*Acta Cryst.* (1979). A35, 266–275

# A Geometrical Approach to Solving Crystal Structures

BY J. NAVAZA* AND A. M. SILVA†

*Universidad Nacional de La Plata, Facultad de Ciencias Exactas, Departamento de Física, CCN* 67, 1900 *La Plata, Argentina*

(*Received* 10 *October* 1977; *accepted* 1 *August* 1978)

## Abstract

The Karle & Hauptman and Goedkoop matrices are regarded as metric tensors in Hilbert space. This enables an $E$ value to be expressed in terms of a finite number of other $E$'s. It is shown that the knowledge of a finite number of $E$'s is theoretically sufficient to determine all the atomic coordinates. This number is smaller than $8(N + 3)$ divided by the order of the point group, counting a complex $E$ twice ($N$ is the number of atoms in the unit cell). The phase problem is analyzed from a geometrical point of view and it is shown how probability becomes certainty with a finite number of data. The $\sum_2$ relationship is obtained as a particular approximation of an exact relationship between $E$'s. This theory enables a criterion to be established which is equivalent to Tsoucaris's maximum determinant rule but more restrictive. The theory is valid for structures with both equal and unequal atoms.

## Introduction

Several authors have studied the information content in Karle & Hauptman matrices by means of Hilbert space (Goedkoop, 1950; Eller, 1955; Kitaigorodsky, 1962; Tsoucaris, 1970). In this paper we discuss the geometrical properties of an $N$-dimensional vector space and show that the usual relationships of direct methods, from Harker & Kasper (1948) inequalities to Tsoucaris's and Podjarny's matrix formulations, can be geometrically derived and interpreted. This approach leads to new results concerning the phase problem and the computation of atomic coordinates.

The theory does not require the usual assumptions of equal atoms, a random atomic distribution, or a large number of atoms in the unit cell. It is assumed that from the $F$ values it is possible to calculate the $E$ values, which in direct space represent point atoms.

*Notation*

$\mathbf{k}$, $\mathbf{h}$, $\mathbf{h}_p$: Reciprocal vectors.
$N$: Number of atoms in the unit cell.
$\mathbf{r}_j$: Vector of coordinates of the $j$th atom.
$z_j$: Number of electrons of the $j$th atom.

$$n_j = z_j \Big/ \left( \sum_{m=1}^{N} z_m^2 \right)^{1/2}.$$

$\delta_{ij}$: Kronecker delta.

$$E(\mathbf{h}) = \sum_{j=1}^{N} n_j \exp(2\pi i \mathbf{h}.\mathbf{r}_j)$$: Normalized structure factor.

$t$: Symmetry number.
$\hat{A}_\alpha = (\mathsf{A}_\alpha | \mathbf{a}_\alpha)$: Each of the symmetry elements of a given space group expressed as an affine transformation, where $\mathsf{A}_\alpha$ is the three-by-three matrix of the point group and $\mathbf{a}_\alpha$ is the translation vector of the transformation.

## 2. The theory

### (*a*) *The Karle & Hauptman matrix as metric tensor*

Consider an $N$-dimensional vector space, $\mathbf{N}$, which associates the vector

$$\mathbf{V}(\mathbf{k}) = \sum_{j=1}^{N} \exp(2\pi i \mathbf{k}.\mathbf{r}_j)\, \mathbf{e}_j \tag{2.1}$$

with each point in reciprocal space. With the scalar product defined as

$$[\mathbf{e}_i | \mathbf{e}_j] = n_j \delta_{ij}, \tag{2.2}$$

then

$$[\mathbf{V}(\mathbf{k}) | \mathbf{V}(\mathbf{h})] = \sum_{i=1}^{N} \sum_{j=1}^{N} \exp(2\pi i \mathbf{h}.\mathbf{r}_j) \times \exp(-2\pi i \mathbf{k}.\mathbf{r}_i)\, [\mathbf{e}_i | \mathbf{e}_j]$$

$$= \sum_{j=1}^{N} n_j \exp(2\pi i (\mathbf{h} - \mathbf{k}).\mathbf{r}_j) = E(\mathbf{h} - \mathbf{k}). \tag{2.3}$$

* Present address: Université de Paris-Sud, Centre Pharmaceutique, Laboratoire de Physique, Tour B, 92290 Chatenay-Malabry, France.

† Present address: Universidad Nacional de Salta, Departamento de Ciencias Exactas, Buenos Aires 177, 4400 Salta, Argentina.

0567-7394/79/020266-10$01.00

For instance, the scalar product $[\mathbf{V}(\mathbf{0})|\mathbf{V}(\mathbf{k})]$ is the normalized structure factor $E(\mathbf{k})$.

Any set of $N$ linearly independent vectors $\mathbf{V}(\mathbf{h}_p)$ $(p = 1, \ldots, N)$ is called a basis; any vector $\mathbf{V}(\mathbf{k})$ can be expressed in a unique way as a linear combination of $\mathbf{V}(\mathbf{h}_p)$'s. The coefficients of such a combination can be put in terms of $E$'s. To do this, the $N$ covariant components $[\mathbf{V}(\mathbf{h}_p)|\mathbf{V}(\mathbf{k})] = E(\mathbf{k} - \mathbf{h}_p)(p = 1, \ldots, N)$ and the metric tensor H, whose general element is

$$[\mathbf{V}(\mathbf{h}_q)|\mathbf{V}(\mathbf{h}_p)] = E(\mathbf{h}_p - \mathbf{h}_q) = H_{qp}, \tag{2.4}$$

are used. This tensor is a Karle & Hauptman matrix (Karle & Hauptman, 1950). Therefore, from the metric tensor and the covariant components the contravariant components can be calculated:

$$C_p = \sum_{q=1}^{N} (\mathsf{H}^{-1})_{pq} E(\mathbf{k} - \mathbf{h}_q). \tag{2.5}$$

Then, the vector $\mathbf{V}(\mathbf{k})$ can be expressed as a linear combination of vectors $\mathbf{V}(\mathbf{h}_p)$. That is,

$$\begin{aligned}\mathbf{V}(\mathbf{k}) &= \sum_{p=1}^{N} C_p \mathbf{V}(\mathbf{h}_p) \\ &= \sum_{p=1}^{N}\sum_{q=1}^{N} (\mathsf{H}^{-1})_{pq} E(\mathbf{k} - \mathbf{h}_q)\, \mathbf{V}(\mathbf{h}_p).\end{aligned} \tag{2.6}$$

Since $[\mathbf{V}(\mathbf{0})|\mathbf{V}(\mathbf{k})] = E(\mathbf{k})$ and $[\mathbf{V}(\mathbf{0})|\mathbf{V}(\mathbf{h}_p)] = E(\mathbf{h}_p)$, the scalar product of (2.6) by $\mathbf{V}(\mathbf{0})$ is

$$E(\mathbf{k}) = \sum_{p=1}^{N}\sum_{q=1}^{N} (\mathsf{H}^{-1})_{pq} E(\mathbf{k} - \mathbf{h}_q) E(\mathbf{h}_p). \tag{2.7}$$

This is the expression of an $E$ value in terms of a finite number of other $E$'s.

The scalar product $[(\mathbf{V}(\mathbf{k})|\mathbf{V}(\mathbf{k})]$ defines the squared norm of $\mathbf{V}(\mathbf{k})$. From (2.3) it is clear that

$$\|\mathbf{V}(\mathbf{k})\|^2 = [\mathbf{V}(\mathbf{k})|\mathbf{V}(\mathbf{k})] = E(\mathbf{0}), \tag{2.8}$$

and from (2.6)

$$\|\mathbf{V}(\mathbf{k})\|^2 = \sum_{p=1}^{N}\sum_{q=1}^{N} (\mathsf{H}^{-1})_{pq} E(\mathbf{k} - \mathbf{h}_q) E(\mathbf{h}_p - \mathbf{k}). \tag{2.9}$$

Therefore, all vectors have equal norm $[E(\mathbf{0})]^{1/2}$, but they are at different 'angles' with $\mathbf{V}(\mathbf{0})$. If these angles are small the vectors have large projections on $\mathbf{V}(\mathbf{0})$. Therefore, they are related to $E$'s of large modulus and *vice versa*.

