Acta Cryst. (1983). A**39**, 47–60

Statistical Geometry. I. A Self-Consistent Approach to the Crystallographic Inversion Problem Based on Information Theory

BY STEPHEN W. WILKINS*

CSIRO, Division of Chemical Physics, PO Box 160, *Clayton, Victoria, Australia* 3168 *and Institut Laue–Langevin, BP*156 *Centre de Tri, Grenoble Cedex* 38042 *France*

JOSEPH N. VARGHESE

CSIRO, Division of Protein Chemistry, Royal Parade, Parkville, Victoria, Australia 3052

AND MOGENS S. LEHMANN

*Institut Laue–Langevin, BP*156 *Centre de Tri, Grenoble Cedex* 38042 *France*

(*Received* 13 *October* 1981; *accepted* 1 *July* 1982)

Abstract

The problem of inverting crystallographic diffraction data to obtain structural information is examined within the maximum-entropy formulation of information theory. The principal features of the present method (termed statistical geometry) are: (i) all predictions of the method are *consistent* with the given information (constraints) and *least biased* with respect to missing information, (ii) the adoption of *weak* (typically non-linear) constraints for incorporating the major part of the structural information guarantees that a solution exists in practice and leads to filtering of the structure maps consistent with the accuracy of the data, (iii) general conditions are established which lead to unique solutions for the structure map, (iv) atomicity is not a prerequisite, (v) other methods of crystallographic inversion may be incorporated *via* the adoption of appropriate constraint relations, and (vi) the task of numerical solution is roughly linear in the number of reflexions and in the number of pixels in the structure, and involves only straightforward numerical techniques. These features suggest that the method is especially well suited to problems such as the structure determination of biological macromolecules, and the determination of high-resolution electron-density maps, although it manifestly provides a general framework for treating a wide class of image-processing problems.

1. Introduction

An inversion procedure is required whenever the available measurements depend on the value of a function of interest at more than one point. In cases where the measurements are incomplete and noisy, the problem essentially becomes one of statistical inference based on partial information. For cases where the number of possible states is extremely large, a very powerful technique for inversion involves the use of the information-theoretic procedure of maximum entropy combined with the adoption of *weak* constraints for the bulk of the data. In such cases the formal structure of the method closely resembles that of equilibrium statistical mechanics and, in the context of the crystallographic inversion problem, is here termed *statistical geometry* (SG).

This work and those following in the series aim at establishing both a conceptual framework and practical numerical methods with which the usual crystallographic inversion problem of determining structural information from kinematical diffraction data can be pursued. In the present context, a *state refers to a particular assignment of scattering density throughout the unit cell.* For high resolution and large-molecule structures, the number of significantly different possible states (structures) becomes astronomical. It is in cases such as these that the information-theoretic approach is capable of providing *optimal* predictions of the structure starting from a partially known or unknown structure. The principal prerequisite of the method is that one's *a priori* and experimental information about the system be expressible in suitable mathematical form. One may consider the present general approach to *subsume all currently practised numerical approaches to the crystallographic inversion problem* (*e.g.* Karle & Hauptman, 1950; Sayre, 1952) as special cases in that they may be incorporated in the method *via* the adoption of appropriate constraint relations.

* Permanent address: CSIRO, Australia.

0567-7394/83/010047-14$01.50

The basic outline of the paper is as follows: In § 2 the fundamental expressions of the general information-theoretic approach are presented (§ 2.1) leading to the presentation of the particular expressions which obtain when the state **x** of the system corresponds to a probability distribution **p** over microstates (§ 2.2). The reader who is primarily interested in the practical details of the present method may proceed directly from the present section to § 2.3 where the fundamental equations, (13*b*), of the method are presented and the existence and uniqueness of solutions to them is discussed. A variational approach to the determination of the Lagrange multiplier vector such that the chosen constraints are satisfied is outlined in § 3 and Appendix 2. § 4 focuses consideration on the particular case of the crystallographic-inversion problem and presents illustrative constraint relations which may serve to invert diffraction data statistically in the form of structure factors with phases which are either: (i) partially known or (ii) essentially unknown. The task of numerically solving the fundamental equations is briefly discussed in § 4.5 and some illustrative results for a simple one-dimensional structure are given in § 4.6. Finally, in § 5 a discussion of the potential advantages of the present approach to crystallographic inversion is presented.

The results presented in this paper provide a very general basis for the future development of mathematical and numerical methods for treating the crystallographic inversion problem which will be described in later papers in this series. Throughout these works, a guiding aim will be to try to establish mathematical methods which are numerically tractable, even when the number of reflexions is extremely large, such as for biological macromolecules.

2. Information theory

Following the lucid work of Jaynes (1957; 1963; 1979), we here give a suitably general statement of the problem to be treated by information theory as follows: Suppose that the quantity (random variable, message, state) **x** can take on the values (states) $(x_1, x_2, \ldots, x_N)$ where N can be finite or infinite, and that the expectation values of several functions $f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_M(\mathbf{x})$ are given by $C_1, \ldots, C_M$, where $M < N$ (typically $M \ll N$). The problem then is to find the probability assignment $P(\mathbf{x})$ which satisfies the given data:

$$\sum_{\mathbf{x}} P(\mathbf{x}) = 1 \tag{1}$$

and

$$\sum_{\mathbf{x}} P(\mathbf{x})\, f_r(\mathbf{x}) = C_r \quad \text{for } r = 1, \ldots, M, \tag{2}$$

and is the *least-biased* estimate possible based on the given information; that is, it is *maximally non-committal* with regard to missing information. The great conceptual advance provided by information theory resides in the discovery by Shannon (1948) that there is a unique and consistent measure of the amount of 'ignorance' (uncertainty, entropy) in a discrete probability distribution and that it is given by

$$S[\mathbf{x}] = -\sum_{\mathbf{x}} P(\mathbf{x}) \ln P(\mathbf{x}) \tag{3}$$

and is immediately seen to correspond to the Boltzmann expression for entropy which arises in statistical mechanics. Shannon showed that it is only this function $S[\mathbf{x}]$ which is (i) *positive*, (ii) *increases with increasing uncertainty* and (iii) is *additive for independent sources of uncertainty*.

To obtain the most *unbiased probability assignment* we must maximize (3) subject to the constraints (1) and (2). The mathematical technique for achieving this is described in § 2.1 below. The function (3) was actually introduced by Shannon in connection with the problem in communications theory of encoding and decoding messages *via* a noisy line, and serves as a very useful measure for comparing the relative efficiency of different coding schemes. The problem of coding structural information in crystallography has been discussed from this viewpoint by Gassmann (1977). Secondly, the great power of the information-theoretic approach to problems of statistical inference resides in the interpretation that the *maximum-entropy distribution describes our state of knowledge in a way which is maximally non-committal with regard to missing information*. As more information is discovered, it may be incorporated, *via* (2), into a new distribution and so the claim repeated. At each stage the information-theoretic predictions are *totally consistent* with our 'state of knowledge' about the system, and thus the method is internally consistent in the sense that it can only lead to 'false' predictions if our input information is false or internally inconsistent. However, the method need not lead to unique solutions or to any solutions, and these matters are discussed in §§ 2.3 and 4.5.

In loose terms, the maximization of $S[\mathbf{x}]$ subject to the constraints (1) and (2) leads to the result that $P(\mathbf{x})$ is the 'smoothest' distribution consistent with the given information [see Wernecke (1977), and § 4.3.3 for a discussion of this property]. Thus, maximizing (3) may be viewed as a form of filtering operation. However, the maximum-entropy approach is much more powerful than just a mathematical smoothing operation, and is perhaps best viewed as a *calculus of inductive reasoning* (see *e.g.* Tribus, 1969; Levine & Tribus, 1979).

2.1. *The distribution of maximal entropy*

The distribution $P(\mathbf{x})$, which is of maximal entropy, subject to the constraints of normalization and to the M

(not necessarily linear) data constraints (2) is readily determined, using the Lagrange undetermined multipliers procedure, to be of the form (see *e.g.* Jaynes, 1957; Levine, 1980)

$$P(\mathbf{x}) = g(\mathbf{x}) \exp\left\{-\sum_{r=0}^{M} \lambda_r f_r(\mathbf{x})\right\}, \tag{4}$$

where $g(\mathbf{x})$ is the degeneracy factor and we have incorporated (1) in (2) by including the case $r = 0$. The $M + 1$ (Lagrange) multipliers, $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_M)$ and λ_0, are determined by the $M + 1$ conditions (2) and (3). The resulting set of equations for λ_0 and $\boldsymbol{\lambda}$

$$\sum_{\mathbf{x}} [f_r(\mathbf{x}) - C_r] P(\mathbf{x}) = 0, \quad \text{for } r = 0, 1, \ldots, M \tag{5}$$

is in general coupled and highly non-linear in $\boldsymbol{\lambda}$ [even when the $f_r(\mathbf{x})$ are linear in $\mathbf{x}$], which in general will have many solutions or none (see also § 4.4).

Formally, the Lagrange multipliers can be computed from the entropy of the distribution $P(\mathbf{x})$ (Levine, 1980)

$$S[\mathbf{x}] = \sum_{r=0}^{M} \lambda_r C_r, \tag{6}$$

since

$$\lambda_r = \frac{\partial S[\mathbf{x}]}{\partial C_r} \quad \text{for } r = 0, 1, \ldots, M. \tag{7}$$

For other properties of the maximal-entropy distribution, the reader is referred in particular to the works of: Shannon (1948); Khinchin (1957); Kikuchi & Soffer (1977); Jaynes (1957, 1963, 1979); Tribus (1969) and those of Levine and co-workers cited in the references.

