HAL
open science

Sparsity of the normal matrix in the refinement of macromolecules at atomic and subatomic resolution

Christian Jelsch

To cite this version:

Christian Jelsch. Sparsity of the normal matrix in the refinement of macromolecules at atomic and subatomic resolution. Acta Crystallographica Section A : Foundations and Advances [2014-..], 2001, A57, pp.558-570. 10.1107/s0108767301007656 . hal-01713044

HAL Id: hal-01713044

https://hal.science/hal-01713044v1

Submitted on 20 Feb 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

electronic reprint

Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Sparsity of the normal matrix in the refinement of macromolecules at atomic and subatomic resolution

Christian Jelsch

Copyright © International Union of Crystallography

Author(s) of this paper may load this reprint on their own web site provided that this cover page is retained. Republication of this article or its storage in electronic databases or the like is not permitted without prior permission in writing from the IUCr.

Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 19 February 2001
Accepted 9 May 2001

Sparsity of the normal matrix in the refinement of macromolecules at atomic and subatomic resolution

Christian Jelsch

Laboratoire de Cristallographie et de Modélisation des Matériaux Minéraux et Biologiques (LCM^{3}B), CNRS, UHP, Faculté des Sciences, BP 239, 54506 Vandoeuvre-les-Nancy, France. Correspondence e-mail: jelsch@lcm3b.uhp-nancy.fr

The normal matrix in the least-squares refinement of macromolecules is very sparse when the resolution reaches atomic and subatomic levels. The elements of the normal matrix, related to coordinates, thermal motion and charge-density parameters, have a global tendency to decrease rapidly with the interatomic distance between the atoms concerned. For instance, in the case of the protein crambin at 0.54 Å resolution, the elements are reduced by two orders of magnitude for distances above 1.5 Å. The neglect *a priori* of most of the normal-matrix elements according to a distance criterion represents an approximation in the refinement of macromolecules, which is particularly valid at very high resolution. The analytical expressions of the normal-matrix elements, which have been derived for the coordinates and the thermal parameters, show that the degree of matrix sparsity increases with the diffraction resolution and the size of the asymmetric unit.

© 2001 International Union of Crystallography
Printed in Great Britain – all rights reserved

1. Introduction

The least-squares method is widely used for the refinement of crystal structures. The optimization procedure, which uses the matrix of normal equations, has a great power of convergence. In addition, the inversion of the complete normal matrix provides the variance/covariance matrix of the refined parameters and permits the determination of the precision of the refined structure (Hamilton, 1964). The conventional least-squares method using the full normal matrix has not, however, been utilized to a great extent in macromolecular refinement, because the calculation and the inversion of the whole matrix is a computationally expensive step, demanding significant time and memory for large systems. The full-matrix method, as implemented in *SHELXL* (Sheldrick & Schneider, 1997), has been applied to a limited number of small to medium-size protein structures at atomic resolution in order to determine the accuracy of the final refined structure. However, because of the aforementioned prohibitive computation times, the refinement packages commonly used for macromolecules apply minimization methods that do not require the complete exact normal matrix (Tronrud, 1999).

For the rapid computation of the full normal matrix, several algorithms have been proposed. In the procedure described by Agarwal (1978), approximate values of the normal-matrix elements are computed by fast Fourier transform (FFT). A fast approximate method for the evaluation of the matrix elements for large structures with many diffraction data has been proposed by Templeton (1999). This efficient calculation uses precalculated tables and is valid for atoms with isotropic thermal displacement factors. For structures with anisotropic thermal displacement parameters, more complicated tables would be required. For the fast computation of the exact normal matrix, Tronrud (1999) proposed an efficient algorithm based on FFT methods.

Dodson (1981) and Templeton (1999) have observed that the normal-matrix elements, related to parameter pairs, have a global tendency to decrease with the distance between the two atoms. Recently, several protein structures have been refined at atomic and subatomic resolution in our laboratory (Housset *et al.*, 2000; Jelsch *et al.*, 2000). The magnitudes of the normal-matrix elements were extensively analysed in the case of crambin at 0.54 Å resolution; the matrix was found to be extremely sparse. To understand which factors contribute to the sparsity of the matrix, the analytical expression of the normal-matrix elements has been derived for several parameters, as described in this paper.

2. The matrix of normal equations

2.1. Least-squares minimization

In a least-squares refinement performed against the reflection intensities, the quantity to minimize is of the form

$$\sum_{\mathbf{H}} W_{\mathbf{H}} (I_{\mathbf{H}} - I_{\mathbf{H}}^{\rm obs})^2, \qquad (1)$$

where $I_{\mathbf{H}}$ and $I_{\mathbf{H}}^{\rm obs}$ are the calculated and observed diffraction intensities, respectively. The quantity to be minimized can also

electronic reprint

be expressed similarly as a function of the structure-factor moduli $|F_{\mathbf{H}}|$ instead of the intensities $I_{\mathbf{H}}$. The factor $W_{\mathbf{H}}$ represents a weight for each reflection **H**, which can be taken as the squared inverse estimated uncertainty of the measured intensity. If the intensities are assumed to be a linear function of each parameter for small changes, the system of normal equations to be solved during one refinement cycle is then of the form

$$\mathbf{A}\,\mathrm{d}\mathbf{X} = \mathbf{b}, \tag{2}$$

where **A** is the n^2 symmetric normal matrix and d**X** is the unknown shift vector to be applied to the n refined parameters. **b** is a vector of dimension n with elements like

$$b_i = \sum_{\mathbf{H}} W_{\mathbf{H}}\,(\partial I_{\mathbf{H}}/\partial p_i)\,(I_{\mathbf{H}}^{\rm obs} - I_{\mathbf{H}}). \tag{3}$$

The normal-matrix elements concerning the refined parameters p_i and p_j are obtained from the summation of the products of the derivatives over the reflections **H**:

$$A_{ii} = \sum_{\mathbf{H}} W_{\mathbf{H}}\,(\partial I_{\mathbf{H}}/\partial p_i)\,(\partial I_{\mathbf{H}}/\partial p_j). \tag{4}$$

The solution of equation (2) can be obtained by inverting the normal matrix. Alternatively, the conjugate-gradient method (Hestenes & Stiefel, 1952; Konnert, 1976) solves the equation by an iterative algorithm which does not require the matrix inversion. There are other least-squares refinement methods that do not assume the linearity of the intensities (or structure factors) with respect to the parameters and that use the Hessian matrix of second derivatives (Lunin & Urzhumtsev, 1985; Murshudov *et al.*, 1999).

2.2. The normalized matrix

The diagonal terms A_{ii} in the normal matrix provide an approximate value of the standard deviation of the variables p_i in the least-squares refinement: $\sigma(p_i) \simeq A_{ii}^{-1/2}$ (Hamilton, 1964; Franklin & Marsh, 1982). The parameters p_i can thus be normalized by dividing them by the quantity $A_{ii}^{1/2}$. The corresponding normalized matrix is obtained by replacing the A_{ij} elements by

$$A'_{ij} = A_{ij}/(A_{ii}A_{jj})^{1/2}. \tag{5}$$

The normalized A'_{ij} value corresponds also to the correlation coefficient between the parameters p_i and p_j in a binary refinement (Hamilton, 1964). In the following parts of the present study, the criterion that defines the sparsity of the normal matrix is the smallness of the elements A'_{ij} relative to unity.