Any set of $s$ linearly independent vectors $\mathbf{V}(\mathbf{h}_p)$ $(p = 1, \ldots, s)$, is a basis for the linear subspace, $\mathbf{S}$, which they span. The projection of $\mathbf{V}(\mathbf{k})$ on $\mathbf{S}$ will be called $\mathbf{V}_s(\mathbf{k})$. Then, $\mathbf{V}(\mathbf{k})$ may be written as the sum of $\mathbf{V}_s(\mathbf{k})$ plus its orthogonal complement $\mathbf{P}(\mathbf{k})$,

$$\mathbf{V}(\mathbf{k}) = \mathbf{V}_s(\mathbf{k}) + \mathbf{P}(\mathbf{k}). \tag{2.10}$$

Then

$$[\mathbf{V}(\mathbf{h}_p)|\mathbf{V}(\mathbf{k})] = [\mathbf{V}(\mathbf{h}_p)|\mathbf{V}_s(\mathbf{k})] \quad (p = 1, \ldots, s), \tag{2.11}$$

because $\mathbf{P}(\mathbf{k})$ is orthogonal to $\mathbf{S}$. Equation (2.11) means that the covariant components of $\mathbf{V}_s(\mathbf{k})$ are equal to the covariant components of $\mathbf{V}(\mathbf{k})$ in the basis which defines $\mathbf{S}$. As $\mathbf{S}$ is a vector space the preceding development is totally applicable to it. Therefore, using (2.11) we can write by analogy with (2.6)

$$\mathbf{V}_s(\mathbf{k}) = \sum_{p=1}^{s}\sum_{q=1}^{s} (\mathsf{H}_s^{-1})_{pq} E(\mathbf{k} - \mathbf{h}_q)\, \mathbf{V}(\mathbf{h}_p), \tag{2.12}$$

where $\mathsf{H}_s$ means that the Karle & Hauptman matrix is of order $s$. The scalar product by $\mathbf{V}(\mathbf{0})$ is

$$[\mathbf{V}(\mathbf{0})|\mathbf{V}_s(\mathbf{k})] = E_s(\mathbf{k}) = \sum_{p=1}^{s}\sum_{q=1}^{s} (\mathsf{H}_s^{-1})_{pq} E(\mathbf{k} - \mathbf{h}_q) E(\mathbf{h}_p). \tag{2.13}$$

$E_s(\mathbf{k})$ is a geometric approximation of order $s$ to $E(\mathbf{k})$, and $E_N(\mathbf{k}) = E(\mathbf{k})$.

To estimate how $E_s(\mathbf{k})$ approaches $E(\mathbf{k})$ it is necessary to calculate the squared norm of $\mathbf{V}_s(\mathbf{k})$. As $\mathbf{V}_s(\mathbf{k})$ and $\mathbf{P}(\mathbf{k})$ are orthogonal vectors it is clear from (2.10) that $[\mathbf{V}_s(\mathbf{k})|\mathbf{V}(\mathbf{k})] = [\mathbf{V}_s(\mathbf{k})|\mathbf{V}_s(\mathbf{k})]$; therefore

$$\begin{aligned}\|\mathbf{V}_s(\mathbf{k})\|^2 &= [\mathbf{V}_s(\mathbf{k})|\mathbf{V}_s(\mathbf{k})] = [\mathbf{V}(\mathbf{k})|\mathbf{V}_s(\mathbf{k})] \\ &= \sum_{p=1}^{s}\sum_{q=1}^{s} (\mathsf{H}_s^{-1})_{pq} E(\mathbf{k} - \mathbf{h}_q) E(\mathbf{h}_p - \mathbf{k}).\end{aligned} \tag{2.14}$$

As $E(\mathbf{k})$ and $E_s(\mathbf{k})$ may be interpreted as the projections of $\mathbf{V}(\mathbf{k})$ and $\mathbf{V}_s(\mathbf{k})$ on $\mathbf{V}(\mathbf{0})$, respectively, then it is natural to think that $\|\mathbf{V}_s(\mathbf{k})\|^2$ compared with its upper bound $E(\mathbf{0})$ may be an estimate of the accuracy of $E_s(\mathbf{k})$ as an approximation of $E(\mathbf{k})$.

A fundamental property of the squared norms $\|\mathbf{V}_s(\mathbf{k})\|^2$ is that they form a non-decreasing sequence when $s$ increases. Consider the subspaces of $\mathbf{N}$, $\mathbf{S}$ and $\mathbf{S} + 1$, being $\mathbf{N} \supset \mathbf{S} + 1 \supset \mathbf{S}$ and the vectors $\mathbf{V}_s(\mathbf{k}) \in \mathbf{S}$ and $\mathbf{V}_{s+1}(\mathbf{k}) \in \mathbf{S} + 1$. As usual $\mathbf{V}_s(\mathbf{k})$ and $\mathbf{V}_{s+1}(\mathbf{k})$ are the projections of $\mathbf{V}(\mathbf{k})$ on $\mathbf{S}$ and $\mathbf{S} + 1$ respectively. But, if $\mathbf{S} + 1 \supset \mathbf{S}$ it is clear that $\mathbf{V}_s(\mathbf{k})$ is also the projection of $\mathbf{V}_{s+1}(\mathbf{k})$ on $\mathbf{S}$; then

$$\mathbf{V}_{s+1}(\mathbf{k}) = \mathbf{V}_s(\mathbf{k}) + \mathbf{W}(\mathbf{k}). \tag{2.15}$$

Since $\mathbf{W}(\mathbf{k})$ must be orthogonal to $\mathbf{S}$, $[\mathbf{V}_s(\mathbf{k})|\mathbf{W}(\mathbf{k})] = 0$. Therefore

$$\|\mathbf{V}_{s+1}(\mathbf{k})\|^2 = \|\mathbf{V}_s(\mathbf{k})\|^2 + \|\mathbf{W}(\mathbf{k})\|^2. \tag{2.16}$$

Then,

$$\|\mathbf{V}_{s+1}(\mathbf{k})\|^2 \geq \|\mathbf{V}_s(\mathbf{k})\|^2, \tag{2.17}$$

which proves that the norms form a non-decreasing sequence when $s$ increases. The upper bound of this

sequence is $E(\mathbf{0})$ for $s = N$ and the lower bound is $|E(\mathbf{h}_1 - \mathbf{k})|^2/E(\mathbf{0})$ for $s = 1$.

*(b) Goedkoop's matrices as metric tensors*

By taking symmetry into account it is possible to reduce the dimensionality of the vector space, defining

$$\mathbf{V}(\mathbf{k}) = \sum_{j=1}^{N/t} \sum_{\alpha=1}^{t} \exp(2\pi i \mathbf{k}\hat{A}_\alpha \mathbf{r}_j)\mathbf{e}_j, \qquad (2.18)$$

where the $\mathbf{e}_j$'s satisfy (2.2). It is shown in Appendix *A* that the scalar product is now

$$[\mathbf{V}(\mathbf{k})|\mathbf{V}(\mathbf{h})] = \sum_{\alpha=1}^{t} \exp(2\pi i \mathbf{a}_\alpha.\mathbf{h})\, E(\mathbf{h}\mathsf{A}_\alpha - \mathbf{k}). \qquad (2.19)$$

That is, the scalar product is equal to a linear combination of $E$'s.

From (2.19) it is evident that $[\mathbf{V}(\mathbf{k})|\mathbf{V}(\mathbf{0})] = tE(-\mathbf{k})$ and it is shown in Appendix *A* that

$$[\mathbf{V}(\mathbf{0})|\mathbf{V}(\mathbf{k})] = tE(\mathbf{k}). \qquad (2.20)$$

Using (2.19) and (2.20) it is possible to develop a set of equations similar to those in the preceding section, with a similar interpretation. The main advantage is that symmetry reduces the dimension of the vector space and therefore the order of the matrices involved in the equations. The main results are as follows.

$\mathbf{V}(\mathbf{h}_p)$: The basis of the $(N/t)$ space, where $\mathbf{h}_p$ belongs to the non-symmetry-related Miller indices set.