2.2 *Expression for the structural entropy in terms of the distribution over microstates*

Thus far, our discussion of the state $\mathbf{x}$ has been quite general. We now consider the situation where $\mathbf{x}$ defines a structure and the different possible structures (macrostates) may be considered to arise by distribution of n identical units of scattering over N identical subcells (pixels, microstates) of the structural volume. One might like to think of these units of scattering as photons, although our approach is essentially geometrical in character and the precise physical nature of the scattering need not be specified and the units of scattering may be made arbitrarily small. If we now conduct an experiment in which the n units of scattering are distributed over the N pixels, then the macrostate of the system (*i.e.* the structure) may be written as $\mathbf{x} = \mathbf{n}$, where $x_j = n_j$ is the number of units of scattering in the jth pixel. Let us then assume that the experiment is repeated many times and that

$$\langle n_j \rangle = np_j \tag{8}$$

is the mean value of the number of scattering units in the jth cell. Then it may readily be shown (see *e.g.* Levine, 1980, p. 99) that maximizing the entropy (3) subject to the N constraints (8) leads to the result

$$S(\mathbf{p}) = -n \sum_{j=1}^{N} p_j \ln p_j \tag{9}$$

for the structural entropy, and is a special case of (3). We have actually expressed S in terms of the constraint values, $\mathbf{p}$ [*cf.* (6)], which are the probability values for the distribution of scattering units over the N microstates or pixels. The relationship between macro- and microstates is the multinomial distribution

$$P(\mathbf{n}) = g(\mathbf{n}) \prod_{j=1}^{N} p_j^{n_j} = n! \prod_{j=1}^{N} p_j^{n_j}/n_j!. \tag{10}$$

The important simplification provided by (9) is that we do not need to enumerate all possible macrostates each time we wish to determine a structure by maximizing S subject to some given structural information, but may proceed directly from (9). Such an approach is an example of sequential inference using maximum entropy (Levine, 1980, p. 98).

Expressions other than (9) (which we shall henceforth denote by S_1) for the entropy in terms of microstates have been used in processing diffraction and spectral data. More particularly, the form $S_2 = -\sum_{j=1}^{N} \ln p_j$ has many adherents (*e.g.* Ables, 1974; Wernecke & D'Addario, 1977) and arose in a field quite different from the present one, namely from consideration of the power spectrum of a Gaussian random process or time series (Shannon, 1948, 1949).

Recently, Kikuchi & Soffer (1977) have attempted to clarify the appropriate choice for the entropy function in different cases by taking a quantum optical approach to the problem and showed that both the S_1 and S_2 expressions may be derived as limiting cases of a single one, and the choice between them depends on a parameter n/z, n being the number of photons received and z the number of degrees of freedom per pixel. They give a simple expression for z which contains a component due to the spatial uncertainty as to the origin of the photon within a pixel and to a temporal uncertainty as to the time of arrival of the photon, and conclude that in general bright radio objects have a large n/z (and therefore should be processed with S_2) while less-bright objects at shorter wavelengths have a small n/z (and therefore should be processed with S_1). From their results it would appear that z is very large for, say, the X-ray diffraction case and so S_1 is also favoured there.

Examination of the mathematical role of S in the MEM method shows it to reside in the following important mathematical properties, namely: (i) preservation of positivity among the p_j and (ii) strict concavity which tends to bias $\mathbf{p}$ towards a uniform distribution and so provide a smooth (see § 4.3.3 and also Wernecke, 1977, Appendix *B*) interpolation when the pixel size becomes smaller than the Shannon–Nyquist interval (see Shannon, 1949) of the data. These properties are common to both S_1 and S_2, and when the data set is large in relation to the resolution sought in the structure determination, the structure is largely determined by the data rather than the interpolation method, and so negligible differences in predictions would be expected (see also Wernecke, 1977, Appendix *B*). On the other hand, when data is sparse or super-resolution is sought, some small, albeit significant, differences in structure may occur depending on the choice for S.

In summary, our basis for adopting (9) in the work to follow is (i) the information-theoretic basis of (3) and (ii) the strictly geometrical argument leading from (3) to (9).

2.3. *The maximum entropy distribution over microstates*

We may now reformulate the maximum entropy determination of $\mathbf{p}$ as follows.

Maximize (9) subject to the constraints

$$\sum_{j=1}^{N} p_j = 1 \qquad (11a)$$

and

$$f_r(\mathbf{p}) = C_r, \quad r = 1, \ldots, M. \qquad (11b)$$

It may be noted that an alternative form of (11*b*) often encountered in the literature, especially in the case of linear constraints, is

$$\langle h_r(\mathbf{p}) \rangle \equiv \sum_{j=1}^{N} p_j h_{rj}(\mathbf{p}) = C_r, \qquad (11b')$$

where for linear constraints h_{rj} is independent of $\mathbf{p}$. It may be noted that (11*b*′) leads to a unique $f_r(\mathbf{p})$, whereas the converse is not true.

Mathematically, the constrained maximization problem formulated above may be achieved by using the method of Lagrange undetermined multipliers and involves unconstrained maximization of

$$Q(\mathbf{p}; \lambda_0, \boldsymbol{\lambda}) = -\sum_j p_j \ln p_j - \lambda_0 \sum_{j=1}^{N} p_j - \boldsymbol{\lambda}.\mathbf{f}(\mathbf{p}) \qquad (12)$$

w.r.t. the p_j (where for convenience we have set the constant $n = 1$) leading to the N equations

$$\frac{\partial Q}{\partial p_j} = -n[\ln p_j + 1] - \lambda_0 - \boldsymbol{\lambda}.\frac{\partial \mathbf{f}(\mathbf{p})}{\partial p_j} = 0 \qquad (13a)$$

$$p_j = \exp\{-\lambda_0 - \boldsymbol{\lambda}.\mathbf{f}_j^1(\mathbf{p})\} \quad \text{for } j = 1, \ldots, N \qquad (13b)$$

for $\mathbf{p}(\lambda_0, \boldsymbol{\lambda})$ with $f^1_{r,j} = \partial f_r(\mathbf{p})/\partial p_j$. The correct values of λ_0 and $\boldsymbol{\lambda}$ in (13) are those which exactly satisfy the constraints (11).

The important questions as to the existence, uniqueness, and practical evaluation of the maximum-entropy solution remain, and may be usefully considered in two stages, namely: (i) the determination of the maximum-entropy solution, $\mathbf{p}(\boldsymbol{\lambda})$, which maximizes (12) for given $\boldsymbol{\lambda}$, and (ii) the determination of $\boldsymbol{\lambda} = \boldsymbol{\lambda}^s$ such that the chosen constraints are satisfied.

Concerning the first stage, one may note that if $Q(\mathbf{p})$ is a strictly concave function (*i.e.* one which is always underestimated by linear interpolation or, equivalently, one which has a negative definite Hessian, see *e.g.* Rockafellar, 1970) on the convex domain of probability distributions, $\mathbf{p}$, then it will have at most one local maximum in its domain (Rockafellar, 1970; Wernecke, 1977). Since $S_1(\mathbf{p})$ [or $S_2(\mathbf{p})$] is a strictly concave function and the sum of concave functions is also concave (Rockafellar, 1970) a sufficient but not necessary condition that $Q(\mathbf{p})$ be a strictly concave function is that each constraint contribution to Q individually be concave. A trivial case is that of only linear constraints (which are necessarily both convex and concave) and which therefore lead to at most one local maximum for $Q(\mathbf{p})$. For non-linear constraints the matter of establishing the concavity or non-concavity of the constraint contributions to Q is more difficult and, for the particular constraint functions introduced in the present work, will be discussed in §§ 4.1 and 4.4. The task of practically evaluating $\mathbf{p}(\boldsymbol{\lambda})$ for non-linear constraints is a central task in the present method and is capable of being approached in many ways (see *e.g.* Gull & Daniell, 1978; Willingale, 1981, and § 4.5), however, it suffices here to note that the function $Q(\mathbf{p})$ offers a possible variational approach to the determination of $\mathbf{p}(\boldsymbol{\lambda})$ and, for the specific constraints considered in the present work, would involve the results given in Table 1.

Concerning the second stage, namely the determination of the value of $\boldsymbol{\lambda} = \boldsymbol{\lambda}^s$ such that the constraints (11) are satisfied by $\mathbf{p}(\boldsymbol{\lambda}^s) \equiv \mathbf{p}^s$, one might choose to try to determine $\boldsymbol{\lambda}^s$ by least-squares fitting of the constraints to the constraint values (see *e.g.* Gull & Daniell, 1978; Willingale, 1981). However, even for convex constraints, one does not have any guarantee of a unique solution to the minimization problem and one may obtain false minima. Furthermore, evaluation of even first derivatives of the sum of squared deviations would in practice seem to require M distinct numerical solutions $\mathbf{p}(\boldsymbol{\lambda})$ for different $\boldsymbol{\lambda}$. For this reason, it seems useful to consider an alternative iterative variational

approach to the determination of λ^s which allows one to make certain restricted claims as to existence and uniqueness of solutions for λ, and also permits evaluation of arbitrary many orders of derivatives of the variational function w.r.t. λ from only one numerical solution for $\mathbf{p}(\lambda)$. For cases involving more than one or two constraints, this latter approach can lead to a very considerable saving in computation indeed.

The λ-variational principle has been introduced in the information-theoretic context by Agmon, Alhassid & Levine (1978, 1979) for the special case of linear constraints and has close ties with some standard results in statistical mechanics (see *e.g.* Landau & Lifshitz, 1980, ch. XII). In the present notation, the basic results for the λ-variational principle and linearized constraints are presented in Appendix 2 (see also § 3). For the case of non-linear constraints, as occur in the present study, it is valuable to investigate to what extent the variational principle may be applied either by linearizing the constraints and treating the problem iteratively or by generalizing the variational principle to cover non-linear constraints. These questions are explored in the following section and, in order to facilitate the discussion, it is helpful to reformulate some of the basic expressions introduced thus far in a manner so as to resemble similar results in statistical mechanics.