2.3. The software *MOPRO*

Protein diffraction at atomic resolution is becoming more and more accessible and accurate with the availability of intense third-generation synchrotron sources (Dauter *et al.*, 1997). Depositions of crystal structures of small and even medium-size proteins at resolutions better than 1.2 Å are becoming more frequent at the Protein Data Bank (Longhi *et al.*, 1998). Therefore, charge-density studies, which require subatomic resolution (d < 0.6 Å), can be performed on molecules of increasing size (Jelsch *et al.*, 1998; Housset *et al.*, 2000). Recently, the valence electron density of the protein crambin was analysed at 0.54 Å resolution (Jelsch *et al.*, 2000). Following from this work, the computer program *MOPRO* was developed for the charge-density analysis of macromolecules (Guillot *et al.*, 2001). This software is derived from the least-squares program *MOLLY* (Hansen & Coppens, 1978), which was extensively modified with the introduction of restraints on the stereochemistry and on the thermal motion.

With *MOLLY*, the least-squares refinement is performed by inverting the normal matrix to solve equation (2). In the case of macromolecules, as the number of variables is large, the storage and inversion of the full matrix is computationally expensive. With the conjugate-gradient algorithm (Hestenes & Stiefel, 1952; Konnert, 1976), it is not necessary to store the full normal matrix (Rae, 1978) as its inversion is not required. Thus, procedures in which the small elements are neglected can be applied efficiently. For instance, the *SHELXL* refinement software (Sheldrick & Schneider, 1997) computes the sparse normal matrix 'on the fly': all the elements of the normal matrix are computed but only the elements with a significant magnitude are stored. The conjugate-gradient method, which has been implemented in *MOPRO* for macromolecular refinement (Guillot *et al.*, 2001), has the additional advantage of being less sensitive to matrix singularity problems compared with the matrix-inversion method. The inversion of the full matrix in the case of crambin (Jelsch *et al.*, 2000) was often hindered by the singularity of the normal matrix, especially when alternative positions of disordered protein atoms or solvent molecules were refined simultaneously. These matrix-inversion problems were avoided with the conjugate-gradient procedure.

In the application to proteins at subatomic resolution, the calculations of the diffraction intensities and of their derivatives are the most consuming parts of the program. These routines have been parallelized in *MOPRO* with respect to the reflections (Guillot *et al.*, 2001). All the crystallographic refinements described in this paper were performed with *MOPRO* and the sparsity of the normal matrix was assessed before adding the stereochemical restraint terms to the matrix. As the restraints concern atoms which are generally close to each other in the molecular structure, their application modifies only a small proportion of elements in the normal matrix.

3. Analytical expression of matrix elements

The expressions of the normal-matrix elements are analysed in this section for different variables used in a crystal structure refinement. The dependencies of the matrix elements on the Patterson vectors observed by Templeton (1999) for the coordinates and isotropic temperature factors are generalized to anisotropic thermal displacement as well as charge-density parameters.

electronic reprint

3.1. Structure-factor amplitude

The structure-factor amplitude $F_{\mathbf{H}}$ can be computed as the sum over the atoms a in the unit cell:

$$F_{\mathbf{H}} = \sum_a G_a \exp(2i\pi\, \mathbf{H}\cdot\mathbf{X}_a). \qquad (6)$$

G_a is a generalized scattering factor of the atom; it incorporates the scattering factor f_a of the atom at rest, the thermal motion and the deformations of the atomic electron density. In the case of spherical atoms with isotropic temperature factors B_a, the factor G_a is simply a function of the resolution $d = 1/2s$:

$$G_a = f_a(s)\exp(-B_a s^2). \qquad (7)$$

When anisotropic thermal displacement parameters are assigned to the atoms, the expression of G_a as a function of the thermal tensor is

$$G_a = f_a(s)\exp(-{}^t\mathbf{H}\,\beta\,\mathbf{H}). \qquad (8)$$

For simplicity of the equations, the β^{ij} coefficients are considered here instead of the more traditional U^{ij} thermal displacement parameters. The occupancy Q of an atom can also be incorporated into the G_a factor, directly as a proportionality factor. If the anomalous scattering is neglected, the generalized scattering factor G generated by a spherical-atom model is a positive real quantity.

3.2. The multipolar pseudo-atom model

At subatomic resolutions ($d \simeq 0.5$ Å), the deformation of the atomic electronic cloud caused by chemical bonds can be derived from the X-ray diffraction experiment (Coppens, 1967). In the programs *MOLLY* (Hansen & Coppens, 1978) and *MOPRO* (Guillot *et al.*, 2001), the atomic electron density is described as

$$\rho_{\rm atom}(\mathbf{r}) = \rho_{\rm core}(r) + P_{\rm val}\kappa^3\rho_{\rm val}(\kappa r) + \sum_l \kappa'^3 R_l(\kappa' r)\sum_{m\pm} P_{lm} y_{lm\pm}. \qquad (9)$$

The first two terms on the right-hand side describe the spherically symmetric core-plus-valence density. The third term describes the non-spherical multipolar distribution of the valence electron density of the atoms. The R_l are exponential Slater-type radial functions and the $y_{lm\pm}$ are real spherical harmonic angular functions. The parameters κ and κ' model the expansion/contraction of the spherical and multipolar parts, respectively, of the valence electron density.

The charge-density parameters relative to the spherical part, such as the population P_v and the expansion/contraction coefficient κ of the valence density, can be readily included in the generalized scattering factor G. In the description of aspherical atoms, the multipole populations $P_{lm\pm}$ can also be incorporated into the G factor. The multipoles are expressed in a local axis system in the programs *MOLLY* and *MOPRO*, and the atomic scattering factor G_a is dependent on the neighbouring atoms defining the axes. The component of the scattering factor arising from the centrosymmetric multipole is a real number which can be negative or positive depending on the relative orientation of the vector $\mathbf{H}$ and the multipole. Antisymmetric multipoles generate purely imaginary scattering factors. For the non-hydrogen atoms, the multipoles have a contribution to the atomic scattering factor that is, however, small compared with the spherical part of the atomic density in all the resolution ranges (Jelsch *et al.*, 1998); the generalized scattering factor G_a is thus not far from being a real positive number in the complex plane.

3.3. Expression of A_{xx} elements

In this section, let $\mathbf{A}(x_i, x_j)$ be a matrix element corresponding to the x coordinates of atoms i and j. The structure is refined against the diffraction intensities. The triclinic space group $P1$ is considered as there is only one symmetry operator (see Appendix A for a generalization to other space groups). The derivative of the calculated structure factor $F_{\mathbf{H}}$ with respect to the coordinate x_i is

$$\partial F_{\mathbf{H}}/\partial x_i = 2i\pi\, h\, G_i \exp(2i\pi\, \mathbf{H}\cdot\mathbf{X}_i). \qquad (10)$$

The intensity is equal to the square of the structure-factor moduli. The derivative of the intensity is thus equal to a real positive quantity:

$$\partial I_{\mathbf{H}}/\partial x_i = (\partial F_{\mathbf{H}}/\partial x_i)\, F^*_{\mathbf{H}} + (\partial F^*_{\mathbf{H}}/\partial x_i)\, F_{\mathbf{H}}. \qquad (11)$$

