

Acta Cryst. (1984). **A40**, 163–164

On information and complementarity in crystal structure determination. By S. STEENSTRUP* and S. W. WILKINS, CSIRO, Division of Chemical Physics, PO Box 160, Clayton, Victoria 3168, Australia

(Received 13 June 1983; accepted 5 October 1983)

Abstract

The roles of prior information, event space and choice of constraint functions in information-theoretic approaches to the crystallographic inversion problem are discussed, with a view to relating recent articles on the inversion problem. In particular, the important distinction between the general expression for the entropy of a probability distribution function and specific derived forms for entropy involving, for example, mean values of the corresponding random variables, is pointed out and illustrated.

Information theory can provide a very general and powerful approach to problems of statistical inference (Jaynes, 1982). However, great power must be tempered by great care in order for any good to result. In this note it is our purpose to focus attention on some of the areas in the information-theoretic approach in which such care is required, with a particular view to elucidating some of the assumptions underlying recent articles (Britten & Collins, 1982; Narayan & Nityananda, 1982; Piro, 1983; Wilkins, Varghese & Lehmann, 1983) on the crystallographic inversion problem.

The key concept and starting point for an information-theoretic approach to a problem in statistical inference is the concept of a unique and consistent measure of the information content, I , in a probability distribution $P(e)$ relative to a given prior $P^0(e)$ (Shannon & Weaver, 1949; Kullback, 1959; Levine, 1980; Jaynes, 1982):

$$I = \sum_{e \in E} P(e) \log \{P(e)/P^0(e)\},$$

where $P(e)$ is the probability distribution of the random variables e , taking values in the event space E . {If the prior P^0 is not normalizable, e.g. a uniform prior on $[0, \infty]$, then $I' = -H = \sum P(e) \log P(e)$ is used; the term 'entropy' is usually used for H .} The precise specification both of the event space E and the prior P^0 is very important and should ideally be such as to incorporate all prior knowledge about the system as efficiently as possible.

In crystallography there are two variables of interest: (i) the electron density $\rho(\mathbf{r})$ in the unit cell; (ii) the set of structure factors $\{F_{\mathbf{H}}\}$.

Working with functions of a continuous variable is often impractical and so, for convenience, the unit cell is sometimes subdivided into N identical pixels, whereby the electron density is turned into a vector ρ with ρ_j the mean electron density in pixel j .

The information-theoretic approach then consists in determining the probability distribution function P which minimizes the information I in a way that is consistent with

the measured data and the prior information. Once this distribution function has been determined an estimate for ρ (or $\{F_{\mathbf{H}}\}$) may be obtained as the mean value with respect to P .

For calculational purposes it is often more convenient to recast the minimization of I with respect to $P(\rho)$ [or $P(\{F_{\mathbf{H}}\})$] into an equivalent minimization (maximization) problem with respect to the mean values, or some other quantity of interest. The precise form of this reformulated problem, however, depends on (1) the event space, (2) the prior information and (3) the form of the constraints—a dependence not often explicitly discussed.

In Table 1 a series of reformulated cases is listed, based largely on results given by Levine (1980). The first column in Table 1 lists the kind of constraints: only linear and quadratic cases are considered. The second column shows the event space, the third column the prior information and the fourth the equivalent form of the information. It must be stressed that this equivalent form is not a general information measure. The point being that it is only under particular assumptions, such as those shown in columns 1, 2 and 3, that the minimization of I and the minimization (maximization) of the forms shown in column 4 are equivalent.

Various possible correspondences may be made between actual physical situations and the entries in the table. For the crystallographic inversion problem, for example, linear constraints could be taken to correspond to the case where a subset of all structure factors is known (including phases), while the case of quadratic constraints could correspond to the phaseless problem, i.e. a subset of intensities is measured.

The equivalence between minimizing the information and maximizing the integral of the logarithm of the power spectrum was first derived by Burg (1967) for the special case where autocorrelations of a discretely sampled time series are the measured quantities, while the quantity to be estimated is the power spectrum. The form 'integral of the logarithm of a quantity' has subsequently (misleadingly) been given a more fundamental significance than it actually deserves and termed 'Burg entropy'. Note that for quadratic constraints the 'Burg form' is only obtained when the variable is unrestricted, i.e. runs from $-\infty$ to $+\infty$; any restriction, like positivity, will give rise to a different equivalent form, which as yet has not been worked out.

The difference between the time-series and crystallographic cases is that, in the latter, the variables are for physical reasons restricted to a subspace determined in real space by the condition $\rho(\mathbf{r}) > 0$, or in reciprocal space by the condition that all the Karle-Hauptmann determinants must be positive. These two conditions are well known to be equivalent (Karle & Hauptmann, 1950; Steenstrup, 1984).