3. The λ-variational principle applied to non-linear constraints

3.1. *The structural freedom*

It is useful to rewrite (13*b*) as

$$p_j = \exp\left\{-\lambda_0 - \boldsymbol{\lambda}.\frac{\partial \hat{\mathbf{f}}(\mathbf{p})}{\partial p_j}\right\} \equiv \exp\{-\boldsymbol{\lambda}.\hat{\mathbf{f}}_j^1\}/z \qquad j = 1, \ldots, N, \tag{14a}$$

where we have redefined λ_0 and $\boldsymbol{\lambda}$, and

$$\hat{f}^1_{r,j} \equiv \frac{\partial \hat{f}_r}{\partial p_j} \equiv \frac{\partial f_r}{\partial p_j} - C'_r \equiv f^1_{r,j} - C'_r, \tag{14b}$$

with C'_r at present denoting an arbitrary constant. It may be noted that the normalizing condition (1) leads directly to the result that λ_0 is a function of the remaining M Lagrange multipliers, *viz*

$$\exp\{\lambda_0(\boldsymbol{\lambda}; \mathbf{C}')\} \equiv \exp\{F(\boldsymbol{\lambda}; \mathbf{C}')\} = z(\boldsymbol{\lambda}; \mathbf{C}') = \sum_{j=1}^{N} \exp\{-\boldsymbol{\lambda}.\hat{\mathbf{f}}_j^1\}, \tag{15}$$

where $z(\boldsymbol{\lambda}; \mathbf{C}')$ would be called the partition function in statistical mechanics and $\lambda_0(\boldsymbol{\lambda}; \mathbf{C}')$ or $F(\boldsymbol{\lambda}; \mathbf{C}')$ the grand potential. Different choices for $\mathbf{C}'$ lead to different forms for the variational function F, but do not alter $\mathbf{p}$ obtained by solving (14*a*). Differentiation of (14*a*) w.r.t. $\boldsymbol{\lambda}$ leads directly to the result

$$\frac{\partial F(\boldsymbol{\lambda}; \mathbf{C}')}{\partial \lambda_r} = -\sum_{j=1}^{N} \hat{f}^1_{r,j}(\mathbf{p})\, p_j - \sum_{j',j=1}^{N} \boldsymbol{\lambda}.\hat{\mathbf{f}}^2_{j'j}(\mathbf{p})\, p_j \frac{\partial p_{j'}}{\partial \lambda_r}, \tag{16}$$

where the first term on the right-hand side of (16) arises from the explicit dependence of $\mathbf{p}$ on $\boldsymbol{\lambda}$ while the second term arises from the implicit dependence of the constraint derivatives on $\boldsymbol{\lambda}$. The quantities $\partial p_j/\partial \lambda_r$ appearing in (16) may be expressed in terms of a set of coupled simultaneous equations (Appendix 3) which are potentially useful in improving numerical solution methods. In (16) we have adopted the notation

$$f^v_{r,j_v j_{v-1} \ldots j_1} \equiv \frac{\partial^v f_r}{\partial p_{j_v} \ldots \partial p_{j_1}} \tag{17}$$

and $\hat{\mathbf{f}}^1_j$ is defined in (14*b*) above. It is the presence of the second term in (16) and, more specifically, its explicit dependence on $\boldsymbol{\lambda}$ which appears to make difficult the development of a general variational principle based on F when the constraint functions, $\mathbf{f}(\mathbf{p})$, are non-linear. If, however, the constraints are linearized, as outlined, say, below, then some very useful results may be invoked.

3.2. *Linearized constraints, a cumulant generating function and a variational principle*

If we consider now the case where the *constraint derivatives*, $\hat{\mathbf{f}}^1_j$, are evaluated in an arbitrary *trial distribution* $\mathbf{p}^t$, which need not be a self-consistent distribution $\mathbf{p}(\boldsymbol{\lambda})$ satisfying (14*a*) for given $\boldsymbol{\lambda}$, we obtain the *linearized* form

$$p_j(\boldsymbol{\lambda}; \mathbf{p}^t) = \exp\{-\boldsymbol{\lambda}.\hat{\mathbf{f}}^1_j(\mathbf{p}^t)\}/z(\boldsymbol{\lambda}; \mathbf{p}^t) \tag{18}$$

and the corresponding form of (15) is

$$\exp\{\lambda_0(\boldsymbol{\lambda}; \mathbf{C}', \mathbf{p}^t)\} \equiv \exp\{F(\boldsymbol{\lambda}; \mathbf{C}', \mathbf{p}^t)\} = z(\boldsymbol{\lambda}; \mathbf{C}', \mathbf{p}^t) = \sum_{j=1}^{N} \exp\{-\boldsymbol{\lambda}.\hat{\mathbf{f}}^1_j(\mathbf{p}^t)\}, \tag{19}$$

where $F(\boldsymbol{\lambda}; \mathbf{C}', \mathbf{p}^t)$ [or $\lambda_0(\boldsymbol{\lambda}; \mathbf{C}', \mathbf{p}^t)$] may be viewed in statistical terms (see *e.g.* Kubo, 1962) as the *cumulant generating function* for the cumulants of the $\mathbf{f}^1(\mathbf{p}^t)$ in the distribution $\mathbf{p}(\boldsymbol{\lambda}; \mathbf{p}^t)$ (see Appendix 2.1), or, equivalently, the cumulants of $\hat{\mathbf{f}}^1(\mathbf{p}^t)$. Secondly, $F(\boldsymbol{\lambda}; \mathbf{C}', \mathbf{p}^t)$ is shown in Appendix 2.2 to be a *strictly convex* function of $\boldsymbol{\lambda}$ yielding a *variational approach* to the determination of $\boldsymbol{\lambda}^s(\mathbf{p}^t)$ and $\mathbf{p}^s(\mathbf{p}^t)$ which corresponds to the maximum-entropy solution satisfying the chosen constraints, but with the derivatives $\hat{\mathbf{f}}^1_j$ evaluated in the *trial*

distribution, $\mathbf{p}^t$. To establish this variational approach one chooses $\mathbf{C}'$ such that

$$\mathbf{C}' = \mathbf{C}^1(\mathbf{p}^t) = \sum_j \mathbf{f}_j^1(\mathbf{p}^t)\, p_j^s$$

$$= [\mathbf{C} - \mathbf{f}(\mathbf{p}^t)] + \sum_{j=1}^{N} p_j^t \mathbf{f}_j^1(\mathbf{p}^t) + (\Delta\mathbf{C}^1), \quad (20a)$$

where

$$\Delta\mathbf{C}^1 = \tfrac{1}{2}\sum_{jj'} (p_j^s - p_j^t)(p_{j'}^s - p_{j'}^t)\, \mathbf{f}_{jj'}^2(\mathbf{p}^t) + \cdots$$

$$= O(|\mathbf{p}^s - \mathbf{p}^t|^2). \quad (20b)$$

In (20) $\mathbf{p}^s$ denotes the sought after, but as yet unknown, maximum entropy solution to the full non-linear problem, so that for given $\mathbf{p}^t$ the error term in (20) is: (i) *small* (for $\mathbf{p}^t$ near $\mathbf{p}^s$), (ii) *constant* [*i.e.* independent of $\boldsymbol{\lambda}(\mathbf{p}^t)$] and (iii) ultimately *tends to zero* on full self-consistent solution of the problem when $\mathbf{p}^t \to \mathbf{p}^s$. From the variational principle for linear constraints outlined in Appendix 2, it immediately follows that, for a given $\mathbf{C}^1(\mathbf{p}^t)$, $\boldsymbol{\lambda}^s(\mathbf{p}^t)$ is *uniquely* determined as the solution of

$$\frac{\partial F(\boldsymbol{\lambda}; \mathbf{C}^1, \mathbf{p}^t)}{\partial \lambda_r} \equiv -\sum_{j=1}^{N} \hat{f}_{r,j}^1(\mathbf{p}^t) \exp\{-\boldsymbol{\lambda}.\hat{\mathbf{f}}_j^1(\mathbf{p}^t)\}$$

$$\equiv -\langle \hat{f}_r^1(\mathbf{p}^t)\rangle = 0, \quad r = 1, \ldots, M \quad (21)$$

and is the value of $\boldsymbol{\lambda}$ which simultaneously *minimizes* $F(\boldsymbol{\lambda}; \mathbf{C}^1, \mathbf{p}^t)$ and leads to the maximum-entropy solution, $\mathbf{p}^s(\mathbf{p}^t)$, satisfying the linearized constraints with constraint values $\mathbf{C}^1$. Working within the linearized formulation of the problem, it immediately follows that one may directly determine $F(\boldsymbol{\lambda}; \mathbf{C}^1, \mathbf{p}^t)$ as a function of $\boldsymbol{\lambda}$ and so determine $\boldsymbol{\lambda}^s(\mathbf{p}^t)$, the value at which F has a global minimum, *without any additional function evaluations* of $\hat{\mathbf{f}}_j^1(\mathbf{p}^t)$ (*i.e.* Fourier transforms of $\mathbf{p}$).

