] \cdot \exp[-\beta \cdot E(\{\mathrm{Ob}\}, \{\mathrm{I}\})] . \quad (4)$$

In analogy with statistical mechanics, we can interpret this maximization procedure as follows. When β is the reciprocal of the temperature of the heat bath, the $\exp[-\beta \cdot E(\{\mathrm{Ob}, \mathrm{I}\})]$ factor represents the relative probability that the object is made from one of the $P(\{\mathrm{Ob}\})$ possible photon states corresponding to the pattern {Ob} when the image {I} is given. The $\exp[S(\{\mathrm{Ob}\})] = P(\{\mathrm{Ob}\})$ factor is the number of ways that the object pattern {Ob} appears, as was stated in Eq. (1). Therefore, the product $P_c(\{\mathrm{Ob}\}, \{\mathrm{I}\})$ is the relative probability that the object pattern {Ob} is made from *any one* of the $P_c(\{\mathrm{Ob}\}, \{\mathrm{I}\})$ possible photon states, and hence is *the relative* probability of finding the pattern {Ob}. Following statistical mechanics, we may define the free energy F as

$$\beta \cdot F(\{\mathrm{Ob}\}, \{\mathrm{I}\}) = \beta \cdot E(\{\mathrm{Ob}\}, \{\mathrm{I}\}) - S(\{\mathrm{Ob}\}) . \quad (5)$$

Then we can write the P_c function as

$$P_c(\{\mathrm{Ob}\}, \{\mathrm{I}\}) = \exp[-\beta \cdot F(\{\mathrm{Ob}\}, \{\mathrm{I}\})] , \quad (6)$$

and the process of finding the most probable pattern {Ob} can be interpreted as finding the minimum free energy state. On the basis of the analogy between Eq. (5) and thermodynamics, it is justifiable to call $E(\{\mathrm{Ob}\}, \{\mathrm{I}\})$ the energy.

There is no unique way of defining the functional form $E(\{\mathrm{Ob}\}, \{\mathrm{I}\})$. Whatever form we may choose, the interpretation holds that $\exp[-\beta \cdot E(\{\mathrm{Ob}\}, \{\mathrm{I}\})]$ is the relative probability of {Ob}. The required properties of E are (i) it becomes zero when I_k is equal to $\sum_i \mathrm{PSF}_{k,i} * \mathrm{Ob}_i$, (ii) it is positive and becomes larger when the difference between the two increases. The square expression in Eq. (2) was chosen because it resembles the potential energy expression of a simple harmonic oscillator, however, it is possible to choose a fourth power of the expression, for example, without violating the requirements (i) and (ii).

In the two-dimensional example we now turn to, the energy constraint is written entirely in the spaces of the object and the image. In this example we use the log B formulation of ME appropriate for Gaussian amplitude or exponential intensity statistics. We chose a convolutional point spread function of the measurement to be of the form $\mathrm{sin}c^2(x)\cdot\mathrm{sin}c^2(y)$ representing diffraction spreading by a limiting square aperture. The scale of diffraction spreading was chosen to be a factor of 4 in each dimension. That is, the number of independent variables needed to completely describe the image was $4\times 4=16$ times less than required to describe the object from which it was mapped. The image can be completely and uniquely represented on a grid 4×4 times coarser than the object.

The example was constructed from the binary black and white alphabetical object shown in Fig. 1(a), constructed on a 20×20 grid. The diffraction image is shown in Fig. 1(b). The 400-dimensional ME estimate of the object is shown in Fig. 1(c) for the case $\beta=10^7$. The convergence was so slow and erratic at this high value of β that Fig. 1(c) shows incompletely converged results; however, super resolution was achieved.

It is interesting to note that the seemingly unrelated method of simulated annealing[5] is also a free energy minimization method, with the Monte Carlo technique automatically providing the entropy contribution to the free energy via the random number portion of the algorithm.

When we introduce the concept of the free energy, we can use the following thermodynamic equation as a consistency check of computations:

$$\frac{\partial}{\partial\beta}[\beta\cdot F(\{\mathrm{Ob}\},\{\mathrm{I}\})]=E(\{\mathrm{Ob}\},\{\mathrm{I}\})\,. \tag{7}$$

When E is defined in the quadratic form of Eq. (2), the left-hand side derivative in Eq. (7) calculates the fluctuation of the object pattern. Noting Eq. (2), this relation can be interpreted as corresponding to the well-known relation for the energy fluctuation in thermodynamics:

$$kT^2c_v=\frac{\partial^2(\beta F)}{\partial\beta^2}=\langle(E-\langle E\rangle)^2\rangle\,, \tag{8}$$

where k is the Boltzmann constant and c_v is the specific heat.

The degree of confidence in the ME estimate can be derived in a general manner by expanding the object probability distribution near its calculated maximum. Only the maximum, not the entire multidimensional distribution, is calculated in the ME method. The expansion is done to second order terms yielding a multivariate Gaussian. A principal axis transformation (i.e., a linear orthogonal transformation of the variables) is made to a new set of vari-

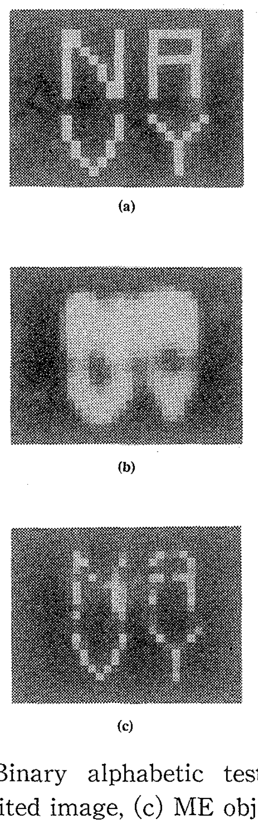

Fig. 1. (a) Binary alphabetic test object, (b) diffraction limited image, (c) ME object estimate.

ables that are stochastically independent and normally distributed. The algebraic form thus obtained is that of a sum of squares of these stochastically independent variables and, therefore, has a chi-square distribution with a number of *statistical degrees of freedom* (not the physical statistical "z" defined earlier) equal to the number of object variables. Tables of the P-fractal of the cumulative chi-square distribution for that number of degrees of freedom give the desired *statistical degree of confidence* estimates. The set of variances thus derived are projected back to the space of the object estimates as confidence regions. A one-dimensional example of the super resolution of two delta function objects was studied to test the method as a function of the super resolution demanded and the fluctuation or noise temperatures assumed. Preliminary results seem intuitively reasonable: the more super resolution demanded or the higher the noise temperature, the smaller the degree of confidence in the estimate. However, further generalization is elusive, as the results of this method are strongly dependent on the particular object, requiring an expansion about the maximum of the particular ME solution at hand. This method is computationally demanding, but provides a useful measure of the degree of confidence in the ME estimate.

References

1) R. Kikuchi and B. H. Soffer, J. Opt. Soc. **A67** (1977), 1656.
2) B. H. Soffer and R. Kikuchi, *Topical Meeting on Quantum-Limited Imaging and Image Processing, Optical Society of America* (Honolulu, March, 1986).
3) B. H. Soffer and R. Kikuchi, Final Report No: AFOSR-TR-81-0324, AD-A097 357/8 (February 1981).
4) B. R. Frieden, J. Opt. Soc. **A73** (1983), 927.
5) S. Kirkpatrick, C. D. Gelatt and M. P. Vecchi, Science **220** (1983), 671.