The element $\mathbf{A}(x_i, x_j)$ of the normal matrix related to the two coordinates x_i and x_j is the sum over all reflections [equation (4)] of the product of the derivatives:

$$\begin{aligned}\frac{\partial I_{\mathbf{H}}}{\partial x_i}\frac{\partial I_{\mathbf{H}}}{\partial x_j} &= U_{\mathbf{H}} + V_{\mathbf{H}}\\ &= \left(\frac{\partial F_{\mathbf{H}}}{\partial x_i}\frac{\partial F^*_{\mathbf{H}}}{\partial x_j}F_{\mathbf{H}}F^*_{\mathbf{H}} + \frac{\partial F^*_{\mathbf{H}}}{\partial x_i}\frac{\partial F_{\mathbf{H}}}{\partial x_j}F_{\mathbf{H}}F^*_{\mathbf{H}}\right)\\ &\quad + \left(\frac{\partial F_{\mathbf{H}}}{\partial x_i}\frac{\partial F_{\mathbf{H}}}{\partial x_j}F^{*2}_{\mathbf{H}} + \frac{\partial F^*_{\mathbf{H}}}{\partial x_i}\frac{\partial F^*_{\mathbf{H}}}{\partial x_j}F^2_{\mathbf{H}}\right). \qquad (12)\end{aligned}$$

The terms $U_{\mathbf{H}}$ and $V_{\mathbf{H}}$ are each sums of two conjugated complex numbers and are therefore real:

$$U_{\mathbf{H}} = 8\pi^2 h^2 I_{\mathbf{H}}\, {\rm Re}(G_i G_j^* E) \qquad (13)$$

and

$$V_{\mathbf{H}} = -8\pi^2 h^2\, {\rm Re}(F^{*2}_{\mathbf{H}}\, G_i G_j E'), \qquad (14)$$

with

$$E = \exp[2\pi\, \mathbf{H}\cdot(\mathbf{X}_i - \mathbf{X}_j)] \quad \text{and} \quad E' = \exp[2i\pi\, \mathbf{H}\cdot(\mathbf{X}_i + \mathbf{X}_j)]. \qquad (15)$$

The summation over the reflections of these terms corresponds to the expressions $H1$ and $H2$ described by Agarwal (1978):

$$H1(x_i, x_j) = \sum_{\mathbf{H}} W_{\mathbf{H}} U_{\mathbf{H}} \quad \text{and} \quad H2(x_i, x_j) = \sum_{\mathbf{H}} W_{\mathbf{H}} V_{\mathbf{H}}; \qquad (16)$$

$$\mathbf{A}(x_i, x_j) = H1(x_i, x_j) + H2(x_i, x_j). \qquad (17)$$

Agarwal (1978) proposed a fast approximate computation of the normal-matrix elements based on an FFT algorithm in which the second terms $H2(x_i, x_j)$ are neglected. For the

electronic reprint

diagonal elements, as the vector $(\mathbf{X}_i - \mathbf{X}_j)$ is nullified, the analytical expression of $\mathbf{A}(x_i, x_i)$ simplifies to

$$A(x_i, x_i) = \sum_{\mathbf{H}} W_{\mathbf{H}}(U_{\mathbf{H}} + V_{\mathbf{H}}) = 8\pi^2 \sum_{\mathbf{H}} W_{\mathbf{H}} h^2 \{|G_i|^2 I_{\mathbf{H}} - \mathrm{Re}[F_{\mathbf{H}}^{*2} G_i^2 \exp(4i\pi\, \mathbf{H}\cdot\mathbf{X}_i)]\}. \quad (18)$$

The summation over the reflections of the first terms $U_{\mathbf{H}}$ is additive in absolute value since the factor $|G_i|^2 I_{\mathbf{H}}$ is always a real and positive number. The second terms $V_{\mathbf{H}}$ are real numbers of random sign that are always smaller than $U_{\mathbf{H}}$ in absolute value. An approximate value of the diagonal elements can be obtained from the first-term summation (Agarwal, 1978):

$$A(x_i, x_i) \simeq 8\pi^2 \sum_{\mathbf{H}} W_{\mathbf{H}}\, |G_i|^2\, h^2\, I_{\mathbf{H}}. \quad (19)$$

For the off-diagonal elements of the normal matrix, the sign of the $U_{\mathbf{H}}$ terms depends on the factor $\mathrm{Re}(G_i G_j^* E)$ in (13). It must be remembered here that, for a spherical atom, the generalized scattering factor G_i is a positive real number, while, for a multipolar atom, G_i is complex but very close to a positive real number. The exponential factor E in (13) is close to unity when the exponent is very small. For $2\pi|\mathbf{H}\cdot(\mathbf{X}_i - \mathbf{X}_j)| < \pi/2$, the factor E is a complex number with a positive real part and the $U_{\mathbf{H}}$ terms are then systematically positive. This is the case for a large number of reflections when the two atoms are close to each other in the molecule ($\mathbf{X}_i \simeq \mathbf{X}_j$). However, as the resolution increases, the indices $(h, k, l) = \mathbf{H}$ can take larger values and the proportion of systematically positive $U_{\mathbf{H}}$ terms diminishes.

As the exponent function has a period of $2i\pi$, the expression of E [equation (15)] is invariant when unit-cell translations are applied to atoms a_i and a_j. More generally, a significant number of $U_{\mathbf{H}}$ terms are positive when one of the Patterson vectors between the two atoms in the crystal is small. The summation $H1 = \sum_{\mathbf{H}} W_{\mathbf{H}} U_{\mathbf{H}}$ is additive in absolute value, totally for the diagonal elements and partially for pairs of atoms close to each other in the crystal. The off-diagonal terms $\mathbf{A}'(x_i, x_j)$ of the normalized matrix can therefore be expected to be small relative to unity when the distance between the atoms in the crystal is large.

As observed by Agarwal (1978), the summation $\sum_{\mathbf{H}} W_{\mathbf{H}} U_{\mathbf{H}}$ represents the Fourier transform of the quantity $8\pi^2 h^2 I_{\mathbf{H}} G_i G_j^*$ evaluated at $(\mathbf{X}_i - \mathbf{X}_j)$, the vector distance between the atoms. This Fourier transform has a very large positive peak at the origin, corresponding to the diagonal terms $\mathbf{X}_i = \mathbf{X}_j$, and then drops rapidly and alternates in sign as the Patterson vector length increases. On the other hand, as the squared conjugated structure factor $F_{\mathbf{H}}^{*2}$ has a random phase in the complex plane (except for centric reflections), the terms $V_{\mathbf{H}}$ are of arbitrary sign.

In the case of structures refined *versus* the structure-factor moduli $|F_{\mathbf{H}}|$ (instead of the intensities), Templeton (1999) showed, in a similar manner, that the elements of the normal matrix depend on the Patterson vector and have a global tendency to decrease with its length.

3.4. Expression of A_{xy} elements

In the case of cross-coordinate pairs $\mathbf{A}(x_i, y_j)$, the quantities $U_{\mathbf{H}}$ and $V_{\mathbf{H}}$ [equations (13) and (14)] are

$$U_{\mathbf{H}} = 8\pi^2\, hk\, I_{\mathbf{H}}\, \mathrm{Re}(G_i G_j^*\, E) \quad (20)$$

and

$$V_{\mathbf{H}} = -8\pi^2\, hk\, \mathrm{Re}(F_{\mathbf{H}}^{*2}\, G_i G_j^*\, E'). \quad (21)$$

The $\mathbf{A}(x_i, x_j)$ elements and the cross-coordinate elements $\mathbf{A}(x_i, y_j)$ differ essentially in that they require respectively the positive factor h^2 [equations (13) and (14)] and the factor hk, which can be positive or negative. At first view, the summation of the terms $U_{\mathbf{H}}$ for small Patterson vectors does not appear this time to be additive in absolute value over an independent part of the reciprocal space (for example, the hemisphere $h \geq 0$ in the triclinic case), because of the non-constant sign of the product hk. The $\mathbf{A}(x_i, y_j)$ elements can thus be expected to be generally smaller than $\mathbf{A}(x_i, x_j)$ and $\mathbf{A}(y_i, y_j)$ in the case of small Patterson vectors.