Alternatively one could choose instead to evaluate only a limited number of derivatives of $F(\boldsymbol{\lambda}; \mathbf{C}^1, \mathbf{p}^t)$ w.r.t. $\boldsymbol{\lambda}$ at $\boldsymbol{\lambda}$, *e.g. via* the cumulant property of $F(\boldsymbol{\lambda}; \mathbf{C}^1, \mathbf{p}^t)$ described in Appendix 2, and this also would not involve any additional Fourier transform operations. If one stopped at second cumulants of $\hat{\mathbf{f}}_j^1(\mathbf{p}^t)$, then numerically the process of estimating $\boldsymbol{\lambda}^s(\mathbf{p}^t)$ would correspond to the Newton–Raphson method, and *convergence would ultimately be quadratic* (*i.e.* very rapid) (see also Alhassid, Agmon & Levine, 1978). Thus, from a given trial solution for $\mathbf{p}$, one may immediately determine a maximum-entropy solution which (approximately) satisfies the linearized constraints if such a solution exists. Conditions for the existence of such a solution are given in Appendix 2.3 and in fact allow one to state the ranges of values of $\mathbf{C}^1$ for which a (unique) solution exists for given $\mathbf{p}^t$.

From a practical point of view, the above results provide the basis for an iterative approach to the determination of $\boldsymbol{\lambda}^s$ either from successive values of $\mathbf{p}(\boldsymbol{\lambda}^t)$, the maximum entropy solution to the non-linear problem for a particular trial $\boldsymbol{\lambda}$, or *via* solutions $\mathbf{p}^s(\mathbf{p}^t)$ which are maximum entropy solutions satisfying linearized constraints. Both these methods have been pursued for particular cases, and only involved iterative refinement of $\hat{\mathbf{f}}_j^1(\mathbf{p}^t)$ without any attempt to use a predictor-corrector approach to refine $\Delta\mathbf{C}^1$. However, an extrapolation procedure on $\lambda^s(\mathbf{p}^t)/\lambda$ was often found useful. The second case above offers an iterative pathway to determining $\mathbf{p}^s$ which proceeds *via* solutions, $\mathbf{p}^s(\mathbf{p}^t)$, which are always maximum-entropy solutions satisfying the linearized constraints, and involves simultaneous refinement of $\boldsymbol{\lambda}$ and $\mathbf{p}(\boldsymbol{\lambda})$ aimed at converging on $\mathbf{p}(\boldsymbol{\lambda}^s)$ (*i.e.* essentially a single-loop refinement).

3.3. *Summary of notation for probability distributions*

For convenience, we here summarize some of the more intricate nomenclature used for labelling the various different probability distributions introduced in this paper, namely:

$\mathbf{p} \equiv \mathbf{p}(\boldsymbol{\lambda})$	the maximum-entropy distribution when the Lagrange-multiplier vector is $\boldsymbol{\lambda}$.
$\mathbf{p}^t$	an arbitrary trial probability distribution.
$\mathbf{p}^s \equiv \mathbf{p}(\boldsymbol{\lambda}^s)$	the maximum-entropy probability distribution which satisfies the chosen constraints and occurs when $\boldsymbol{\lambda} = \boldsymbol{\lambda}^s$.
$\mathbf{p}(\boldsymbol{\lambda}; \mathbf{p}^t)$	the maximum-entropy distribution for arbitrary $\boldsymbol{\lambda}$ and linearized constraints evaluated in the trial distribution $\mathbf{p}^t$.
$\mathbf{p}^s(\mathbf{p}^t) \equiv \mathbf{p}^{s,t} \equiv \mathbf{p}[\boldsymbol{\lambda}^s(\mathbf{p}^t); \mathbf{p}^t]$	the maximum-entropy distribution satisfying the linearized constraints evaluated in the trial distribution $\mathbf{p}^t$.

4. Application of information theory to the crystallographic inversion problem – statistical geometry

In conventional crystallography, the primary data are in the form of diffraction amplitudes, while it is the Fourier transform of the structure factors (amplitudes times a phase factor) which is of physical interest. Not only is the phase factor for each reflexion usually not known, but the extent and precision of the data in reciprocal space is inevitably limited by measurement errors and physical limitations such as finite resolution

of the measuring apparatus. The essential problem to be tackled is the *optimal* (least-biased) determination of the structure (say electron density) that one can make on the basis of: (i) limited and noisy data (structure amplitudes), and (ii) other *a priori* assumptions about the system. Such a problem is immediately suitable for treatment by information theory (or more specifically the particular application developed here called the *statistical geometric method*, and henceforth abbreviated to SGM), provided one can express one's *knowledge and assumptions* about the system in tractable mathematical form.

For example, following the recent work of Gull & Daniell (1978), who considered the closely-related problem of processing radio-astronomical data, the first stage in the approach is to consider the unit cell to be divided up into N equal subcells (pixels), and for $\mathbf{p}_j$ to denote the fraction of the unit-cell scattering density in cell j. For simplicity we will restrict our development to the one-dimensional case (the approach immediately extends to arbitrary higher dimension). The Fourier transform of the p_i will be denoted by P_k, where these quantities are related by (A1.1) and (A1.2). Concerning the available information, we will firstly suppose that we have measurements of the P_k, including phase information, denoted by E_k, for a subset of reflexions, D_1. Secondly, we will assume that we have measurements of the structure amplitudes, A_k, for the subset of reflexions D_2, which may overlap D_1. Finally, we will introduce some *a priori* information about the expected structure, namely, that we wish the structure-factor map to be at a uniform level for the subset of cells $j \in B$. Such a constraint is sometimes used by protein crystallographers to assist in phase determination (Colman, private communication).

4.1. *Information in the form of weak constraints*

At first glance, one might think that the most efficacious way to proceed would be to enforce the distribution $\mathbf{p}$ to satisfy *exactly* the datum for each individual reflexion or piece of information (this would correspond to a *method of strong constraints*). However, the data are themselves known only approximately and so exact fitting to such noisy data would lead to spurious detail in our structure map $\mathbf{p}$. Moreover, the dimensionality of our problem in $\boldsymbol{\lambda}$ space would be of order $(N_1 + N_2 + N_3)$ and so would rapidly become intractable in numerical terms for data sets involving many reflexions and, even worse, the problem may become intractable in principle since a solution may not even exist. By contrast, a much more powerful and simpler approach to the problem of structure determination in the present framework is to adopt *weak constraints* (at least in treating any information which is *not known exactly* or to very high accuracy). Such an approach using weak constraints has the following advantages:

(i) filtering of the map to avoid spurious detail (see *e.g.* Ables, 1974; Wernecke, 1977; Gull & Daniell, 1978).

(ii) existence of a solution in practice, since M is very small and N is very large so that the number of degrees of freedom is very large.

(iii) ease of treatment by numerical methods since the dimensionality in $\boldsymbol{\lambda}$ space is low (§ 3).

There are many parallels one might draw between the method of weak constraints adopted here and similar approaches adopted in statistical mechanics. For example, in the canonical ensemble of equilibrium statistical mechanics one assumes only that the expectation value of the Hamiltonian, $H(\mathbf{x})$, over the states $\mathbf{x}$ for an n-particle system is equal to the energy, E, *i.e.*

$$\sum_{\mathbf{x}} P(\mathbf{x})\, H(\mathbf{x}) \equiv \langle H \rangle = E, \tag{22}$$

and *not* that, say, the expectation value of the Hamiltonian for each type of particle is equal to E/n_1. It is helpful to appreciate that equilibrium statistical mechanics in, say, the canonical ensemble (Jaynes, 1957) involves only one piece of information in the information-theoretic sense, namely the weak constraint (22), and yet is routinely used to predict *macroscopic and average microscopic properties* of systems where N is typically of order 10^{23}. The Lagrange multiplier corresponding to (22) is, of course, the inverse temperature $1/k_B T$, where k_B is Boltzmann's constant.

In the present case we introduce the following appropriate, but by no means necessary or exhaustive, types of information in the form of weak constraints.

4.1.1. *The starting or known structure factors.*

If one has some *a priori* information on the phases of certain reflexions (*e.g.* from dynamical scattering measurements, Hurley & Moodie, 1980; heavy-atom techniques; anomalous scattering; envelope functions; guesswork; *etc.*) and if this is combined with the measured X-ray diffraction amplitudes to give phased structure factors, E_k, for a set of reflexions, D_1, then an appropriate weak constraint function to incorporate this data, taking into account the errors in the data, is (see also Gull & Daniell, 1978)

$$f_1(\mathbf{p}) = \frac{1}{2N_1} \sum_{k \in D_1} \frac{|P_k - E_k|^2}{\sigma_{k,1}^2}, \tag{23}$$

where N_1 is the number of reflexions in D_1 (taken to include all Friedel pairs k and $-k$), $\sigma_{k,1}$ represents the standard deviation in the datum E_k, while P_k is the trial structure factor and is given by (A1.2). To be

Table 1. *Table of constraint derivatives*

$$f^{v}_{r,j_v,j_{v-1},\ldots,j_1} \equiv \frac{\partial^v f_r}{\partial p_{j_v} \ldots \partial p_{j_1}}.$$

The various derivatives of P_k and $|P_k|$ w.r.t. p_j are given in Appendix 1 and all derivatives of the f_r here are zero for $v > 2$ and $r = 1$ and 3.