$[\mathbf{V}(\mathbf{h}_p)|\mathbf{V}(\mathbf{k})]$: Covariant components of $\mathbf{V}(\mathbf{k})$ in the basis $\mathbf{V}(\mathbf{h}_p)$; they can be calculated from (2.19).

$$G_{pq} = [\mathbf{V}(\mathbf{h}_p)|\mathbf{V}(\mathbf{h}_q)] = \sum_{\alpha=1}^{t} \exp(2\pi i \mathbf{a}_\alpha.\mathbf{h}_q)\, E(\mathbf{h}_q\mathsf{A}_\alpha - \mathbf{h}_p):$$

An element of the metric tensor, which is a Goedkoop's matrix corresponding to the totally symmetric representation (Goedkoop, 1950). The expression of $\mathbf{V}(\mathbf{k})$ in the $\mathbf{V}(\mathbf{h}_p)$ basis is

$$\mathbf{V}(\mathbf{k}) = \sum_{p=1}^{N/t} \sum_{q=1}^{N/t} \sum_{\alpha=1}^{t} (\mathsf{G}^{-1})_{pq} \exp(2\pi i \mathbf{a}_\alpha.\mathbf{k}) \times E(\mathbf{k}\mathsf{A}_\alpha - \mathbf{h}_q)\, \mathbf{V}(\mathbf{h}_p). \qquad (2.21)$$

The projection on $\mathbf{V}(\mathbf{0})$ is

$$E(\mathbf{k}) = \sum_{p=1}^{N/t} \sum_{q=1}^{N/t} \sum_{\alpha=1}^{t} (\mathsf{G}^{-1})_{pq} \exp(2\pi i \mathbf{a}_\alpha.\mathbf{k}) \times E(\mathbf{k}\mathsf{A}_\alpha - \mathbf{h}_q)\, E(\mathbf{h}_p). \qquad (2.22)$$

The squared norm of $\mathbf{V}(\mathbf{k})$ is

$$\|\mathbf{V}(\mathbf{k})\|^2 = \sum_{\alpha=1}^{t} \exp(2\pi i \mathbf{a}_\alpha.\mathbf{k})\, E[\mathbf{k}(\mathsf{A}_\alpha - \mathsf{I})], \qquad (2.23)$$

$\mathsf{I}$ being the three-by-three identity matrix. It is shown in Appendix *A* that (2.23) is real and non-negative as expected. From (2.23) it is evident that the vectors have equal norms only in $P_1$.

The projection of $\mathbf{V}(\mathbf{k})$ on a subspace gives rise to the following equations:

$$\mathbf{V}_s(\mathbf{k}) = \sum_{p=1}^{s} \sum_{q=1}^{s} \sum_{\alpha=1}^{t} (\mathsf{G}_s^{-1})_{pq} \exp(2\pi i \mathbf{a}_\alpha.\mathbf{k}) \times E(\mathbf{k}\mathsf{A}_\alpha - \mathbf{h}_q)\, \mathbf{V}(\mathbf{h}_p), \qquad (2.24)$$

$$E_s(\mathbf{k}) = \sum_{p=1}^{s} \sum_{q=1}^{s} \sum_{\alpha=1}^{t} (\mathsf{G}_s^{-1})_{pq} \exp(2\pi i \mathbf{a}_\alpha.\mathbf{k}) \times E(\mathbf{k}\mathsf{A}_\alpha - \mathbf{h}_q)\, E(\mathbf{h}_p), \qquad (2.25)$$

$$\|\mathbf{V}_s(\mathbf{k})\|^2 = \sum_{p=1}^{s} \sum_{q=1}^{s} \sum_{\alpha=1}^{t} \sum_{\beta=1}^{t} (\mathsf{G}_s^{-1})_{pq} \exp[2\pi i \mathbf{k}.(\mathbf{a}_\alpha - \mathbf{a}_\beta)] \times E(\mathbf{k}\mathsf{A}_\alpha - \mathbf{h}_q)\, E(\mathbf{h}_p - \mathbf{k}\mathsf{A}_\beta). \qquad (2.26)$$

The property that the norms form a non-decreasing sequence when $s$ increases is also valid.

## 3. The phase determination

*(a) A geometric approach to the phase extension problem*

Equation (2.25) – which includes (2.22), (2.7) and (2.13) as particular cases – may be used for phase extension. Then, the phase of $E_s(\mathbf{k})$ would be accepted if $|E_s(\mathbf{k})|$ approaches $|E(\mathbf{k})|$. But the natural geometric estimate is $\|\mathbf{V}_s(\mathbf{k})\|^2$. Therefore the squared norm of the orthogonal complement of $\mathbf{V}_s(\mathbf{k})$,

$$\|\mathbf{V}(\mathbf{k}) - \mathbf{V}_s(\mathbf{k})\|^2 = \|\mathbf{V}(\mathbf{k})\|^2 - \|\mathbf{V}_s(\mathbf{k})\|^2, \qquad (3.1)$$

and the difference $|E(\mathbf{k})| - |E_s(\mathbf{k})|$ are two possible estimates of the accuracy of the phase of $E_s(\mathbf{k})$. It is worth noticing that if $E_s(\mathbf{k})$ is calculated from (2.25), no new phased $E$'s are needed for calculating $\|\mathbf{V}_s(\mathbf{k})\|^2$.

Even if the norm of the orthogonal complement of $\mathbf{V}_s(\mathbf{k})$ is small, the accuracy of the phase prediction is not guaranteed if $|E(\mathbf{k})|$ is small. Since, if $|E(\mathbf{k})|$ is small, $\mathbf{V}(\mathbf{k})$ makes a large angle with $\mathbf{V}(\mathbf{0})$ and then the orthogonal complement of $\mathbf{V}_s(\mathbf{k})$ is not necessarily compelled to maintain the direction of $\mathbf{V}_s(\mathbf{k})$ [with respect to $\mathbf{V}(\mathbf{0})$] in order to obtain the correct projection on $\mathbf{V}(\mathbf{0})$.

Recently, equation (2.25) has been used for phase extension in proteins (Podjarny, Yonath & Traub, 1976) and nucleic acids (Podjarny & Yonath, 1977). These authors derived it from probabilistic arguments, but they cannot justify the validity in unequal-atom structures (Podjarny & Yonath, 1977).

*(b) The minimal norm criterion*

Inequality (2.17) provides new constraints on the phases. With the correct phases, the norm of a vector must be non-decreasing when $s$ increases. If a method were derived from this principle, its maximum sensitivity would be obtained when the gap between the upper and lower bounds of the sequence is a minimum. Then, for $s = 1$ we should choose $E(\mathbf{h}_1 - \mathbf{k})$ as the largest $E$.

It is interesting to note that the condition $\|\mathbf{V}_2(\mathbf{0})\|^2 \leq E(\mathbf{0})$ leads to $|U(\mathbf{h})|^2 \leq (\frac{1}{2}) + (\frac{1}{2})\, U(2\mathbf{h})$ if the base vectors $\mathbf{V}(\mathbf{h})$ and $\mathbf{V}(-\mathbf{h})$ are chosen in a centrosymmetric structure (Harker & Kasper, 1948).

Another important result can be derived by equating the phases which minimize $\|\mathbf{V}_s(\mathbf{0})\|^2$. The phases of $\mathsf{H}_s$ are known. From (2.14) the squared norm of $\mathbf{V}_s(\mathbf{0})$ is

$$\|\mathbf{V}_s(\mathbf{0})\|^2 = \sum_{p,q}^{s} D_{pq} E_p E_q^*, \tag{3.2}$$

where $\mathsf{H}_s^{-1} = \mathsf{D}$ and $E(\mathbf{h}_p) = E_p$. The phases of $E_p$ which minimize this quadratic form can be calculated by means of the partial derivative of $\|\mathbf{V}_s(\mathbf{0})\|^2$ with respect to $\varphi_m$

$$\begin{aligned}\frac{\partial \|\mathbf{V}_s(\mathbf{0})\|^2}{\partial \varphi_m} &= i\left(\sum_{p,q}^{s} D_{pq}\,\delta_{pm} E_p E_q^* - \sum_{p,q}^{s} D_{pq}\,\delta_{qm}\, E_p E_q^*\right) \\ &= 2i\,\mathrm{Im}\left(\sum_{\substack{q=1 \\ q\neq m}}^{s} D_{mq} E_m E_q^*\right).\end{aligned} \tag{3.3}$$

All the first partial derivatives must be zero in an extremum, then

$$\mathrm{Im}\left(\sum_{\substack{q=1 \\ q\neq m}}^{s} D_{mq} E_m E_q^*\right) = 0 \quad (m = 1, \ldots, s). \tag{3.4}$$

The solutions of these equations are

$$\text{phase}\,(E_m) = \text{phase}\left(\sum_{\substack{q=1 \\ q\neq m}}^{s} D_{qm} E_q\right) (m = 1, \ldots, s) \tag{3.5}$$

and

$$\text{phase}\,(E_m) = \text{phase}\left(-\sum_{\substack{q=1 \\ q\neq m}}^{s} D_{qm} E_q\right) \quad (m = 1, \ldots, s). \tag{3.6}$$

In Appendix $B$ it is shown that the matrix of the second-order partial derivatives, C, is non-negative at the point defined by (3.6). Therefore, the extremum is a relative minimum.