In this summation, in order to identify partial contributions that might be additive in absolute value, approximate values of $W_{\mathbf{H}} U_{\mathbf{H}}$ shall be considered. The quantity $hk\, \mathrm{Re}(G_i G_j^*\, E)$ is highly anisotropic in three-dimensional space and the summation over $W_{\mathbf{H}} U_{\mathbf{H}}$ merits detailed analysis. The summation can be performed over a positive fraction of the reciprocal space ($h \geq 0, k \geq 0, l \geq 0$) by considering the average value of $W_{\mathbf{H}} U_{\mathbf{H}}$ over the reflections $(+h, \pm k, \pm l)$:

$$\sum_{\mathbf{H}} W_{\mathbf{H}} U_{\mathbf{H}} = \sum_{\mathbf{H}\geq 0} n_{\mathbf{H}} \langle W_{\mathbf{H}} U_{\mathbf{H}} \rangle. \quad (22)$$

The factor $n_{\mathbf{H}}$ accounts for the $\sum_{\mathbf{H}}$ number of regrouped reflections. The reflections with zero indices ($n_{\mathbf{H}} < 4$) deserve special treatment at this point. In this summation, in order to identify partial contributions that might be additive in absolute value, an approximate value of the average $\langle W_{\mathbf{H}} U_{\mathbf{H}} \rangle$ can be obtained by considering the average value of $W_{\mathbf{H}} I_{\mathbf{H}}$ over the $n_{\mathbf{H}}$ reflections [equation (20)]. The deviation from this approximate value is by definition non-additive in absolute value. In a further approximation, the generalized scattering factors G_i can be considered as real positive; the average over the four reflections is then simplified:

$$\langle W_{\mathbf{H}} U_{\mathbf{H}} \rangle \simeq 2\pi^2\, \langle W_{\mathbf{H}} I_{\mathbf{H}} \rangle\, G_i G_j \times \sum_{(+h,\pm k,\pm l)} hk \cos[2\pi\, \mathbf{H}\cdot(\mathbf{X}_i - \mathbf{X}_j)]/n_{\mathbf{H}}. \quad (23)$$

The cosine functions can be summed over the reflections $(+h, \pm k, \pm l)$; an approximate value of $\langle W_{\mathbf{H}} U_{\mathbf{H}} \rangle$ is then given by

$$\langle W_{\mathbf{H}} U_{\mathbf{H}} \rangle \simeq \langle W_{\mathbf{H}} I_{\mathbf{H}} \rangle\, G_i G_j\, U_{4\mathbf{H}}^{xy}, \quad (24)$$

with

electronic reprint

Table 1
Parameter pairs and expressions for the product of the structure-factor derivatives.

The value of the predominant terms $U_{\mathbf{H}}$ (divided by the common factor $I_{\mathbf{H}}G_iG_j$) in space group $P1$ is also given. When the $U_{\mathbf{H}}$ terms are non-additive in absolute value for short Patterson vectors, the symbol $\simeq$ precedes an approximate value of the average of $U_{\mathbf{H}}/I_{\mathbf{H}}G_iG_j$ over the four reflections $(+h, \pm k, \pm l)$. By convention, the indices h, k and l are all positive (≥ 0) here. For the expression of $U_{\mathbf{H}}$ in this table, the generalized structure factors G_i have been approximated to be real quantities. $E = \exp(2i\pi\,\mathbf{H}\cdot\mathbf{X}_i)\exp(-2i\pi\,\mathbf{H}\cdot\mathbf{X}_j) = \exp[2i\pi\,\mathbf{H}\cdot(\mathbf{X}_i-\mathbf{X}_j)]$.

Parameter pair $p_i\,p_j$	Derivatives product $(\partial F_{\mathbf{H}}/\partial p_i)(\partial F^*_{\mathbf{H}}/\partial p_j)$	Value of $U_{\mathbf{H}}/I_{\mathbf{H}}G_iG_j$
$X_i\,X_j$	$4\pi\,h^2\,G_iG_j^*\,E$	$8\pi^2\,h^2\,\cos[2\pi\,\mathbf{H}\cdot(\mathbf{X}_i-\mathbf{X}_j)]$
$X_i\,Y_j$	$4\pi\,hk\,G_iG_j^*\,E$	$\simeq -8\pi^2\,hk\,\sin 2\pi h(x_i-x_j)\,\sin 2\pi k(y_i-y_j)\,\cos 2\pi l(z_i-z_j)$
$B_i^{\rm iso}\,B_j^{\rm iso}$	$s^4\,G_iG_j^*\,E$	$2s^4\,\cos[2\pi\,\mathbf{H}\cdot(\mathbf{X}_i-\mathbf{X}_j)]$
$B_i^{\rm iso}\,\beta_j^{11}$	$s^2\,h^2\,G_iG_j^*\,E$	$2s^2\,h^2\,\cos[2\pi\,\mathbf{H}\cdot(\mathbf{X}_i-\mathbf{X}_j)]$
$B_i^{\rm iso}\,\beta_j^{12}$	$s^2\,hk\,G_iG_j^*\,E$	$\simeq -2s^2\,hk\,\sin 2\pi h(x_i-x_j)\,\sin 2\pi k(y_i-y_j)\,\cos 2\pi l(z_i-z_j)$
$\beta_i^{11}\,\beta_j^{11}$	$h^4\,G_iG_j^*\,E$	$2h^4\,\cos[2\pi\,\mathbf{H}\cdot(\mathbf{X}_i-\mathbf{X}_j)]$
$\beta_i^{11}\,\beta_j^{22}$	$h^2k^2\,G_iG_j^*\,E$	$2h^2k^2\,\cos[2\pi\,\mathbf{H}\cdot(\mathbf{X}_i-\mathbf{X}_j)]$
$\beta_i^{12}\,\beta_j^{12}$	$h^2k^2\,G_iG_j^*\,E$	$2h^2k^2\,\cos[2\pi\,\mathbf{H}\cdot(\mathbf{X}_i-\mathbf{X}_j)]$
$\beta_i^{11}\,\beta_j^{12}$	$h^3k\,G_iG_j^*\,E$	$\simeq -2h^3k\,\sin 2\pi h(x_i-x_j)\,\sin 2\pi k(y_i-y_j)\,\cos 2\pi l(z_i-z_j)$
$\beta_i^{11}\,\beta_j^{23}$	$h^2kl\,G_iG_j^*\,E$	$\simeq -2h^2kl\,\cos 2\pi h(x_i-x_j)\,\sin 2\pi k(y_i-y_j)\,\sin 2\pi l(z_i-z_j)$
$\beta_i^{12}\,\beta_j^{13}$	$h^2kl\,G_iG_j^*\,E$	$\simeq -2h^2kl\,\cos 2\pi h(x_i-x_j)\,\sin 2\pi k(y_i-y_j)\,\sin 2\pi l(z_i-z_j)$
$X_i\,B_j^{\rm iso}$	$2i\pi h\,s^2\,G_iG_j^*\,E$	$\simeq 4\pi\,h\,s^2\,\sin 2\pi h(x_i-x_j)\,\cos 2\pi k(y_i-y_j)\,\cos 2\pi l(z_i-z_j)$
$X_i\,\beta_j^{11}$	$2i\pi\,h^3\,G_iG_j^*\,E$	$\simeq 4\pi h^3\,\sin 2\pi h(x_i-x_j)\,\cos 2\pi k(y_i-y_j)\,\cos 2\pi l(z_i-z_j)$
$X_i\,\beta_j^{22}$	$2i\pi\,hk^2\,G_iG_j^*\,E$	$\simeq 4\pi\,hk^2\,\sin 2\pi h(x_i-x_j)\,\cos 2\pi k(y_i-y_j)\,\cos 2\pi l(z_i-z_j)$
$X_i\,\beta_j^{12}$	$2i\pi\,h^2k\,G_iG_j^*\,E$	$\simeq 4\pi\,h^2k\,\cos 2\pi h(x_i-x_j)\,\sin 2\pi k(y_i-y_j)\,\cos 2\pi l(z_i-z_j)$
$X_i\,\beta_j^{23}$	$2i\pi hkl\,G_iG_j^*\,E$	$\simeq -4\pi\,hkl\,\sin 2\pi h(x_i-x_j)\,\sin 2\pi k(y_i-y_j)\,\sin 2\pi l(z_i-z_j)$