Constraint number r	f^0_r	C_r	$f^1_{r,j}$	$f^2_{r,j'j}$	$f^2_{r,jj}$
0	$\sum_{j=1}^{N} p_j$	1	1	0	0
1	$\frac{1}{2N_1}\sum_{k\in D_1}\frac{\lvert P_k - E_k\rvert^2}{\sigma^2_{k,1}}$	1	$\frac{1}{N_1}\sum_{k\in D_1}\frac{\exp\{-2\pi ijk/N\}}{\sigma^2_{k,1}} \times (P_k - E_k)$	$\frac{1}{N_1}\sum_{k\in D_1}\frac{1}{\sigma^2_{k,1}} \times \exp\{2\pi i(j'-j)k/N\}$	$\frac{1}{N_1}\sum_{k\in D_1}\frac{1}{\sigma^2_{k,1}}$
2	$\frac{1}{2N_2}\sum_{k\in D_2}\frac{(\lvert P_k\rvert - A_k)^2}{\sigma^2_{k,2}}$	1	$\frac{1}{N_2}\sum_{k\in D_2}\frac{1}{\sigma^2_{k,2}} \times \exp\{-2\pi ijk/N\} \times (P_k - A_k \exp\{i\varphi_k\})$	$\frac{1}{N_2}\sum_{k\in D_2}\frac{1}{\sigma^2_{k,2}} \times \left\{\left(1-\frac{1}{2}\frac{A_k}{\lvert P_k\rvert}\right) \times \exp[2\pi i(j'-j)k/N] + \frac{1}{2}\frac{A_k}{\lvert P_k\rvert} \times \exp[2i\varphi_k - 2\pi i(j+j')k/N]\right\}$	$\frac{1}{N_2}\sum_{k\in D_2}\frac{1}{\sigma^2_{k,2}} \times \left\{\left(1-\frac{1}{2}\frac{A_k}{\lvert P_k\rvert}\right) + \frac{1}{2}\frac{A_k}{\lvert P_k\rvert} \times \exp[2i\varphi_k - 4\pi ijk/N]\right\}$
3	$\frac{1}{N_3}\sum_{j\in B}\frac{\left(p_j - \sum_{j'} p_{j'}/N_3\right)^2}{\sigma^2_b}$	$\frac{(N_3-1)^2}{N_3^2}$	$\frac{2}{\sigma^2_b}\frac{(N_3-1)}{N_3^2} \times \left[p_j - \sum_{j'} p_{j'}/N_3\right]$	$\frac{2}{\sigma^2_b}\frac{(N_3-1)^2}{N_3^2}\left(\delta_{jj'} - \frac{1}{N_3}\right)$	$\frac{2}{\sigma^2_b}\frac{(N_3-1)^3}{N_3^3}$

consistent, we assume that the data E_k have been scaled such that $E_0 = 1$, so that both the E_k and P_k are *unitary structure factors* [see *e.g.* Tsoucaris & de Rango, 1970]. If the E_k have Gaussian errors with standard deviation $\sigma_{k,1}$ and the P_k are assumed to be true values, then $f_1(\mathbf{p})$ possesses a reduced χ^2 distribution with $C_1 = 1$.

It should be noted that the constraint function (23) together with its associated Lagrange multiplier, λ_1^s, leads to a concave contribution to Q *via* (12), since $f_1(\mathbf{p})$ is a convex function (see Appendix 4 or an alternative proof by Wernecke, 1977) and λ_1^s is positive (Appendix 5). Thus, the introduction of information into the statistical geometrical method *via* this constraint does not lead to any problems of non-uniqueness of $\mathbf{p}(\boldsymbol{\lambda})$ (see §§ 2.3 and 4.4).

The Lagrange multiplier associated with constraint, λ_1, may be viewed as a *generalized contrast parameter* since for $\lambda_1 = 0$ it can be seen *via* (14*a*) that no contrast arises in $\mathbf{p}$ due to this information, whereas as $\lambda_1 > 0$ becomes larger, more and more contrast due to the information $f_1(\mathbf{p})$ is introduced into the structure.

4.1.2. *The measured structure amplitudes.* For reflexions where there is no *a priori* information on the phase or where the phase assignment is highly uncertain, an appropriate weak constraint to incorporate such data, taking into account the errors in the amplitudes, is (see also Gull & Daniell, 1978)

$$f_2(\mathbf{p}) \equiv \frac{1}{2N_2}\sum_{k\in D_2}\frac{(\lvert P_k\rvert - A_k)^2}{\sigma^2_{k,2}} \equiv \frac{1}{2N_2}\sum_{k\in D_2}\frac{\lvert P_k - A_k\exp\{i\varphi_k\}\rvert^2}{\sigma^2_{k,2}}, \tag{24}$$

where N_2 is the number of reflexions in D_2 (taken to include all Friedel pairs), $\sigma_{k,2}$ is the standard deviation in the measured unitary structure amplitudes, A_k. The quantity $f_2(\mathbf{p})$ does not have a χ^2 distribution; however, the number of structure amplitudes is usually so large that we may apply the central limit theorem to prove that the distribution of f_2 is approximately normal with expectation value $C_2 \simeq 1$.

Unlike the previous constraint function, the amplitude constraint function (24) is *not* in general convex, as can readily be demonstrated by trying to pursue an argument similar to that given in Appendix 4, whereupon it may be shown that if there exists an $A_k \neq 0$, then one can find a $\mathbf{p}$ such that $|P_k|$ is sufficiently small that the quadratic form associated with $f^2_{2,jj'}$ (see Table 1) becomes negative. Thus, the introduction of information into the statistical geometrical method *via* (24) leads to the possibility that Q given by (12) is no longer strictly concave and hence of non-unique solutions for $\mathbf{p}(\boldsymbol{\lambda})$, as might have been expected on other grounds. As a matter of consistency, we note at this point that constraints (23) and (24) treat zero-

amplitude reflexions identically (provided one assumes the same errors).

As regards the Lagrange multiplier λ_2, this too may be viewed as a *generalized contrast parameter*, only corresponding here to the amplitude information $f_2(\mathbf{p})$.

4.1.3. *Background flattening.* In order to enforce local flatness in a structure, the following weak constraint is introduced:

$$f_3(\mathbf{p}) = \frac{1}{N_3} \sum_{j \in B} \frac{\left[p_j - \sum_{j' \in B} p_{j'}/N_3 \right]^2}{\sigma_b^2}, \qquad (25)$$

where N_3 is the number of cells in B. This constraint simply expresses the requirement that the expectation value of p_j for all $j \in B$ is equal to the mean value of p_j taken over the N_3 cells in B. The standard deviation for the distribution of p_j is taken to be σ_b^2, and so the expectation value of (25) is given by $C_3 = [(N_3 - 1)/N_3]^2$.

The constraint $f_3(\mathbf{p})$ given by (25) may readily be shown to be convex by an argument closely resembling that given in Appendix 4 and using the $f^2_{3,jj'}$ listed in Table 1. For constraint function (25), the associated Lagrange multiplier λ_3 may be viewed as controlling the level of noise or contrast in the background density of the map.

4.2. *Information in the form of strong constraints*

The use of weak constraints is very desirable for building the bulk of the diffraction information into the determination of **p**; however, in addition to information arising from weak constraints, one might want to build certain other information into **p** in the form of *strong constraints*. For example, one might consider that the distribution is given by $\mathbf{p}(\gamma)$, where γ is some smaller set of parameters than the p_j and that the maximum entropy distribution $\mathbf{p}^s(\gamma^s)$ is to be determined. In practice, γ might be the atomic coordinates or the various parameters in a bonding-electron density basis set.

Another form of information which may be introduced into **p** in the form of strong constraints is that of non-crystallographic symmetry (*e.g.* see Bricogne, 1974), which involves the averaging of p_j over cells which correspond to equivalent elements of the basic unit of structure (*e.g.* a virus particle).

4.3. *Information automatically incorporated in the structure map via the maximum-entropy formalism*

4.3.1 *Normalization.* The requirement that **p** be normalized, *i.e.* $\sum_{j=1}^{N} p_j = 1$, is guaranteed in the present formalism *via* the introduction of λ_0 or the partition function z in (14*a*).

4.3.2. *Positivity.* The maximum-entropy formulation of the problem and the introduction of **p** as a discrete probability distribution *automatically guarantee* that

$$1 > p_j > 0 \quad \text{for } j = 1, \ldots, N, \qquad (26)$$

as can readily be seen from (14*a*).

4.3.3. *Smoothness.* From a fundamentalist viewpoint, smoothness of the maximum-entropy structure map arises from the maximum-entropy principle and the concept of entropy as a measure of the amount of information in the structure map. The larger the entropy, the smaller the amount of information in the map. More specifically, the maximum-entropy structure, $\mathbf{p}^s$, tends to be smooth in the sense of having the most nearly uniform distribution of **p** values satisfying the given information (Gull & Daniell, 1979). This has been shown mathematically by Wernecke (1977) for both the S_1 and S_2 definitions of entropy, in the sense that application of a general linear smoothing operator to a structure never decreases the value of the entropy and constitutes a form of global rather than local smoothness. Local smoothness may be enforced by additional constraints, if necessary.

4.4. *Uniqueness and symmetry*

The question of the uniqueness of SGM solutions was discussed to some extent in § 2.3 where it was divided into two parts, namely: Firstly, that which concerns the maximum-entropy solution, $\mathbf{p}(\boldsymbol{\lambda})$, for given $\boldsymbol{\lambda}$, which is unique if Q given by (12) is strictly concave w.r.t. the convex domain of possible structures, **p**, for which a sufficient condition is that each of the constraint functions multiplied by its associated Lagrange multiplier makes a concave contribution to Q. As we have seen from § 4.1, both constraint functions $f_1(\mathbf{p})$ and $f_3(\mathbf{p})$ make concave contributions to Q, whereas $f_2(\mathbf{p})$ does not. The non-concavity of the amplitude constraint contribution to Q is an expression in the present context of the well-known general non-uniqueness of solutions to crystal structure determination (see *e.g.* Dainty, Fiddy & Greenaway, 1979). Nevertheless, it can be seen from the present discussion and the properties of convex functions that, for given values of λ_1 and λ_3, there will exist a range of values of λ_2 including zero for which Q will remain strictly concave and so have a unique solution. This observation suggests the following approach to the determination of SGM solutions when non-concave contributions to Q are involved (see also § 4.6 and Fig. 1), namely: (i) establish the unique solution, $\mathbf{p}(\boldsymbol{\lambda})$, which maximizes the corresponding Q and then gradually introduce non-concave contributions to Q by gradually increasing their associated Lagrange multiplier, so as to try to avoid jumping to distant solutions when uniqueness breaks down; (ii) to choose constraints which give as strongly or nearly concave contributions to Q as possible (*e.g.* tend to minimize the maximum eigenvalue of the Hessian of Q, see also Wilkins, 1983*a*).