In Appendix $C$ it is shown that

$$-\sum_{\substack{q=1 \\ q\neq m}}^{s} D_{qm} E_q = \sum_{p=1}^{s-1}\sum_{q=1}^{s-1} (\mathsf{H}_{s-1}^{-1})_{pq}\, E(\mathbf{h}_m - \mathbf{h}_q)\, E(\mathbf{h}_p) \times \frac{\det(\mathsf{H}_{s-1})}{\det(\mathsf{H}_s)}, \tag{3.7}$$

where $\mathsf{H}_{s-1}$ is obtained from $\mathsf{H}_s$ suppressing in it the $m$th row and the $m$th column. Therefore, the phases which minimize the quadratic form (3.2) are solutions of the system

$$\text{phase}\,[E(\mathbf{h}_m)] = \text{phase}\left[\sum_{p=1}^{s-1}\sum_{q=1}^{s-1} (\mathsf{H}_{s-1}^{-1})_{pq}\, E(\mathbf{h}_m - \mathbf{h}_q) \times E(\mathbf{h}_p)\right] \quad (m = 1, \ldots, s). \tag{3.8}$$

This system expresses one phase in terms of the $s - 1$ other phases. In other words, the maximum consistencies between phases compatible with the information content in the Karle & Hauptman matrix is obtained when the phases minimize the quadratic form (3.2). For $s$ large it would be expected that 'maximum consistence between phases' means 'correct phases'. But the non-decreasing norm property provides the auxiliary criterion to test if these phases are the correct ones. This may be synthesized in the minimal norm criterion. 'A necessary condition for the phases that minimize $\|\mathbf{V}_s(\mathbf{0})\|^2$ to be the correct ones is that the sequence of norms be non-decreasing' (see Fig. 1).

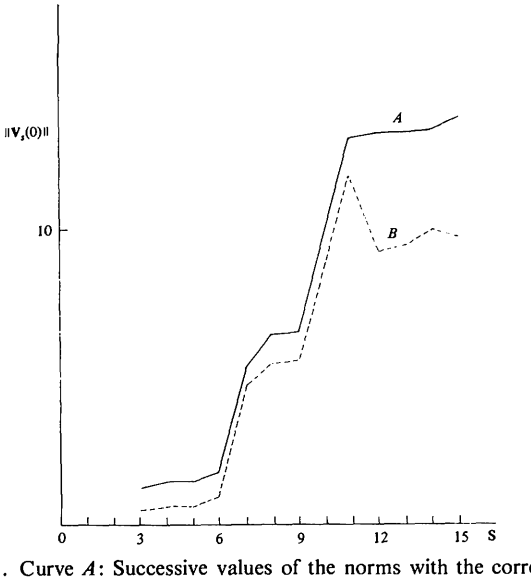


Fig. 1. Curve $A$: Successive values of the norms with the correct signs for the $E$'s, corresponding to a centrosymmetric structure with $N = 52$. Curve $B$: The same as $A$ except that the signs are those which correspond to a minimum of the norms. We note that curve $B$ fails to be a non-decreasing sequence, a certain indication that something is wrong.

As was pointed out before, the extremum is a relative minimum because C is non-negative, but it would be a strict relative minimum only if C is positive definite. Then, if C is positive definite the function takes its minimum value at one unique point and if C is non-negative the function could take the same minimum value at a set of points. This means that more than one set of phases could satisfy system (3.8). This is a reasonable result bearing in mind that the equations involved in this system are transcendental. Uniqueness of the solution should not then be expected. Let us consider $s = 2$. In this case $\det(\mathsf{C}) = 0$ and system (3.8) reduces to

$$\begin{aligned}\text{phase}[E(\mathbf{h}_1)] &= \text{phase}[E(\mathbf{h}_1 - \mathbf{h}_2)E(\mathbf{h}_2)], \\ \text{phase}[E(\mathbf{h}_2)] &= \text{phase}[E(\mathbf{h}_2 - \mathbf{h}_1)E(\mathbf{h}_1)].\end{aligned} \tag{3.9}$$

This system does not unequivocally determine $\varphi(\mathbf{h}_1)$ and $\varphi(\mathbf{h}_2)$ because it can be reduced to the relationship $-\varphi(\mathbf{h}_1) + \varphi(\mathbf{h}_1 - \mathbf{h}_2) + \varphi(\mathbf{h}_2) = 0$.

### (c) *Sayre's equation*

If the Karle & Hauptman matrix is written as

$$\mathsf{H} = E(\mathbf{0})(\mathsf{I} + \mathsf{B}), \tag{3.10}$$

where I is the identity matrix, $\mathsf{H}^{-1}$ may be expanded in powers of B; that is

$$\mathsf{H}^{-1} = \frac{1}{E(\mathbf{0})} \sum_{m=0}^{\infty} (-1)^m \mathsf{B}^m. \tag{3.11}$$

This series will be convergent if the absolute values of the eigenvalues of B are less than one. Calling $\Lambda_{\mathsf{B}}$ the eigenvalue matrix of B, this condition reads $-\mathsf{I} < \Lambda_{\mathsf{B}} < \mathsf{I}$, where the inequality thus written means inequalities between the corresponding diagonal terms. If B is an Hermitian matrix, A is its modal matrix and $\mathsf{A}^\dagger$ the Hermitian transpose of A, then $\mathsf{B} = \mathsf{A}\,\Lambda_{\mathsf{B}}\,\mathsf{A}^\dagger$. From (3.10) it follows that

$$\mathsf{H} = E(\mathbf{0})(\mathsf{I} + \mathsf{A}\,\Lambda_{\mathsf{B}}\,\mathsf{A}^\dagger), = \mathsf{A}E(\mathbf{0})(\mathsf{I} + \Lambda_{\mathsf{B}})\,\mathsf{A}^\dagger, \tag{3.12}$$

because $\mathsf{A}^\dagger\mathsf{A} = \mathsf{A}\mathsf{A}^\dagger = \mathsf{I}$. Therefore, calling $\Lambda_{\mathsf{H}}$ the eigenvalue matrix of H, from (3.12) it follows that

$$\mathsf{A}^\dagger\mathsf{H}\mathsf{A} = E(\mathbf{0})(\mathsf{I} + \Lambda_{\mathsf{B}}) = \Lambda_{\mathsf{H}}. \tag{3.13}$$

Then, the convergence condition of series (3.11), $-\mathsf{I} < \Lambda_{\mathsf{B}} < \mathsf{I}$, can be put in terms of $\Lambda_{\mathsf{H}}$:

$$0 < \Lambda_{\mathsf{H}} < 2E(\mathbf{0})\mathsf{I}. \tag{3.14}$$

These inequalities mean that the series will be convergent to $\mathsf{H}^{-1}$ if H is positive definite and its eigenvalues are less than $2E(\mathbf{0})$. The convergence will be fast if the eigenvalues approach $E(\mathbf{0})$, which is what would be expected if the off-diagonal elements of H were small. In this case $\mathsf{H}^{-1} \simeq [1/E(\mathbf{0})]\mathsf{I}$ and then (2.13) reads

$$E_s(\mathbf{k}) = \sum_{p=1}^{s} \frac{1}{E(\mathbf{0})} E(\mathbf{k} - \mathbf{h}_p)E(\mathbf{h}_p), \tag{3.15}$$

which resembles a partial sum of a Sayre's (1952) equation. However, there is an essential difference between both approximations. Sayre's hypothesis assumes equal atoms and if the atoms are very sharp, which is the case for (3.15), the sum would contain a very large number of terms. Several authors have derived relations similar to (3.15); for example, Hauptman & Karle (1953) and Karle & Hauptman (1956), giving probabilistic meaning to this partial sum, called it $\sum_2$. But, in general, these approaches need to assume that some sums have zero average and that the atoms are equal; then as the partial sums do not obey these requirements they were interpreted as probable relations. Our approximation only assumes small off-diagonal elements in the Karle & Hauptman matrix related to the $E(\mathbf{h}_p)$'s; then the sum may be convergent although it includes a few terms. That is, by a simple inspection of the Karle & Hauptman matrix we would have a notion about the accuracy of the approximation and we could improve it by choosing suitable $E(\mathbf{h}_p)$'s.