$$U_{4\mathbf{H}}^{xy} = -\,8\pi^2\,hk\,\sin 2\pi h(x_i-x_j)\,\sin 2\pi k(y_i-y_j) \times \sin 2\pi l(z_i-z_j) \qquad (25)$$

for a general reflection with no zero index. The indices h, k and l are all positive here and, as the sine function has a constant sign for arguments in the interval $]0, \pi[$ or in the interval $]-\pi, 0[$, the quantity $U_{4\mathbf{H}}^{xy}$ (Table 1) has a constant sign for a large number of reflections when the Patterson vector length is small.

The summation of the approximated $\langle W_{\mathbf{H}}U_{\mathbf{H}}\rangle$ values is then partially additive in absolute value. However, when the Patterson vector length becomes very short, the factor $U_{4\mathbf{H}}^{xy}$ tends to zero.

For short non-zero Patterson vector lengths, the normalized matrix elements $\mathbf{A}'(x_i, y_j)$ can take significant values relative to unity. This is in agreement with the results found by Templeton (1999) for cross-coordinate elements in the case of a structure refined against the structure-factor moduli. The expression of A_{xx} and A_{xy} elements for a triclinic $P1$ crystal can be generalized to space groups with several symmetry operators (see Appendix A). For simplicity of the equations, the expressions of the matrix elements related to thermal parameters in the following paragraphs are treated in space group $P1$ only; they can similarly be generalized for all space groups.

3.5. Isotropic thermal parameters

The thermal displacement parameters can be incorporated in the generalized scattering factor G [equation (6)]. The derivative of the structure factor with respect to a thermal parameter b_i is then

$$\partial F_{\mathbf{H}}/\partial b_i = (\partial G_i/\partial b_i)\exp(2i\pi\,\mathbf{H}\cdot\mathbf{X}_i). \qquad (26)$$

The values of the previously defined $U_{\mathbf{H}}$ and $V_{\mathbf{H}}$ quantities relating to a pair of thermal parameters (b_i, b_j) are

$$U_{\mathbf{H}} = 2I_{\mathbf{H}}\,{\rm Re}[(\partial G_i/\partial b_i)\,(\partial G_j^*/\partial b_j)\,E] \qquad (27)$$

and

$$V_{\mathbf{H}} = 2{\rm Re}[F_{\mathbf{H}}^{*2}\,(\partial G_i/\partial b_i)\,(\partial G_j/\partial b_j)\,E']. \qquad (28)$$

The second terms $V_{\mathbf{H}}$ have obviously a random sign, as in the case of coordinate parameters. The sign of $U_{\mathbf{H}}$ depends on the value of the product of the derivatives and has to be considered in detail.

The derivative of G_i with respect to an isotropic thermal motion parameter is

$$\partial G_i/\partial b_i = -s^2\,G_i. \qquad (29)$$

The quantity $G_iG_j^*$ and consequently the product of the derivatives are close to positive real numbers (Table 1). The expression of the $U_{\mathbf{H}}$ terms is then

$$U_{\mathbf{H}} = 2s^4\,I_{\mathbf{H}}\,{\rm Re}(G_iG_j^*\,E). \qquad (30)$$

A large number of $U_{\mathbf{H}}$ terms [equations (27) and (30)] are positive when the Patterson vector is short and the summation of an element A_{bb} is then partly additive in absolute value. The normalized matrix elements A'_{bb} concerning two isotropic thermal parameters have thus the same behaviour as the A'_{xx} elements: they tend globally to decrease from unity to small values as the interatomic distance increases.

3.6. Anisotropic thermal parameters

When the anisotropic motion is modelled, the diagonal β^{ii} and non-diagonal β^{ij} elements of the thermal tensor have to be distinguished. The derivatives with respect to the parameters

electronic reprint

Table 2
Crystallographic data on the high-resolution structures used for the analysis of the normal matrix in practical cases.

	Crambin	Toxin	Aldose reductase	NAD^+
Unit-cell volume (10^3 Å^3)	16.9	55.3	154	0.77
Space group	$P2_1$	$P2_12_12_1$	$P2_1$	$P1$
Molecular weight (kDa)	4.7	7.1	36	0.73
No. of amino acids	46	64	315	–
No. of atoms in asymmetric unit	1000	1080	5884	83
Resolution (Å)	0.54	0.96	0.65	0.50
No. of reflections with $I > \sigma(I)$	111350	32890	410000	12590

β_i^{11} and β_j^{12} concerning the atoms a_i and a_j are, for instance, respectively

$$\partial G_i/\partial\beta_i^{11} = -h^2\, G_i \quad \text{and} \quad \partial G_j/\partial\beta_j^{12} = -2hk\, G_j. \qquad (31)$$

The additivity in absolute value of the $U_{\mathbf{H}}$ terms depends on the sign of the fourth degree polynomial function of the reflection indices (h, k and l) in the product of the derivatives (Table 1). For simplicity of the following equations, the generalized scattering factors G are here approximated to be positive real quantities.

When the factors h, k and l all appear with an even exponent, the $U_{\mathbf{H}}$ terms are positive for a large number of reflections and their summation is partly additive in absolute value for atoms close to each other in the crystal. This is clearly the case for homologous parameter pairs like $(\beta_i^{11}, \beta_j^{11})$ or $(\beta_i^{12}, \beta_j^{12})$ and the corresponding normalized matrix elements are large for short Patterson vector length, tending to unity when the length tends to zero. For non-homologous parameter pairs of type $(\beta_i^{11}, \beta_j^{22})$ or $(B_i^{\rm iso}, \beta_j^{11})$, the summation is also partly additive in absolute value, but the normalized matrix elements take variable values below unity for zero Patterson vector lengths.