Clearly, the constraints should, where possible, be chosen to contain sufficient information so as to distinguish between different symmetry cases. For example, if only the constraint f_2 were imposed (which is translationally invariant) then the origin of the unit cell in direct space would not be defined. This could readily be overcome by introducing a maximum of three phased structure factors into the constraint f_1. A second example of degeneracy is that of distinguishing centro- and non-centrosymmetric structures. Additional phase information may be needed to fix the handedness of the structure. The general problem of homometric structures (*i.e.* different structures having the same diffraction intensities) also means that one may have to assign some additional phases to produce unique solutions. If sufficient information is included in the constraints, then the SGM may be expected to pick the smoothest structure which fits the data (*e.g.* centrosymmetric rather than non-centrosymmetric structures).

An immediate practical consequence of case (ii) is that, although mathematically zero measured structure amplitudes are treated identically by constraints f_1 and f_2, it is preferable to include zero and near-zero structure amplitude cases in f_1 than in f_2, since phase errors for such amplitudes will largely be irrelevant, whereas inclusion of such near-zero structure amplitudes in f_2 can lead to large departures from convexity of f_2. Thus, inclusion of zero-measured structure factors in the SGM offers to provide valuable additional information in helping initial structure determination (also J. White, private communication) by contrast with conventional methods (see *e.g.* Woolfson, 1980) which tend to ignore these cases in the initial structure determination stage.

Secondly, we briefly consider the uniqueness of maximum-entropy solutions $\mathbf{p}(\boldsymbol{\lambda}^s)$ or $\mathbf{p}[\boldsymbol{\lambda}^s(\mathbf{p}^t); \mathbf{p}^t]$ as functions of $\boldsymbol{\lambda}$. The interpretation in § 4 of λ_1 and λ_2 as generalized contrast parameters and the knowledge that $\mathbf{p}(\boldsymbol{\lambda}^s)$ and $\mathbf{p}[\boldsymbol{\lambda}^s(\mathbf{p}^t); \mathbf{p}^t]$ satisfy the available information in some sense suggests that the uniqueness of $\boldsymbol{\lambda}^s$ is not as serious a matter as that of $\mathbf{p}(\boldsymbol{\lambda})$ discussed earlier. From the work of § 3, Appendix 2 and Levine (1980), it follows that $\boldsymbol{\lambda}^s(\mathbf{p}^t)$ is unique for given $\mathbf{p}^t$, $\mathbf{C}^1(\mathbf{p}^t)$ and linearly independent constraints (*i.e.* $\partial C_r^1/\partial C_s^1 = \delta_{rs}$). Thus one may seek to determine $\boldsymbol{\lambda}^s$ from an initial trial structure, $\mathbf{p}^t$, by an iterative procedure using the $\boldsymbol{\lambda}$-variational principle, with the knowledge that, at each step in the iterative procedure, an essentially unique structure is determined which differs in generalized contrast from the previous one.

4.5. *Numerical solution for the scattering density*

The equations (14*a*) may be expressed in a form similar to that given by Gull & Daniell (1978) by substituting for $\hat{f}_{r,j}^1$ using the results given in Table 1 and Appendix 1. For a given trial $\boldsymbol{\lambda}$, the resulting set of non-linear equations $\mathbf{p}(\boldsymbol{\lambda})$ are amenable to solution by iteration substitution (see *e.g.* Gull & Daniell, 1978; Willingale, 1981), although some smoothing of successive solutions was found necessary in order to achieve convergence. Self-consistency of the solution may be checked by comparing the left- and right-hand sides of (14*a*). In principle, the solution may be evaluated to arbitrary numerical accuracy and does not involve any approximations of a mathematical kind. The value of $\boldsymbol{\lambda}$, namely $\boldsymbol{\lambda}^s$, which satisfies the chosen constraints may be established by the $\boldsymbol{\lambda}$-variational principle outlined in § 3, and this has been found to work quite well in practice.

In following papers of this series (Wilkins, 1983*a*,*b*) other approaches to the determination of the structure map (*i.e.* $\mathbf{p}$) will be presented.

4.6. *Illustrative example*

In following papers of this series (Wilkins, 1983*a*,*b*) the SGM may be used, we have taken a simple one-dimensional trial structure consisting of three Gaussian peaks with different heights and half-widths (see thick solid curve in Fig. 1). The two lowest-order structure amplitudes (including random errors) were then assigned phases (one corresponding to a choice of origin in the unit cell and the second obtained by adding a random error to the true phase) which on Fourier inversion yield the initial trial structure shown as a dotted curve in Fig. 1. Using only constraint 1, with the two phased structure factors, we find that the SGM yields the two-peaked structure shown as a dashed curve in Fig. 1. This structure shows: (i) better resolution than the initial trial structure, (ii) no regions

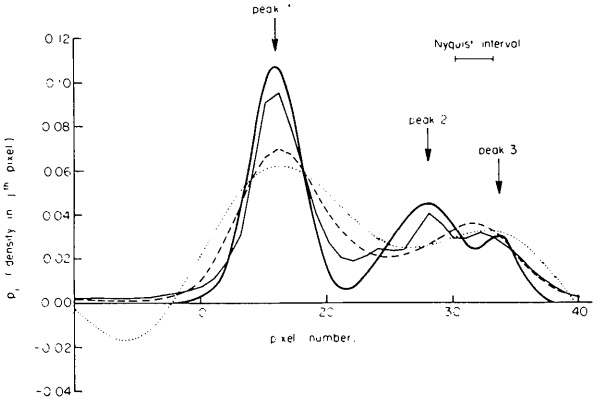

Fig. 1. One-dimensional structure distributions corresponding to: (i) true structure (thick solid curve), (ii) trial structure using two phased structure factors with errors (dotted curve), (iii) SGM determination of structure based on two phased structure factors with errors (dashed curve), (iv) SGM determination of structure based on two phased structure factors and ten structure amplitudes all with errors (thin solid curve).

of negative density, and (iii) slightly less bias in the assignment of density to peaks 2 and 3. Addition of the next ten amplitudes (with random errors) into the amplitude constraint alone and continued refinement with the SGM using two constraints yields the faint solid curve shown in Fig. 1. This structure shows extremely good resolution and agreement with the true structure and has three main peaks with near true peak-to-peak separations and low noise in the background. A very weak false peak arises between peaks 1 and 2, consisting of only one point, but otherwise the SGM solution is extremely smooth and is a conservative estimate of the true structure. Also shown in Fig. 1 is the Shannon–Nyquist interval (Shannon, 1949) corresponding to 14 degrees of freedom in the data (*i.e.* two phased structure factors and ten amplitudes). Structural information inside this interval corresponds to super-resolution. It should be noted that the determination of structure in the present case involved the assignment of only one non-arbitrary phase and no assumption was made in the refinements as to peak shape. Precise details of the model and method of solution will be given elsewhere (Wilkins & Varghese, in preparation).

5. Discussion

The statistical geometrical method offers a powerful and general approach to the crystallographic inversion problem. It is not practical to discuss here in detail all the advantages which this method offers. However, we briefly outline below some of the advantages which we expect for the method and which will be explored in subsequent papers.

(i) The statistical geometrical approach to the crystallographic inversion problem is an optimal method in a clearly defined statistical sense, and is capable of *exact* solution to arbitrary numerical accuracy (Gull & Daniell, 1978; Willingale, 1981).

(ii) The method operates in *direct space* where information is easily interpreted.

(iii) It allows one gradually to increase the number of reflexions and/or resolution in direct space, *i.e.* it can be used to treat the problems of *phase determination* and *phase extension* by continuous degrees, without any fundamental need to distinguish these cases. The method of statistical geometry even allows a certain (small) amount of 'super-resolution' to be achieved (*e.g.* see Frieden, 1972, and Gull & Daniell, 1978), *i.e.* some resolution to be obtained inside the Nyquist–Shannon interval (Shannon, 1949) due to the additional assumption in the SGM of smoothness and positivity of **p**.

(iv) It does not necessarily involve the concept of atomicity and so is also applicable to: (*a*) low-angle neutron studies of protein structures, (*b*) radio-astronomy, (*c*) electron-micrograph and lattice-image processing, (*d*) optical-image processing, and (*e*) almost any image processing problem.

(v) The method is not necessarily restricted to the analysis of Bragg reflexions but may in principle be used to invert the *total* measured scattering data (for the case of diffuse scattering alone, see also Clapp, 1969, 1971; Wilkins, 1972). However, this would in practice seem to involve the introduction of clusters or 'supercells' in direct space.

(vi) The method incorporates information about the estimated *accuracy* of each reflexion (data point).

(vii) The method incorporates a diagnostic for 'bad' reflexions, since one may investigate the contribution of each reflexion to constraints such as 1 and 2 in § 4.

(viii) The method can incorporate such information as background flattening and non-crystallographic symmetry (see *e.g.* Bricogne, 1974) which is used in the structure determination of biological macromolecules.

(ix) It does *not* involve the inversion of very large matrices or other weighty numerical tasks.

(x) Provided the errors $\sigma_{k,1}$ and $\sigma_{k,2}$ are correctly chosen, it leads to the *smoothest* map consistent with the original data (*i.e.* the method incorporates a *filtering procedure*) and does not lead to spurious detail due to series terminations (see Gull & Daniell, 1978, 1979; Wernecke, 1977).

(xi) The constraints may be interactively modified at any stage in the solution process so as to 'bootstrap' to a solution, *e.g.*, as certain features of the solution are ascertained, D_1 can be enlarged.