If the terms for $m = 0$ and $m = 1$ are considered in (3.11), equation (2.13) becomes

$$\begin{aligned}E_s(\mathbf{k}) = 2\sum_{p=1}^{s} \frac{1}{E(\mathbf{0})} E(\mathbf{k} - \mathbf{h}_p)E(\mathbf{h}_p) \\ - \sum_{p=1}^{s}\sum_{q=1}^{s} \frac{1}{E(\mathbf{0})^2} E(\mathbf{k} - \mathbf{h}_p)E(\mathbf{h}_p - \mathbf{h}_q)E(\mathbf{h}_q).\end{aligned} \tag{3.16}$$

This resembles equation (2) of the phase-correction method of Hoppe & Gassmann (1968), but differs from it in multiplicative constants. The main remark is that the difference between the hypothesis leading to (2) and (3.16) is analogous to that pointed out between Sayre's hypothesis and the approximation which leads to (3.15).

### (d) *The statistical theory*

The usual probabilistic formulae may be obtained, with a Karle & Hauptman or a Goedkoop matrix as covariance matrix, from the multi-dimensional distribution law

$$\mathsf{P}(E_1, \ldots, E_s) = K_1 \exp\left(-\sum_{i,j}^{s} D_{ij}E_iE_j^*\right) \tag{3.17}$$

(Tsoucaris, 1970; Castellano, Podjarny & Navaza, 1973). These authors build the covariance matrix with $U$'s. The covariance matrix in (3.17) is built with $E$'s, and differs by a constant which is unimportant in the following analysis of the distribution laws.

Assuming that $E_i (i = 1, \ldots, s;\ i \neq m)$ and $\mathsf{H}_s$ are known, the distribution law of the $m$th structure factor is

$$P[E_m | E_1, \ldots, E_i, \ldots, E_s;\ (i \neq m)] = K_2 \exp\left(-\frac{|E_m - \overline{E_m}|^2}{\sigma^2}\right), \quad (3.18)$$

where $K_1$ and $K_2$ are normalization constants, and

$$\overline{E_m} = -\frac{1}{D_{mm}} \sum_{\substack{p=1 \\ p \neq m}}^{s} D_{pm} E_p, \quad (3.19)$$

$$\sigma^2 = \frac{1}{D_{mm}} \quad (3.20)$$

(de Rango, Tsoucaris & Zelwer, 1974). In Appendix *C* it is proved that

$$\sigma^2 = \frac{1}{D_{mm}} = \frac{\det(\mathsf{H}_s)}{\det(\mathsf{H}_{s-1})} = \|\mathbf{V}(\mathbf{h}_m) - \mathbf{V}_{s-1}(\mathbf{h}_m)\|^2 \quad (3.21)$$

and from (*C*.6) it follows that

$$\overline{E_m} = \sum_{p=1}^{s-1} \sum_{q=1}^{s-1} (\mathsf{H}_{s-1}^{-1})_{pq} E(\mathbf{h}_m - \mathbf{h}_q) E(\mathbf{h}_p) = E_{s-1}(\mathbf{h}_m). \quad (3.22)$$

Therefore, the probabilistic parameters may be geometrically interpreted. The expected value is the geometric $s - 1$ approximation and the variance is the norm of the orthogonal complement of $\mathbf{V}_{s-1}(\mathbf{h}_m)$. $\sigma^2$ forms a non-increasing sequence when $s$ increases and for $s = N + 1$, $E_m = E_N(\mathbf{h}_m) = E(\mathbf{h}_m)$, $\sigma^2 = 0$ and *probability becomes certainty.*

From (3.18) it is evident that the distribution law of the phase of $E_m$ is

$$P(\varphi_m) = K_2 \exp\left\{\frac{2|E_m||\overline{E_m}|\cos[\varphi_m - \varphi(\overline{E_m})]}{\sigma^2}\right\} \quad (3.23)$$

(De Rango *et al.*, 1974). The reliability of the phase prediction depends not only on the orthogonal complement norm but also on the modulus of the $E$'s. Equation (3.23) contains these two magnitudes in the correct way because the numerator in the argument is of the order of $|E(\mathbf{h}_m)|^2$ and the denominator is equal to the orthogonal complement squared norm.

On the other hand, the quadratic form in (3.17) is $\|\mathbf{V}_s(\mathbf{0})\|^2$. There is a close connexion between Tsoucaris's (1970) maximum determinant rule and our criterion of norm minimization. It follows that this rule is derivable and interpretable geometrically; that is, without probabilistic arguments. It is notable that Tsoucaris's formulation leads to a Karle & Hauptman matrix whose elements correspond to the squared density; therefore this formulation is strictly valid only in the equal-atoms case. Moreover, the properties of norms impose a far more restrictive condition than that of the non-increasing sequence of Tsoucaris's $D_m$ determinants.

As the Tsoucaris's $D_s$ determinants are built with $U$'s, then $D_s = [\det(\mathsf{H}_s)]/[E(\mathbf{0})]^s$. Therefore, from (3.21) it follows that

$$\frac{D_s}{D_{s-1}} = \frac{\det(\mathsf{H}_s)}{\det(\mathsf{H}_{s-1}^m) E(\mathbf{0})} = \frac{\|\mathbf{V}(\mathbf{h}_m) - \mathbf{V}_{s-1}(\mathbf{h}_m)\|^2}{\|\mathbf{V}(\mathbf{h}_m)\|^2}, \quad (3.24)$$

where $\mathsf{H}_{s-1}^m$ is the matrix obtained from $\mathsf{H}_s$ by suppressing the $m$th row and the $m$th column. In this case $\mathsf{H}_s$ is built in such a way that the $m$th row and the $m$th column are the last. Tsoucaris's condition $(D_s/D_{s-1}) \leq 1$ is equivalent to $\|\mathbf{V}(\mathbf{h}_m) - \mathbf{V}_{s-1}(\mathbf{h}_m)\|^2 \leq \|\mathbf{V}(\mathbf{h}_m)\|^2$; that is, the squared norm of the orthogonal complement is compared with its upper bound. But, inequality (2.17) also implies that

$$\|\mathbf{V}(\mathbf{h}_m) - \mathbf{V}_s(\mathbf{h}_m)\|^2 \leq \|\mathbf{V}(\mathbf{h}_m) - \mathbf{V}_{s-1}(\mathbf{h}_m)\|^2; \quad (3.25)$$

therefore

$$\frac{\det(\mathsf{H}_{s+1})}{\det(\mathsf{H}_s^m)} \leq \frac{\det(\mathsf{H}_s)}{\det(\mathsf{H}_{s-1}^m)}. \quad (3.26)$$

Then, not only is the sequence of determinants non-increasing but the sequence of quotients is also non-increasing. Note that the matrices from which the quotients are built differ always by the $m$th row and the $m$th column; therefore $\mathsf{H}_s^m \neq \mathsf{H}_s$.

Karle's (1971) generalized tangent formula can now be interpreted from a geometrical point of view. This formula is based on the quotient of determinants $\delta_{m,p}(\mathbf{h}) = \Delta'_{m,p}(\mathbf{h})/\Delta_{m,p}$ [equations (6), (8) and (9) of Karle (1971)]. $\Delta_{m,p}$ differs from $\Delta'_{m,p}(\mathbf{h})$ by the suppression of the first column and the last row. Then, developing $\Delta'_{m,p}(\mathbf{h})$ along the first column and along the last row, and bearing in mind the reasonings in Appendix *C*, it follows that

$$\delta_{m,p}(\mathbf{h}) = \frac{\Delta'_{m,p}(\mathbf{h})}{\Delta_{m,p}} = (-1)^m \sum_{u=1}^{m-2} \sum_{v=1}^{m-2} (\mathsf{H}_{m-2}^{-1})_{uv} E(\mathbf{h} - \mathbf{k}_v) E(\mathbf{k}_u) = (-1)^m E_{m-2}^p(\mathbf{h}). \quad (3.27)$$

Therefore, $\delta_{m,p}(\mathbf{h})$ is proportional to $E_{m-2}^p(\mathbf{h})$; that is, it is proportional to the projection on $\mathbf{V}(\mathbf{0})$ of the projection of $\mathbf{V}(\mathbf{h})$ on the $p$th subspace of dimension $m - 2$. The index $p$ labels a particular subspace of dimension $m - 2$ (Karle, 1971).