In the case of parameter pairs of type $(\beta_i^{12}, \beta_j^{13})$, $(\beta_i^{11}, \beta_j^{23})$ or $(\beta_i^{11}, \beta_j^{12})$, which result in odd exponents for some reflection indices (Table 1), the derivative products can have a negative or a positive sign, depending on the reflection indices. Like for the cross-coordinate elements A_{xy}, the summation of $U_{\mathbf{H}}$ can be regrouped over the four $(+h, \pm k, \pm l)$ reflections in order to identify a partial additivity in absolute value. Similar to the case of the quantity $U_{4\mathbf{H}}^{xy}$ defined by equation (24), the approximate average of $U_{\mathbf{H}}$ over the four reflections leads to the quantity $U_{4\mathbf{H}}^{\beta_i\beta_j}$ (see Table 1). The normal-matrix elements can be expected to be large for short non-zero Patterson vectors only, as a large number of $U_{4\mathbf{H}}^{\beta_i\beta_j}$ terms then have the same sign. The magnitudes of the normal-matrix elements relative to β^{ij} parameters have been analysed in a real case, as described in §4.6.

Some finer effects in three dimensions may also be described. The Patterson vectors in some directions are more likely to produce large normal-matrix elements depending on the thermal displacement parameters considered. For example, the $(\beta_i^{11}, \beta_j^{23})$ or $(\beta_i^{12}, \beta_j^{13})$ parameter pairs generate the same indices polynomial function (Table 1) and can be expected to have similar behaviours. These matrix elements are likely to be large when the Patterson vector between the two considered atoms is small, with a zero length in the x direction and non-zero lengths in the y and z directions.

3.7. Coordinates and temperature factors

For the coordinate/thermal displacement parameter pairs (x_i, b_j), the values of $U_{\mathbf{H}}$ and $V_{\mathbf{H}}$ have the following expressions:

$$U_{\mathbf{H}} = -4\pi\, h\, I_{\mathbf{H}}\, \mathrm{Im}[G_i\, (\partial G_j^*/\partial b_j)\, E] \qquad (32)$$

and

$$V_{\mathbf{H}} = -4\pi\, h\, \mathrm{Im}[F_{\mathbf{H}}^{*2}\, G_i\, (\partial G_j/\partial b_j)\, E']. \qquad (33)$$

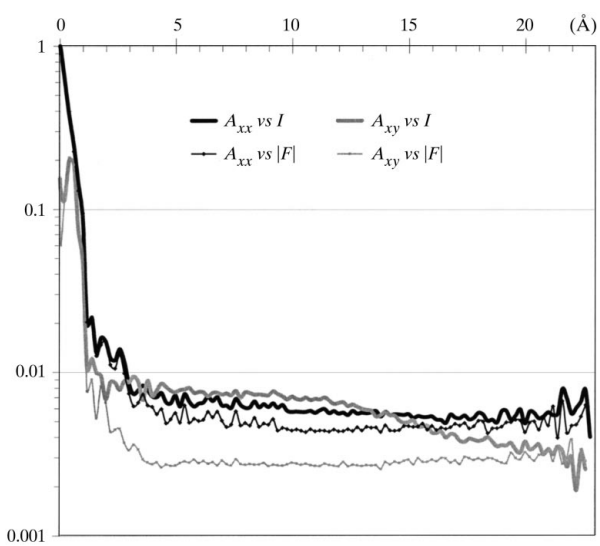

Figure 1
Evolution of the root mean square (r.m.s.) values of the normalized matrix elements $\mathbf{A}'(x_i, x_j)$ (black lines) and $\mathbf{A}'(x_i, y_j)$ (grey lines) as a function of the Patterson vector length in the case of crambin at 0.54 Å resolution. The x coordinates of the non-hydrogen atoms of the protein were refined *versus* the diffraction intensities I (thick lines) or *versus* the structure-factor moduli $|F|$ (thin lines).

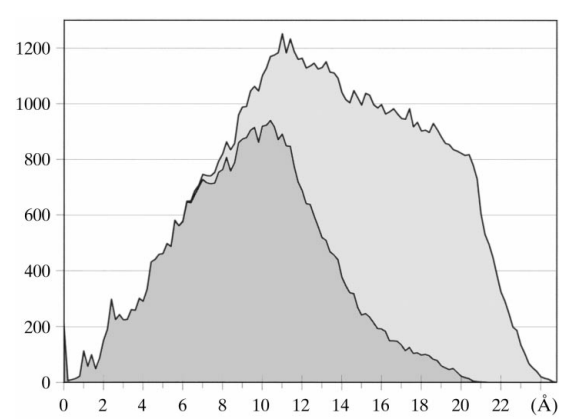

Figure 2
Frequency of the interatomic distances between protein atoms in the crambin crystal. The x axis is divided into shells of 0.1 Å. There are two symmetry operators in space group $P2_1$. The two shortest (different) Patterson vectors between atom pairs and their symmetry equivalents are represented.

electronic reprint

An approximate value of $U_{\mathbf{H}}$ can be obtained by assuming that the generalized scattering factors are real:

$$U_{\mathbf{H}} \simeq -4\pi\ h\, I_{\mathbf{H}}\, G_i\, (\partial G_j/\partial b_j)\ \sin[2\pi\, \mathbf{H}\cdot(\mathbf{X}_i - \mathbf{X}_j)]. \qquad (34)$$

The reflection indices h, k and l, as a third degree polynomial function, are factors in the expression for $U_{\mathbf{H}}$, which does not have a constant sign. Because of the sine factor, the quantity $U_{\mathbf{H}}$ is very small if the two parameters concern the same atom. To identify a partial additivity in absolute value in the $\sum W_{\mathbf{H}} U_{\mathbf{H}}$ summation, the terms can again be regrouped for the four reflections $(+h, \pm k, \pm l)$. The quantity $U_{4\mathbf{H}}^{xb}$ (Table 1) can then be defined similarly to $U_{4\mathbf{H}}^{xy}$ [equation (24)]. The summation of $U_{4\mathbf{H}}^{xb}$ displays some additivity in absolute value (Table 1) and the normal-matrix elements A_{xb} have a propensity to be large for short non-zero Patterson vectors only, similarly to A_{xy}. In three-dimensional space, the most favourable directions for large matrix elements are directly related to the presence of cosine and sine functions in the $U_{4\mathbf{H}}^{xb}$ factor (Table 1).

3.8. Charge-density parameters

The analytical expressions for $U_{\mathbf{H}}$ and $V_{\mathbf{H}}$ by (27) and (28) are valid for all parameters p which can be expressed in the generalized scattering factor G [equation (6)]. When a charged spherical-atom model is used (Coppens *et al.*, 1979), the population P_v and the expansion coefficient κ of the valence electron density generate $\partial G/\partial p$ derivatives which are real and isotropic. The derivatives of the G factors with respect to the multipolar parameters P_{lm} and their expansion coefficient κ' are, on the other hand, complex quantities and highly anisotropic. These derivatives have more complicated expressions than those for the coordinates and thermal parameters and are therefore not specified here.

The expressions for $U_{\mathbf{H}}$ by (27) and (32) are still valid for charge-density parameters; therefore, the exponential E [equation (15)] is still a factor of the $U_{\mathbf{H}}$ terms and the matrix elements depend on the Patterson vector.