(xii) The method may be used to try to improve resolution in one region of a structure map while treating the remainder of the map in a constrained manner (see Gull & Daniell, 1978), *e.g.* some groups of pixels in the structure may be set equal. This could be of value, say, if one is particularly interested in certain 'active sites' of a large molecule and wishes to try to develop this region to higher resolution or, conversely, if the structure is known to higher resolution in one region, this may be used to help structure determination elsewhere. Also, one might choose to refine the structure in alternate horizontal and vertical slices (a 'multislice' approach) which may have some computational advantages.

(xiii) The SGM is well suited to problems with large N such as the determination of the structures of biological macromolecules or the determination of high-resolution electron-density maps in small-molecule systems. The SGM should not become greatly more difficult to evaluate in practice as the number of reflexions and the number of pixels, N, increases, since the problem of numerical solution is approxmately linear in the number of reflexions and in N, and is limited by the speed of the Fourier transform operation.

(xiv) The close formal correspondence between the crystallographic inversion problem and equilibrium statistical mechanics, which is apparent from the present work, should enable many of the very powerful techniques and important results of statistical mech-

anics and the many-body problem to be brought to bear on to the statistical-geometrical problem. This may in fact help in the derivation of *explicit mathematical results* for the statistical-inversion problem.

(xv) The statistical-geometric formalism guarantees the positivity of the p_j's as a *minimal condition* (see § 4.3.2).

(xvi) In addition to purely geometrical information from diffraction measurements, one might also choose to introduce information of other types, *e.g.* energetic considerations, into the formalism, given some knowledge of the dependence of configurational energy on **p** or a subset of **p**.

(xvii) Although mathematically one seeks $\boldsymbol{\lambda}^s$ as providing the optimal determination of the structure based on the given information, values of $\boldsymbol{\lambda}$ near $\boldsymbol{\lambda}^s$ in practice are sufficient to determine a structure to high accuracy, since the interpretation of, say, λ_1 and λ_2 in § 4.1 as *generalized contrast parameters* means that variation of these away from λ_1^s and λ_2^s, respectively, will primarily *change* the *contrast* in the maps but *not* the *underlying 'structure', i.e.* the spatial relationships in the structure such as coordination numbers, bond lengths, and angles.

6. Conclusions and prospects

The statistical-geometric method, because of its generality and numerical tractability, offers a powerful approach to the crystallographic inversion problem, even when very large and complicated structures are being studied by X-ray and other diffraction techniques. In the long term, one may view the method as providing a mathematical pathway to microscopy starting from diffraction data and any *a priori* information or assumptions about the structure.

In later papers in this series we will concern ourselves more with practical aspects of the method and develop both mathematical and numerical approaches to the solution of the central equations in the method, and apply these methods to simulated and real structure determinations.

The authors wish to thank Drs Colman, Fraser, Head, Johnson, Mathieson and Moodie for helpful discussions. One of us (SWW) wishes to thank the ILL for its hospitality during a stay where this work was initiated.

APPENDIX 1

Some useful definitions and results

$$p_j = N^{-1} \sum_k P_k \exp\{-2\pi ijk/N\} \qquad (A1.1)$$

$$P_k = |P_k| \exp\{i\varphi_k\} = \sum_{j=1}^{N} p_j \exp\{2\pi ijk/N\} \qquad (A1.2)$$

$$P_k^* = P_{-k} \qquad (A1.3)$$

$$\frac{\partial P_k}{\partial p_j} = \exp\{2\pi ijk/N\}; \frac{\partial^2 P_k}{\partial p_{j'}\,\partial p_j} = 0 \qquad (A1.4)$$

$$\frac{\partial |P_k|}{\partial p_j} = \cos(\varphi_k - 2\pi jk/N) \qquad (A1.5)$$

$$\frac{\partial^2 |P_k|}{\partial p_{j'}\,\partial p_j} = \frac{1}{|P_k|}\{\cos|2\pi(j'-j)| - \cos|\varphi_k - 2\pi jk/N| \cos|\varphi_k - 2\pi j' k/N|\}. \qquad (A1.6)$$

APPENDIX 2

Properties of the scalar potential $F(\boldsymbol{\lambda}; \mathbf{C}')$

The scalar potential $F(\boldsymbol{\lambda}; \mathbf{C}')$ was defined in (15), for arbitrary $\boldsymbol{\lambda}$, $\mathbf{C}'$ and $\mathbf{p}'$, by

$$F(\boldsymbol{\lambda}; \mathbf{C}', \mathbf{p}') \equiv \ln\left[\sum_j \exp\{-\boldsymbol{\lambda}.\hat{\mathbf{f}}_j^{1,t}\}\right] \equiv \ln\left[\sum_j \exp\{-\boldsymbol{\lambda}.\hat{\mathbf{f}}_j^{1,t}\}\right] + \boldsymbol{\lambda}.\mathbf{C}'. \qquad (A2.1)$$

A2.1. Cumulant generating function

Considering now the derivatives of F with respect to $\boldsymbol{\lambda}$ evaluated at $\boldsymbol{\lambda}$, one may readily show that:

First derivative:

$$\frac{\partial F(\boldsymbol{\lambda}; \mathbf{C}', \mathbf{p}')}{\partial \lambda_r} = -\sum_j \hat{f}_{rj}^{1,t} p_j = -\langle \hat{f}_r^{1,t} \rangle = -\sum_j (f_{rj}^{1,t} - C_r') p_j = -\langle f_r^{1,t} - C_r' \rangle. \qquad (A2.2)$$

Second derivative:

$$\frac{\partial^2 F(\boldsymbol{\lambda}; \mathbf{C}', \mathbf{p}')}{\partial \lambda_s\,\partial \lambda_r} = \sum_j \hat{f}_{rj}^{1,t} \hat{f}_{sj}^{1,t} p_j - \left(\sum_j \hat{f}_{rj}^{1,t} p_j\right)\left(\sum_j \hat{f}_{sj}^{1,t} p_j\right) = \langle (\hat{f}_r^{1,t} - \langle \hat{f}_r^{1,t} \rangle)(\hat{f}_s^{1,t} - \langle \hat{f}_s^{1,t} \rangle) \rangle = \langle \hat{f}_r^{1,t} \hat{f}_s^{1,t} \rangle - \langle \hat{f}_r^{1,t} \rangle \langle \hat{f}_s^{1,t} \rangle, \qquad (A2.3)$$

where all expectation values are taken in the distribution $\mathbf{p}(\boldsymbol{\lambda}; \mathbf{p}')$. That is, in statistical terms (see *e.g.* Kubo, 1962), $F(\boldsymbol{\lambda}; \mathbf{C}', \mathbf{p}')$ is the cumulant-generating function for cumulants of $\hat{f}_r^{1,t}$ in the distribution $\mathbf{p}(\boldsymbol{\lambda}; \mathbf{p}')$.

*A*2.2. *Variational function (structural freedom)*

It follows from (*A*2.3) that $F(\boldsymbol{\lambda}; \mathbf{C}', \mathbf{p}^t)$ is a *strictly convex* function whenever the constraints are *linearly independent* (Levine, 1980). From (*A*2.2) it follows that the problem of satisfying the M constraints (11*b*) can equivalently be expressed as determining the value of $\boldsymbol{\lambda}$ such that

$$\frac{\partial F(\boldsymbol{\lambda}; \mathbf{C}', \mathbf{p}^t)}{\partial \lambda_r} \equiv \frac{\partial F}{\partial \lambda_r}(\boldsymbol{\lambda}; \mathbf{C}^1, \mathbf{p}^t) = -\langle \hat{f}_r^1 \rangle = 0, \quad \text{for } r = 1, \ldots, M \tag{A2.4}$$

where $\mathbf{C}^1$ is given by (20).

Thus the *problem of maximizing* (3) *subject to* (11) *is equivalent to minimizing the strictly convex* (since *A*2.3 is a variance–covariance matrix for the variable $\hat{\mathbf{f}}^1$) *function F as a function of* $\boldsymbol{\lambda}$. This result forms the basis for a very direct and powerful variational method for determining $\boldsymbol{\lambda}$ such that the M constraints are satisfied, since knowledge of $\hat{\mathbf{f}}^{1,t}$ for a trial $\mathbf{p}^t$ immediately leads to knowledge of arbitrarily many higher derivatives of F w.r.t. $\boldsymbol{\lambda}$ *via* (*A*2.3) or to $F(\boldsymbol{\lambda}; \mathbf{C}^1)$ itself *via* (*A*2.1). For linear constraints $C_r^1 = C_r$ and this method has been discussed by Agmon, Alhassid & Levine (1978, 1979). It may be noted that even for extremely large data sets the task of solving for $\boldsymbol{\lambda}$ (*i.e.* minimizing F in $\boldsymbol{\lambda}$) involves only simple numerical operations and the inversion of an $M \times M$ matrix, where M in the weak-constraint method is typically less than, say, 10.

It may also be seen from (9), (*A*2.1), and the above variational principle for F (see also Agmon, Alhassid & Levine, 1979) that

$$S[\mathbf{p}(\boldsymbol{\lambda}; \mathbf{p}^t)] \leq S[\mathbf{p}(\boldsymbol{\lambda}^s; \mathbf{p}^t)] = F(\boldsymbol{\lambda}^s(\mathbf{p}^t); \mathbf{p}^t) \leq F(\boldsymbol{\lambda}; \mathbf{p}^t). \tag{A2.5}$$

so that $F(\boldsymbol{\lambda}; \mathbf{C}^1, \mathbf{p}^t)$ is an *upper bound to the entropy* of any distribution which is consistent with the same set of constraints.