The assumption $E(\mathbf{h}) \propto \langle \delta_{m,p}(\mathbf{h}) \rangle_p$ [equation (11) of Karle (1971)] is equivalent to $E(\mathbf{h}) \propto \langle E^p_{m-2}(\mathbf{h}) \rangle_p$. The generalized tangent formula is an average of geometric approximations on subspaces of the same dimension. From the geometric approach it is evident that the element of the set $\{E^1_m(\mathbf{h}), E^2_m(\mathbf{h}), \ldots, E^p_m(\mathbf{h}), \ldots, E^n_m(\mathbf{h})\}$ which gives the best phase indication is the element associated with the vector $\mathbf{V}^p_m(\mathbf{h})$ with the largest norm. The advantage of averaging this set is not clear, except to take into account inaccuracies in the data.

## 4. Determination of atomic coordinates

Equation (2.6) is a vectorial relationship; it is therefore an equality between components

$$\exp(2\pi i \mathbf{k}.\mathbf{r}) = \sum_{p=1}^{N} \sum_{q=1}^{N} (\mathsf{H}^{-1})_{pq} E(\mathbf{k} - \mathbf{h}_q) \times \exp(2\pi i \mathbf{h}_p.\mathbf{r}), \qquad (4.1)$$

where $\mathbf{r} = \mathbf{r}_j (j = 1, \ldots, N)$. This is the equation of a surface in the variables $(x,y,z) = \mathbf{r}$, which contains all the atomic positions. If three such surfaces are constructed, changing $\mathbf{k}$ in (4.1), the atoms will be at their intersections. These intersections will define, except in special cases, a finite number of points not all of which will correspond to atomic coordinates. The number of spurious points depends strongly on the basis selected and on the vector chosen to be expanded. It has been observed that this number increases with increasing Miller indices (see Fig. 2).

In linear structures the highest frequency in the trigonometric function in (4.1) is at least $N/t$. As we see in Fig. 2 all the zeros correspond to atomic positions with the lowest frequencies. For this structure the way of choosing the minimal frequencies is unique, while for two-dimensional (see Fig. 3) and three-dimensional structures there is more freedom in the selection of Miller indices smaller than $N/t$.

Choosing the lowest frequencies the smallest number of different $E$'s needed to determine the atomic coordinates in linear structures is $N$. For three-dimensional structures the minimum is also obtained choosing the lowest frequencies for the basis vectors; that is, selecting those $N + 3$ reciprocal vectors which are nearest to the origin. It means that they are all within a sphere of radius $R$, equal to the greatest modulus of the selected reciprocal vectors. As the density of points in reciprocal space is a constant equal to the volume $V$ of the unit cell in direct space, then

$$N + 3 = (\tfrac{4}{3})\pi V R^3. \qquad (4.2)$$

The differences among these vectors are all within a sphere of radius $2R$, which contains

$$(\tfrac{4}{3})\pi V (2R)^3 = 8(N + 3) \qquad (4.3)$$

points. This is the number of different $E$'s needed to construct a metric tensor and three sets of covariant components. In the presence of symmetry (4.3) is divided by $t$. Actually, Friedel's law reduces (4.3) to

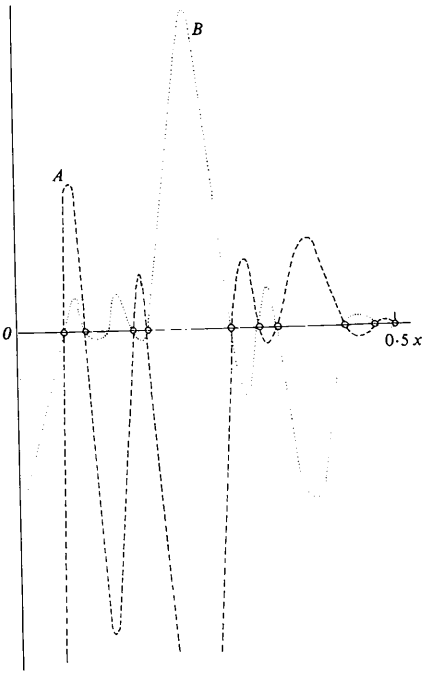


Fig. 2. Centrosymmetric linear structure with $N = 20$ and fractional coordinates $x_j = \pm 0{\cdot}06, \pm 0{\cdot}09, \pm 0{\cdot}15, \pm 0{\cdot}17, \pm 0{\cdot}28, \pm 0{\cdot}32, \pm 0{\cdot}34, \pm 0{\cdot}43, \pm 0{\cdot}47, \pm 0{\cdot}49$. Plot of the difference between the two members of (4.1). The metric tensor is a Goedkoop matrix. Curve $A$: $k = 0$, $h_p = 1, 2, \ldots, 10$. All the zeros correspond to atomic positions. Curve $B$: $k = 1$, $h_p = 2, 3, \ldots, 11$. One zero does not correspond to atomic positions.

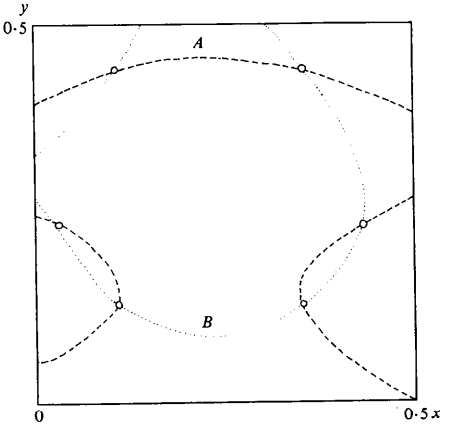


Fig. 3. Centrosymmetric two-dimensional theoretical structure with $N = 12$ and atomic coordinates $(x_j,y_j) = (0{\cdot}43, 0{\cdot}23), (0{\cdot}355, 0{\cdot}439), (0{\cdot}105, 0{\cdot}439), (0{\cdot}03, 0{\cdot}23), (0{\cdot}105, 0{\cdot}135), (0{\cdot}355, 0{\cdot}135)$ and the symmetry-related ones. Plot of the curves represented by (4.1), with base vectors $\mathbf{h}_p = 10, 01, 11, 1\bar{1}, 20, 02$ and metric tensor $\mathbf{G}$, for two different $\mathbf{k}$'s. Curve $A$: $\mathbf{k} = 03$. Curve $B$: $\mathbf{k} = 2\bar{2}$. The intersection of both curves corresponds to atomic positions. There are not spurious intersections in the remainder of the unit cell.

half, but in order to distinguish between centro- and noncentrosymmetric structures a complex $E$ is counted twice.

It would be possible to reduce the number of $E$'s needed for atomic-coordinate determination to $3N$, the number of unknowns, only if it were possible to select a set of $N$ **V**'s which form an orthogonal basis. In this way the metric tensor that arises would be diagonal. For constructing one surface only $N$ $E$'s would be required.

It is interesting to note that von Eller's (1962) forbidden-domains method can be easily derived geometrically. The covariant component of the vector $\mathbf{e}_i$ in the basis $\mathbf{V}(\mathbf{h}_p)$ is $[\mathbf{V}(\mathbf{h}_p)|\mathbf{e}_i] = n_i \exp(-2\pi i \mathbf{h}_p.\mathbf{r}_i)$; then the squared norm of the projection of $\mathbf{e}_i$ on **S** is

$$\|\mathbf{e}_{s,i}\|^2 = \sum_{p=1}^{s}\sum_{q=1}^{s} (\mathsf{H}_s^{-1})_{pq} n_i^2 \exp[2\pi i(\mathbf{h}_p - \mathbf{h}_q).\mathbf{r}_i]. \quad (4.4)$$

On the other hand, from (2.2) $\|\mathbf{e}_i\|^2 = n_i$. Then, the inequality $\|\mathbf{e}_{s,i}\|^2 \leq \|\mathbf{e}_i\|^2$ leads to

$$\sum_{p=1}^{s}\sum_{q=1}^{s} (\mathsf{H}_s^{-1})_{pq} n_i \exp[2\pi i(\mathbf{h}_p - \mathbf{h}_q).\mathbf{r}] \leq 1. \quad (4.5)$$

The left-hand side of (4.5) as a function of **r** defines regions in the unit cell forbidden for atoms of weight $n_i$ or higher. These are von Eller's forbidden domains. From the geometrical approach it is clear that the inequality $\|\mathbf{e}_{s,i}\|^2 \leq \|\mathbf{e}_{s+1,i}\|^2$ could also be considered and that for $s = N$ (4.5) is an equality.