The conclusions drawn on the coordinates and on the thermal parameters can be generalized to the multipole parameters. For all pairs of parameters describing the molecular structure or the charge density, there is a certain degree of additivity in absolute value for short Patterson vectors, depending on the behaviour of the product of the derivatives $\partial[G_i \exp(2i\pi\, \mathbf{H}\cdot\mathbf{X}_i)]/\partial p$. The normal matrix elements can be expected to be large for small Patterson vectors only. Examples of normal-matrix elements related to charge-density parameters are given in the case of the NAD^+ molecule in §4.6. For a given pair of multipolar parameters, the $\sum W_{\mathbf{H}} U_{\mathbf{H}}$ summations in three-dimensional space around the Patterson vector origin have a degree of additivity that is intermediate between A_{xx} and A_{xy}, depending on the relative orientation of the multipoles. For instance, the A'_{DxDy} normalized matrix elements, corresponding to two orthogonal dipoles of a given atom, can be expected to be generally smaller than unity, similarly to the A'_{xy} elements.

4. Matrix sparsity in real cases

The degree of matrix sparsity has been investigated for four systems of different size and resolution (Table 2): (i) a small protein at subatomic resolution, crambin at 0.54 Å (Jelsch *et al.*, 2000); (ii) a small protein at atomic resolution, the toxin of the scorpion *Androctonus Australis* Hector II at 0.94 Å (Housset *et al.*, 2000); (iii) a medium-size protein at subatomic resolution, human aldose reductase at 0.65 Å (Mitschler *et al.*, 2000); (iv) a small molecule at subatomic resolution, nicotinamide adenine dinucleotide (NAD^+) at 0.50 Å (Guillot *et al.*, 2000). The crystallographic data on these systems are

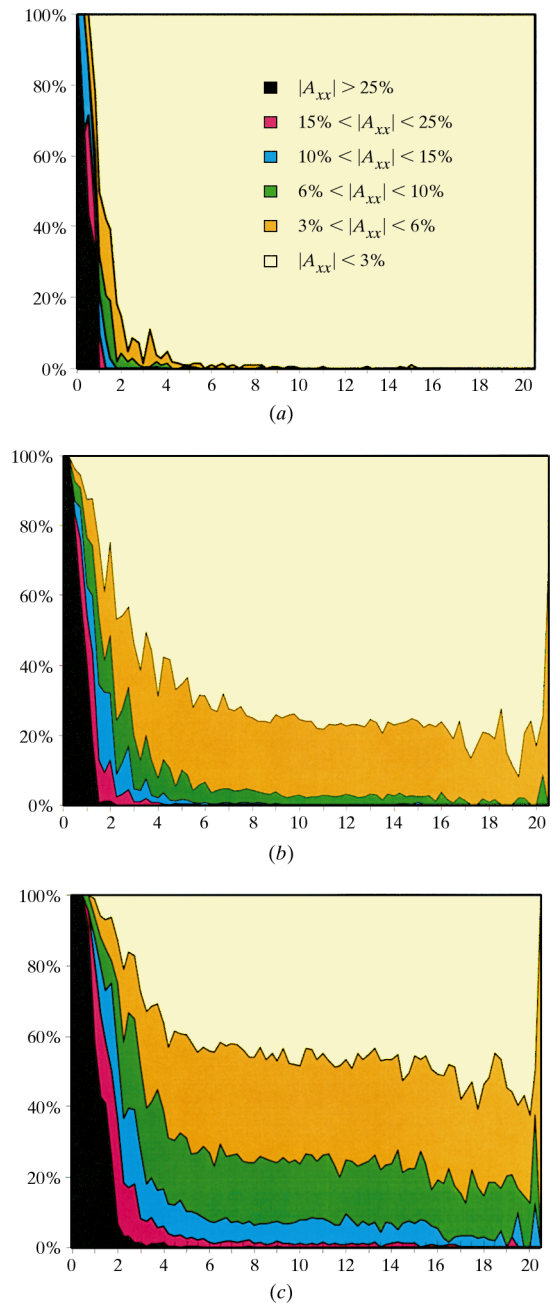

Figure 3
Distribution of the normalized matrix elements $|A'_{xx}|$ in the [0, 1] segment, when applying a resolution cutoff of (*a*) 1 Å, (*b*) 1.5 Å or (*c*) 2 Å to the diffraction data of crambin. The x coordinates of the non-disordered non-hydrogen atoms of the protein were refined.

electronic reprint

summarized in Table 2. These structures were refined with the software *MOPRO* (Guillot *et al.*, 2001) using a model for charged non-spherical multipolar atoms [equation (9)] to describe the molecular electron-density distribution accurately. All the matrix computations were performed using the refined molecular structures unless otherwise specified.

4.1. Crambin at 0.54 Å resolution

In the charge-density analysis of this protein (Jelsch *et al.*, 2000), the normal matrix turned out to be extremely sparse, as can be seen in Fig. 1. The root mean square (r.m.s.) magnitudes of the normalized matrix elements during a cycle of refinement of the atomic coordinates are shown. The value of the A'_{xx} and A'_{xy} matrix elements have a global tendency to decrease when the distance in the crystal between the concerned atoms increases, as pointed out by Templeton (1999). Interatomic distances shorter than 1 Å correspond to the disordered atoms in crambin, which represent as many as 30% of the protein atoms (Yamano *et al.*, 1997).

The r.m.s. magnitudes of the normalized matrix elements are shown for crystallographic refinements performed *versus* the structure-factor moduli and *versus* the diffraction intensities. The two curves have similar shapes. The r.m.s. values of $\mathbf{A}'(x_i, x_j)$ decrease from unity by two orders of magnitude as the Patterson distance varies from 0 to 1.5 Å. On the other hand, the $\mathbf{A}'(x_i, y_j)$ elements have r.m.s. values of 0.15 for pairs of the same atom and a maximum of 0.2 is reached for distances around 0.5 Å. The only large A'_{xx} and A'_{xy} matrix elements correspond to parameter pairs relating to the same atom or to alternative positions of disordered atoms. Even for atoms linked by a covalent bond ($d \simeq 1.5$ Å) in crambin, the related matrix elements are already very small. At longer distances ($d_{ij} > 2$ Å), the r.m.s. values of the matrix elements A'_{xx} and A'_{xy} are relatively stable. In the case of a refinement *versus* the structure-factor moduli, they plateau at the low values of 0.4 and 0.2%, respectively (Fig. 1). On average, the A'_{xy} elements are systematically smaller than the A'_{xx} elements.

These computations were performed on the refined crambin structure $\{wR = [\sum W_{\mathbf{H}}(|F_{\mathbf{H}}^{\rm obs}| - |F_{\mathbf{H}}|)^2 / \sum W_{\mathbf{H}}|F_{\mathbf{H}}^{\rm obs}|^2]^{1/2} = 9\%\}$. In order to assess if these results are still valid for a molecular model far from the converged structure, errors were introduced into the coordinates (up to ±1 Å) and into the anisotropic thermal motion parameters (up to ±10%). The generated molecular structure has a high crystallographic residual (wR = 65%). The elements A'_{xx} and A'_{xy} of the resulting normal matrix display a similar curve to that in Fig. 1.