*A*2.3. *Asymptotic behaviour of F w.r.t.* $\boldsymbol{\lambda}$

From (*A*2.1) one can see that for unique upper and lower bounds, $\hat{\mathbf{f}}_{\max}^{1,t}$ and $\hat{\mathbf{f}}_{\min}^{1,t}$ respectively, on $\hat{\mathbf{f}}_j^{1,t}$ as a function of j, the structural freedom, F, asymptotes to

$$-\boldsymbol{\lambda}.\hat{\mathbf{f}}_{\min}^{1,t} \quad \text{as } \boldsymbol{\lambda} \to +\infty$$

and (*A*2.6)

$$-\boldsymbol{\lambda}.\hat{\mathbf{f}}_{\max}^{1,t} \quad \text{as } \boldsymbol{\lambda} \to -\infty.$$

Given the above assumption, a necessary and sufficient condition that $F(\boldsymbol{\lambda}; \mathbf{C}')$ has a minimum [*i.e.* that a solution for $\boldsymbol{\lambda}^s(\mathbf{p}^t)$ exists] is that

$$\left.\begin{aligned} \hat{f}_{r,\min}^{1,t} < 0 \\ \text{and} \\ \hat{f}_{r,\max}^{1,t} > 0 \end{aligned}\right\} \text{for all } r, \tag{A2.7}$$

which from (20) is equivalent to the requirement that

$$\mathbf{f}_{\min}^{1,t} < \mathbf{C}^1(p^t) < \mathbf{f}_{\max}^{1,t}. \tag{A2.8}$$

APPENDIX 3

Variation of p_j with $\boldsymbol{\lambda}$

With (14*a*), the quantities $\partial p_j/\partial \lambda_r$ appearing in (16) may be expressed without approximation as

$$\begin{aligned} \frac{\partial p_j}{\partial \lambda_r} &= \left[\left(\sum_{j'} p_{j'} \hat{f}_{r,j'}^1\right) p_j - \hat{f}_{r,j}^1 p_j\right] \\ &\quad + \boldsymbol{\lambda}.\left[\sum_{j'j''} \hat{\mathbf{f}}_{j'j''}^2(\mathbf{p}) \frac{\partial p_{j''}}{\partial \lambda_r} p_{j'}\right] p_j \\ &\quad - \boldsymbol{\lambda}.\left[\sum_{j'} \hat{\mathbf{f}}_{jj'}^2(\mathbf{p}) \frac{\partial p_{j'}}{\partial \lambda_r}\right] p_j, \end{aligned} \tag{A3.1}$$

$$\begin{aligned} &= \left[\langle \hat{f}_r^1 \rangle - \hat{f}_{r,j}^1\right] p_j \\ &\quad + \boldsymbol{\lambda}.\sum_{j''}\left[\left\langle \hat{\mathbf{f}}_{j''}^2 \frac{\partial p_{j''}}{\partial \lambda_r}\right\rangle - \hat{\mathbf{f}}_{jj''}^2 \frac{\partial p_{j''}}{\partial \lambda_r}\right] p_j, \end{aligned} \tag{A3.2}$$

which, for given $\mathbf{p}(\boldsymbol{\lambda})$ and $\boldsymbol{\lambda}$, may be viewed as a set of linear simultaneous equations for the $\partial p_j/\partial \lambda_r$.

Any exact or approximate solutions to these equations are potentially important since they would yield $\partial P_k/\partial \lambda_r$, and so enable exploration of values of constraints such as 1 and 2 in Table 1 in $\boldsymbol{\lambda}$ space, without the need to carry out a Fourier transformation operation for each new value of $\boldsymbol{\lambda}$.

APPENDIX 4

Convexity of $f_1(\mathbf{p})$

To prove that $f_1(\mathbf{p})$ is a convex function over the convex domain of possible structure vectors, $\mathbf{p}$, it is sufficient (see Rockafellar, 1970, theorem 4.5) to prove that $f_{1,jj'}^2(\mathbf{p})$ is a positive semi-definite quadratic form. To show this, consider arbitrary $\mathbf{p}$ for which

$$\sum_{jj'} p_j f_{1,jj'}^2 p_{j'} = \sum_k \frac{1}{\sigma_k^2} P_k^* P_k \geq 0, \tag{A4.1}$$

and we have used the explicit form for $f_{1,jj'}^2$ given in Table 1.

APPENDIX 5

Sign of λ_r^s

In discussing the convexity of constraint functions and for initializing numerical solution, it is helpful to know the sign of λ_r^s. To this end we note that one always has one explicit solution for $\mathbf{p}(\boldsymbol{\lambda})$, namely the trivial case $\boldsymbol{\lambda} = \mathbf{0}$ in which case $p_j = 1/N$ for all N. Using the $\boldsymbol{\lambda}$-variational principle and (16), one can write

$$\text{sign}\,(\lambda_r^s) = -\,\text{sign}\left[\left(\frac{\partial F}{\partial \lambda_r}\right)_{\lambda=0}\right] = \text{sign}\left[\sum_{j=1}^{N} \hat{f}_{r,j}^{1}(\mathbf{0})\right] \tag{A5.1}$$

and with (20) this becomes

$$\text{sign}\,(\lambda_r^s) = \text{sign}\left[f_r\left(\frac{1}{\mathbf{N}}\right) - C_r\right], \tag{A5.2}$$

which simply says that the sign of λ_r^s is positive if the constraint function value must decrease from the flat map to the desired constraint value and negative otherwise.

References

ABLES, J. G. (1974). *Astron. Astrophys. Suppl. Ser.* **15**, 383–393.

AGMON, N., ALHASSID, Y. & LEVINE, R. D. (1978). In *The Maximum Entropy Formalism*, edited by R. D. LEVINE & M. TRIBUS, pp. 207–210. Massachusetts: MIT Press.

AGMON, N., ALHASSID, Y. & LEVINE, R. D. (1979). *J. Comput. Phys.* **30**, 250–258.

ALHASSID, Y., AGMON, N. & LEVINE, R. D. (1978). *Chem. Phys. Lett.* pp. 22–26.

BRICOGNE, G. (1974). *Acta Cryst.* **A30**, 395–405.

CLAPP, P. C. (1969). *J. Phys. Chem. Solids*, **30**, 2589–2598.

CLAPP, P. C. (1971). *Phys. Rev. B*, **4**, 255–270.

DAINTY, J. C., FIDDY, M. A. & GREENAWAY, A. H. (1979). In *Image Formation from Coherence Functions in Astronomy*, pp. 95–101. Amsterdam: Reidel.

FRIEDEN, R. (1972). *J. Opt. Soc. Am.* **62**, 511–518.

GASSMANN, J. (1977). *Acta Cryst.* A**33**, 474–479.

GULL, S. F. & DANIELL, G. J. (1978). *Nature (London)*, **272**, 686–690.

GULL, S. F. & DANIELL, G. J. (1979). In *Image Formation from Coherence Functions in Astronomy*, edited by C. VAN SCHOONEFELD. Amsterdam: Reidel.

HURLEY, A. C. & MOODIE, A. F. (1980). *Acta Cryst.* A**36**, 737–738.

JAYNES, E. T. (1957). *Phys. Rev.* **106**, 620–630.

JAYNES, E. T. (1963). *Statistical Physics*, pp. 181–187. New York: Benjamin.

JAYNES, E. T. (1979). In *The Maximum Entropy Formalism*, edited by R. D. LEVINE, & M. TRIBUS, pp. 15–118. Massachusetts: MIT Press.

KARLE, J. & HAUPTMANN, H. A. (1950). *Acta Cryst.* **3**, 181–187.

KHINCHIN, A. (1957). *Mathematical Foundations of Information Theory*. New York: Dover.

KIKUCHI, R. & SOFFER, B. H. (1977). *J. Opt. Soc. Am.* **67**, 1656–1665.

KUBO, R. (1962). *J. Phys. Soc. Jpn*, **17**, 1100–1120.

LANDAU, L. D. & LIFSHITZ, E. M. (1980). *Statistical Physics*, ch. XII. Oxford: Pergamon.

LEVINE, R. D. (1980). *J. Phys. A: Math. Nucl. Gen.* **13**, 91–108.

LEVINE, R. D. & TRIBUS, M. (1970). (Editors.) *The Maximum Entropy Formalism*. Massachusetts: MIT Press.

ROCKAFELLAR, R. T. (1970). *Convex Analysis*. Princeton Univ. Press.

SAYRE, D. (1952). *Acta Cryst.* **5**, 60–65.

SHANNON, C. E. (1948). *The Mathematical Theory of Communication, Bell Syst. Tech. J.* **27**, 379–423; 623–656.

SHANNON, C. E. (1949). *Proc. IRE*, **37**, 10–21.

TRIBUS, M. (1969). *Rational, Descriptions, Decisions and Design*. New York: Pergamon.

TSOUCARIS, G. & DE RANGO, C. (1970). *Direct and Patterson Methods of Solving Crystal Structures*. (Introductory Material for NATO Advanced Study Institute held at Parma, Italy.)

WERNECKE, S. J. (1977). *Radio Sci.* **12**, 831–845.

WERNECKE, S. J. & D'ADDARIO, L. R. (1977). *IEEE Trans. Comput.* C-**26**, 351–364.

WILKINS, S. W. (1972). PhD Thesis, Univ. of Melbourne (unpublished).

WILKINS, S. W. (1983*a*). *Acta Cryst.* A. Submitted.

WILKINS, S. W. (1983*b*). *Acta Cryst.* A. Submitted.

WILLINGALE, R. (1981). *Mon. Not. R. Astron. Soc.* **194**, 359–364.

WOOLFSON, M. (1980). In *Computing in Crystallography*, edited by R. DIAMOND, S. RAMASESHAN & K. VENKATAN, pp. 901–911. Bangalore: Indian Academy of Sciences.