Britten & Collins (1982) and Piro (1983) use reciprocal space as the primary event space with quadratic constraints

* Permanent address: Physics Laboratory II, Universitetsparken 5, DK-2100 Copenhagen 0, Denmark.

Table 1. Relations between assumed input information (columns 1, 2 and 3) and derived expressions for the information function I (column 4)

Notation: R set of real numbers, R_+ set of positive real numbers, N set of integers, x set of random variables, e.g. electron densities, or set of structure factors.

Data constraints	Event space E	Prior probability p°	Equivalent Information I /entropy H
Linear	$x \in R_+^N$	—	$H = \sum_{i=1}^N \log \langle x_i \rangle$
$A_r = \left\langle \sum_{i=1}^N a_{ri} x_i \right\rangle$	$x \in N^N$	—	$H = \sum_{i=1}^N (1 + \langle x_i \rangle) \log (1 + \langle x_i \rangle) - \langle x_i \rangle \log \langle x_i \rangle$
$r = 1, \dots, m$	$x_i \in \{0, \dots, n\}$ $\sum_{i=1}^N x_i = n$	$N^{-n} \frac{n!}{x_1! \dots x_N!}$	$I = n \log N + \sum_{i=1}^N p_i \log p_i$ $p_i = \langle x_i \rangle / n$
Quadratic	$x \in R^N$	—	$H = \int_{-\pi}^{\pi} \log P(\omega) d\omega$ $P(\omega) = \sum_{r=1}^m A_r e^{i\omega r}$
$A_r = \left\langle \sum_{i=1}^N x_i x_{r+i} \right\rangle$			

— i.e. assumptions corresponding to the fourth row — to obtain a multivariate Gaussian probability distribution function in the structure factors and obtain the normalization by extending integration over all space ignoring the positivity constraint on $\rho(r)$. The ‘complementary’ result to that of Burg is thereby obtained, namely that the autocorrelations between structure factors in reciprocal space yield the corresponding ‘power spectrum’ in direct space, i.e. electron density squared.

The equivalence between the maximum determinant method and the maximum of the ‘integral of logarithm of electron density’ form is established by Narayan & Nityananda (1982). However, it is not explicitly stated whether the identification of this form with entropy is based on the assumptions corresponding to those in the first row in our table or to those in the fourth.

We believe that for the crystallographic inversion problem the assumptions corresponding to neither the first row nor the fourth row are really appropriate. The nature of the problem indicates that:

(i) only a finite number of distinct sets ρ corresponds to distinguishable measurements (structure factors, intensities), by arguments of ‘complementarity’;

(ii) the total number of electrons in the unit cell is fixed and usually known; and

(iii) the electron density is bounded.

The assumptions in the third row are therefore, in our view,

the appropriate ones and are the ones used by Wilkins *et al.* (1983). With these assumptions, however, there is then no direct relation between the maximum determinant method and the maximum entropy principle.

One of us (SS) thanks the CSIRO for hospitality and the Danish Science Foundation for financial help.

References

- BRITTEN, P. L. & COLLINS, D. M. (1982). *Acta Cryst.* **A38**, 129–132.
 BURG, J. P. (1967). *Maximum-Entropy Spectral Analysis*. Paper presented at 37th Annual International SEG Meeting, Oklahoma City. Reprinted in *Modern Spectrum Analysis*, edited by D. G. CHILDERS. New York: Wiley (1978).
 JAYNES, E. T. (1982). *Papers on Probability, Statistics, and Statistical Physics*. Dordrecht-Holland: Reidel.
 KARLE, J. & HAUPTMANN, H. (1950). *Acta Cryst.* **3**, 181–187.
 KULLBACK, S. (1959). *Information Theory and Statistics*. New York: Wiley.
 LEVINE, R. D. (1980). *J. Phys. A*, **13**, 91–108.
 NARAYAN, R. & NITYANANDA, R. (1982). *Acta Cryst.* **A38**, 122–128.
 PIRO, O. E. (1983). *Acta Cryst.* **A39**, 61–68.
 SHANNON, C. E. & WEAVER, W. (1949). *The Mathematical Theory of Communication*. The Univ. of Illinois Press, Urbana.
 STEENSTRUP, S. (1984). *Acta Cryst.* **A40**, 66–67.
 WILKINS, S. W., VARGHESE, J. & LEHMANN, M. S. (1983). *Acta Cryst.* **A39**, 47–60.