## 5. Summary

This geometrical approach suggests that the computation of atomic coordinates can be addressed independently of Fourier series. The main proposition is that the phase problem can be reduced to the determination of those phases which allow the construction of three surfaces.

Inaccuracies in experimental data may require an increase in the number of surfaces because of instability in matrix inversions. Some successful phase extensions involving high-order matrix inversions have been reported (Podjarny *et al.*, 1976), and it is conceivable that this will not be an insurmountable obstacle. On the other hand it is interesting to note the difference between the theoretical requirements of (4.1) and the Fourier series. A method derived from (4.1) requires a finite number of $E$'s and the Fourier series requires an infinite number. From Fig. 4 it is evident that, with theoretical data, (4.1) is better than the Fourier series. At present we do not know which method would require the lesser amount of *experimental* data. (Fig. 5 is as Fig. 2 but with an incomplete basis.)

The authors thank Dr E. E. Castellano for many helpful discussions and Dr D. Avalos for his critical reading of the manuscript. We also thank Drs H. Hauptman and J. Karle for their letters in support of this paper and their comments and criticisms.

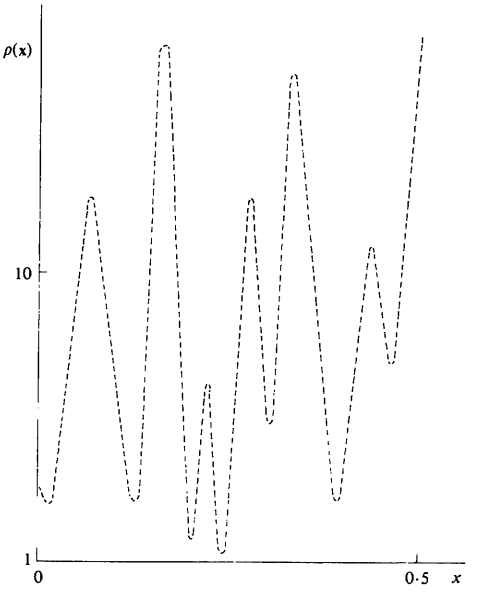


Fig. 4. Plot of $\rho(x)$ corresponding to the structure described in Fig. 2, constructed with a Fourier series partial sum that includes the same coefficients which suffice to construct the curve $A$ in Fig. 2.

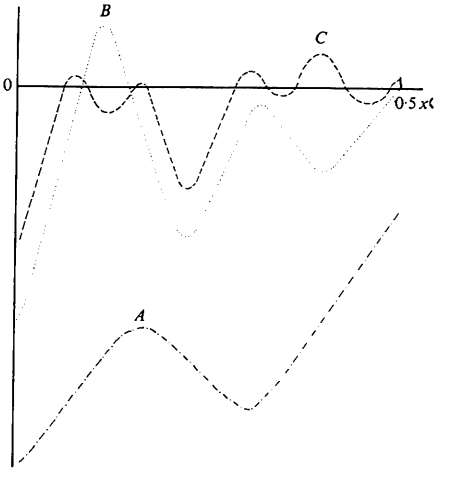


Fig. 5. As in Fig. 2 but with an incomplete basis. Curve $A$: $k = 0$, $h_p = 1, 2, 3$. Curve $B$: $k = 0$, $h_p = 1, 2, \ldots, 6$. Curve $C$: $k = 0$, $h_p = 1, 2, \ldots, 9$. The atomic positions are approximately determined by curve $C$.

## APPENDIX *A*

(*a*) *Calculus of the scalar product defined by* (2.19)

From (2.18) and (2.2) we can write

$$[\mathbf{V}(\mathbf{k})|\mathbf{V}(\mathbf{h})] = \sum_{\alpha=1}^{t}\sum_{\beta=1}^{t}\sum_{i=1}^{N/t}\sum_{j=1}^{N/t} \exp[2\pi i(\mathbf{h}\hat{A}_\alpha \mathbf{r}_i - \mathbf{k}\hat{A}_\beta \mathbf{r}_j)]\,[\mathbf{e}_j|\mathbf{e}_i]$$

$$= \sum_{\alpha=1}^{t}\sum_{\beta=1}^{t}\sum_{i=1}^{N/t} \exp[2\pi i(\mathbf{h}\hat{A}_\alpha - \mathbf{k}\hat{A}_\beta).\mathbf{r}_i]\,n_i. \qquad (A.1)$$

Applying the rearrangement theorem in the sum over $\alpha$ it follows that

$$[\mathbf{V}(\mathbf{k})|\mathbf{V}(\mathbf{h})] = \sum_{\alpha=1}^{t}\sum_{\beta=1}^{t}\sum_{i=1}^{N/t} \exp[2\pi i(\mathbf{h}\hat{A}_\alpha - \mathbf{k}).\hat{A}_\beta \mathbf{r}_i]\,n_i$$

$$= \sum_{\alpha=1}^{t} \exp(2\pi i \mathbf{a}_\alpha.\mathbf{h})E(\mathbf{h}\mathsf{A}_\alpha - \mathbf{k}). \qquad (A.2)$$

(*b*) *Proof of* (2.20)

The normalized structure factor $E(\mathbf{h})$ is

$$E(\mathbf{h}) = \sum_{\alpha=1}^{t}\sum_{j=1}^{N/t} n_j \exp(2\pi i \mathbf{h}\hat{A}_\alpha \mathbf{r}_j). \qquad (A.3)$$

Applying the rearrangement theorem

$$E(\mathbf{h}) = \sum_{\alpha=1}^{t}\sum_{j=1}^{N/t} n_j \exp(2\pi i \mathbf{h}\hat{A}_\beta\hat{A}_\alpha \mathbf{r}_j)$$

$$= \exp(2\pi i \mathbf{a}_\beta.\mathbf{h})\,E(\mathbf{h}\mathsf{A}_\beta). \qquad (A.4)$$

Then, replacing (*A*.4) in (2.19) with $\mathbf{k} = \mathbf{0}$ if follows that

$$[\mathbf{V}(\mathbf{0})|\mathbf{V}(\mathbf{h})] = \sum_{\beta=1}^{t} \exp(2\pi i \mathbf{a}_\beta.\mathbf{h})\,E(\mathbf{h}\mathsf{A}_\beta) = tE(\mathbf{h}). \qquad (A.5)$$

(*c*) *Calculus of* $\|\mathbf{V}(\mathbf{k})\|^2$

Equation (2.23) can be derived from (*A*.1) by putting $\mathbf{h} = \mathbf{k}$; then

$$\|\mathbf{V}(\mathbf{k})\|^2 = \sum_{\alpha=1}^{t}\sum_{\beta=1}^{t}\sum_{j=1}^{N/t} n_j \exp[2\pi i \mathbf{k}(\hat{A}_\alpha - \hat{A}_\beta)\mathbf{r}_j]. \qquad (A.6)$$

Defining

$$g_i(\mathbf{k}) = \sum_{\alpha=1}^{t} \exp(2\pi i \mathbf{k}\hat{A}_\alpha \mathbf{r}_i), \qquad (A.7)$$

it follows that

$$\|\mathbf{V}(\mathbf{k})\|^2 = \sum_{j=1}^{N/t} n_j|g_j|^2, \qquad (A.8)$$

which proves that $\|\mathbf{V}(\mathbf{k})\|^2$ is real and non-negative, as expected.

## APPENDIX *B*

An element of the second partial derivatives matrix is

$$C_{mn} = \frac{\partial^2\|\mathbf{V}_s(\mathbf{0})\|^2}{\partial\varphi_m\partial\varphi_n} = D_{mn}E_mE_n^* + D_{nm}E_nE_m^* - \delta_{mn}\left(\sum_q D_{mq}E_mE_q^* + D_{qm}E_m^*E_q\right). \qquad (B.1)$$

Let $\mathbf{X}$ be an arbitrary $s$-dimensional vector such that $\mathbf{X} \neq \mathbf{0}$. if the quadratic form $\mathbf{X}\mathrm{C}\mathbf{X}^\dagger$ is non-negative the extremum is a minimum:

$$\mathbf{X}\mathrm{C}\mathbf{X}^\dagger = \sum_{m,n} D_{mn}E_mE_n^*X_mX_n^* + \sum_{m,n} D_{mn}E_nE_m^*X_mX_n^* - \sum_{q,m} D_{mq}E_mE_q^*X_m^2 - \sum_{q,m} D_{mq}E_m^*E_qX_m^2. \qquad (B.2)$$

Changing the index names and multiplying by $(-1)$, it follows that

$$-\mathbf{X}\mathrm{C}\mathbf{X}^\dagger = \sum_{m,n} D_{mn}E_mE_n^*(X_m^2 + X_n^2 - X_mX_n^* - X_nX_m^*) = \sum_{m,n} D_{mn}E_mE_n^*|X_m - X_n|^2 \leq \operatorname*{Max}_{u,v}|X_u - X_v|^2\left(\sum_{\substack{m,n\\ m\neq n}} D_{mn}E_mE_n^*\right). \qquad (B.3)$$

But, with solution (3.6):

$$\sum_{\substack{n\\ n\neq m}} D_{mn}E_mE_n^* = -|E_m|\left|\sum_{\substack{n\\ n\neq m}} D_{nm}E_n\right|. \qquad (B.4)$$

Therefore,

$$\mathbf{X}\mathrm{C}\mathbf{X}^\dagger \geq \operatorname*{Max}_{u,v}|X_u - X_v|^2 \sum_m |E_m|\left|\sum_{\substack{n\\ n\neq m}} D_{nm}E_n\right| \geq 0. \qquad (B.5)$$

This inequality proves that C is non-negative.

## APPENDIX C

Calculate the sum

$$\sum_{\substack{p=1\\p\neq m}}^{s} (H_s^{-1})_{pm} E(\mathbf{h}_p). \qquad (C.1)$$

It is clear that

$$(H_s^{-1})_{pm} = (-1)^{p+m} \frac{\text{Minor}(H_s)_{mp}}{\det(H_s)}, \qquad (C.2)$$

where the Minor is the determinant of the matrix obtained from $H_s$ suppressing the $m$th row and the $p$th column. If this matrix is $H'_s$, then $\text{Minor}(H_s)_{mp} = \det(H'_s)$. Assume that the $m$th row and the $m$th column are the last ones. If they are not, it is always possible to rearrange the matrix without changing its meaning. On the other hand, it is clear that this rearrangement does not change the determinant sign because rows and columns are interchanged twice.

$\text{Det}(H'_s)$ can be calculated by developing it along its last column; that is, the $m$th. Therefore,

$$\det(H'_s) = \sum_{q=1}^{s-1} H_{qm}(-1)(-1)^{q+m} \text{Minor}(H'_s)_{qm}. \qquad (C.3)$$

The factor $(-1)$ has been introduced to take into account the suppression of the $p$th column. Again, $\text{Minor}(H'_s)_{qm}$ is the determinant of the matrix obtained from $H'_s$ suppressing the $q$th row and the $m$th column. Then, $\text{Minor}(H'_s)_{qm}$ is the determinant of the matrix obtained from $H_s$ suppressing the $m$th and the $q$th rows, and the $m$th and the $p$th columns. Therefore,

$$\text{Minor}(H'_s)_{qm} = \text{Minor}(H_{s-1})_{qp}, \qquad (C.4)$$

where $H_{s-1}$ is the matrix obtained from $H_s$ suppressing the $m$th row and the $m$th column. From (C.4), (C.3) and (C.2) it follows that

$$(H_s^{-1})_{pm} = \sum_{q=1}^{s-1} H_{qm}(-1)^{p+m+q+m+1} \frac{\text{Minor}(H_{s-1})_{qp}}{\det(H_s)}$$
$$= -\sum_{q=1}^{s-1} (H_{s-1}^{-1})_{pq} E(\mathbf{h}_m - \mathbf{h}_q) \frac{\det(H_{s-1})}{\det(H_s)}. \qquad (C.5)$$

Therefore,

$$\sum_{\substack{p=1\\p\neq m}}^{s} (H_s^{-1})_{pm} E(\mathbf{h}_p)$$
$$= -\sum_{p=1}^{s-1}\sum_{q=1}^{s-1} (H_{s-1}^{-1})_{pq} E(\mathbf{h}_m - \mathbf{h}_q) E(\mathbf{h}_p) \frac{\det(H_{s-1})}{\det(H_s)} \qquad (C.6)$$

where

$$\frac{\det(H_{s-1})}{\det(H_s)} = (H_s^{-1})_{mm} = D_{mm}. \qquad (C.7)$$

Since $\det(H_s)$ can be calculated developing it along the $m$th row and then along the $m$th column, by reasoning analogous to that just given, it is clear that

$$\det(H_s) = \left[\sum_{p=1}^{s-1} \text{Minor}(H_s)_{mp} (-1)^{m+p} \times E(\mathbf{h}_p - \mathbf{h}_m)\right] + \det(H_{s-1}) H_{mm}$$
$$= \left[-\sum_{p=1}^{s-1}\sum_{q=1}^{s-1} \text{Minor}(H_{s-1})_{qp} E(\mathbf{h}_p - \mathbf{h}_m) \times E(\mathbf{h}_m - \mathbf{h}_q)(-1)^{p+q}\right] + \det(H_{s-1}) H_{mm}. \qquad (C.8)$$

Then,

$$\frac{1}{D_{mm}} = \frac{1}{(H_s^{-1})_{mm}} = \frac{\det(H_s)}{\det(H_{s-1})}$$
$$= H_{mm} - \sum_{p=1}^{s-1}\sum_{q=1}^{s-1} (H_{s-1}^{-1})_{pq} E(\mathbf{h}_p - \mathbf{h}_m) E(\mathbf{h}_m - \mathbf{h}_q)$$
$$= \|\mathbf{V}(\mathbf{h}_m)\|^2 - \|\mathbf{V}_{s-1}(\mathbf{h}_m)\|^2$$
$$= \|\mathbf{V}(\mathbf{h}_m) - \mathbf{V}_{s-1}(\mathbf{h}_m)\|^2. \qquad (C.9)$$

The relationships obtained in this Appendix are also valid if symmetry is considered. For example, the quotient $\det(G_s)/\det(G_{s-1})$ may be interpreted as the squared norm of the orthogonal complement of $\mathbf{V}_{s-1}(\mathbf{h}_m)$.

### References


CASTELLANO, E. E., PODJARNY, A. D. & NAVAZA, J. (1973). *Acta Cryst.* A**29**, 609–615.

ELLER, G. VON (1955). *Acta Cryst.* **8**, 641–645.

ELLER, G. VON (1962). *Acta Cryst.* **15**, 590–595.

GOEDKOOP, J. A. (1950). *Acta Cryst.* **3**, 374–378.

HARKER, D. & KASPER, J. S. (1948). *Acta Cryst.* **1**, 70–75.

HAUPTMAN, H. & KARLE, J. (1953). *Solution of the Phase Problem. I. The Centrosymmetric Crystal.* ACA Monograph N. 3. Pittsburgh: Polycrystal Book Service.

HOPPE, W. & GASSMANN, J. (1968). *Acta Cryst. B***24**, 97–107.

KARLE, J. (1971). *Acta Cryst.* B**27**, 2063–2065.

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

KARLE, J. & HAUPTMAN, H. (1956). *Acta Cryst.* **9**, 635–651.

KITAIGORODSKY, A. I. (1962). *The Theory of Crystal Structure Analysis.* New York: Consultants Bureau.

PODJARNY, A. D. & YONATH, A. (1977). *Acta Cryst.* A**33**, 655–661.

PODJARNY, A. D. & YONATH, A. & TRAUB, W. (1976). *Acta Cryst.* A**32**, 281–292.

RANGO, C. DE, TSOUCARIS, G. & ZELWER, CH. (1974). *Acta Cryst.* A**30**, 342–353.

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

TSOUCARIS, G. (1970). *Acta Cryst.* A**26**, 492–499.