It must be remembered that the density of interatomic distances in the crystal (represented in Fig. 2) is not uniform. For a given pair of atoms, there are four different Patterson vectors (modulo the translations of unit-cell vectors) in the case of the space group $P2_1$. In the crambin crystal, the β angle happens to be close to 90°; therefore, the four Patterson vectors can be grouped into two pairs of nearly equal lengths, as represented in Fig. 2. In this paper, the magnitudes of the normal-matrix elements are always analysed (Fig. 1) with respect to the shortest interatomic distance in the crystal. The shortest distance in the crystal between two atoms of crambin is always in the range 0–25 Å and the highest frequencies are found at intermediate distances around 11 Å. The distances shorter than 3 Å represent a small proportion (5%) of the total number of atom pairs in the case of the crambin crystal.

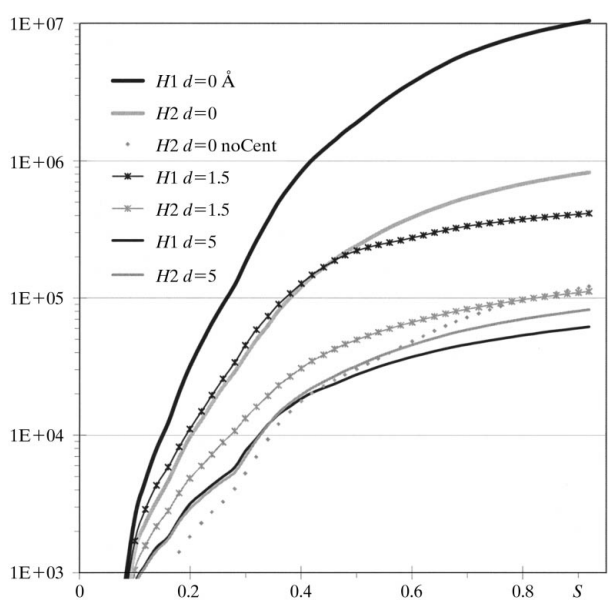

Figure 4
Evolution of the $H1 = \sum_{\mathbf{H}} W_{\mathbf{H}} U_H$ (black lines) and $H2 = \sum_{\mathbf{H}} W_{\mathbf{H}} V_{\mathbf{H}}$ (grey lines) summations as a function of resolution ($d = 1/2s$) in the case of crambin. The x coordinates of the 423 non-disordered carbon atoms of crambin were refined. The r.m.s. $H1$ and $H2$ values for the three different Patterson vector lengths 0, 1.5 and 5 Å have been computed for samples of 423, 164 and 67 atom pairs, respectively. To compute the terms $U_{\mathbf{H}}$ and $V_{\mathbf{H}}$ according to equations (13) and (14), the selected carbon atoms were each set to be spherical and electrically neutral with an isotropic temperature factor B of 3 Å^2. The other atoms were kept identical to those of the final refined structure. The $H2$ summation without the centric reflections $h0l$ is also shown for the zero Patterson vector length.

4.2. Crambin at different resolutions

The effect of the resolution on the magnitudes of the normalized matrix elements has been analysed by applying several resolution cutoffs to the crambin diffraction data (Fig. 3). When all the reflections up to 0.54 Å resolution are used, the r.m.s. values of the matrix elements decrease rapidly with the Patterson vector length to a low level (Fig. 1). This decrease becomes less abrupt when the higher-resolution reflections are omitted. At a 1 Å resolution cutoff, the proportion of matrix elements with magnitudes greater than 6% is still negligible at an interatomic distance of 4 Å (Fig. 3*a*). At a 2 Å resolution cutoff, a significant amount of matrix elements have magnitudes greater than 6% for even the largest Patterson distances. As a consequence, the normal matrix in protein crystallography can be considered very sparse at subatomic and atomic resolution only, whereas, for more usual resolutions (2 Å), a significant number of matrix elements can reach non-negligible values at any interatomic distance.

electronic reprint

From an analytical point of view, the exponential factor E [equation (15)] in the terms of $U_{\mathbf{H}}$ is always equal to unity for diagonal elements ($\mathbf{X}_i = \mathbf{X}_j$). For non-diagonal elements, its phase becomes arbitrary at high resolution and for long interatomic distances. For short Patterson vectors, the summation of the $U_{\mathbf{H}}$ terms is additive in absolute value up to a limited resolution. As a consequence, when the resolution is increased, the off-diagonal elements increase more slowly than the diagonal elements and the normal matrix becomes sparser.

4.3. Summation decomposition

The evolution of the r.m.s. magnitudes of the normal-matrix elements with the resolution and with the Patterson vector length can be investigated further by looking at the decomposition into two terms, $A_{xx} = H1 + H2$ [equation (17)]. In Fig. 4, the variations of the r.m.s. values of the two summations $H1$ and $H2$ with the resolution are represented for crambin. Since the crambin crystals are of space group $P2_1$, the quantities $U_{\mathbf{H}}$ and $V_{\mathbf{H}}$ are each a sum of four terms related to the symmetry-equivalent atoms in the unit cell [see equations (39) and (40) in Appendix A].

For a diagonal element (zero interatomic distance), the summation over $H1 = \sum W_{\mathbf{H}} U_{\mathbf{H}}$ is preponderant compared with $H2 = \sum W_{\mathbf{H}} V_{\mathbf{H}}$ and the ratio between the two sums increases with the resolution, reaching a value of 13 when all the reflections are used (Fig. 4). For the interatomic distance of 1.5 Å, $H1$ is still nearly one order of magnitude larger than $H2$. The summation $H2$ [equation (16)] is thus small compared with $H1$ for short Patterson vectors, as discussed previously in §3.3. The FFT computation of the normal-matrix elements proposed by Agarwal (1978) which neglected the term $H2$ is thus a fully valid approximation.

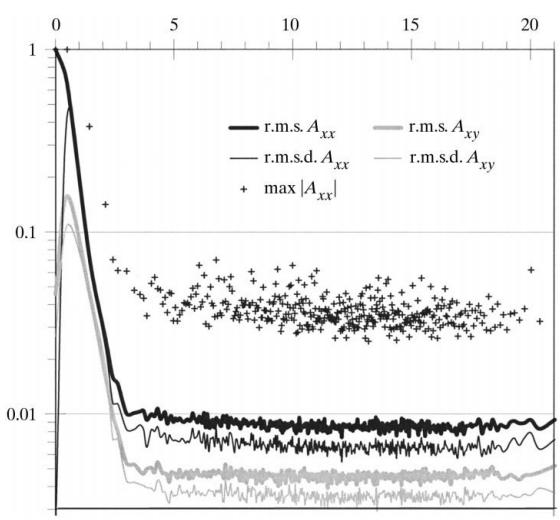

Figure 5
Evolution of the r.m.s. normalized-matrix elements A'_{xx} and A'_{xy} as a function of the Patterson vector length in the case of scorpion toxin at 0.96 Å resolution. The r.m.s. deviation (r.m.s.d.) over samples of 300 $|A'|$ elements and the maximum of $|A_{xx}|$ are also shown. The coordinates of the 481 non-hydrogen non-disordered atoms of the protein were refined *versus* the structure-factor moduli.

The $H2$ summation is not additive in absolute value as the terms $V_{\mathbf{H}}$ contain the factor $F_{\mathbf{H}}^{*2}$, which has an arbitrary phase (for acentric reflections). For zero Patterson vector lengths, the $H2$ summation is not, however, as small as could be expected, as it is one order of magnitude larger (at the highest resolution in Fig. 4) than $H1$ and $H2$ computed for a long interatomic distance (d = 5 Å). This could be caused by the centric reflections ($h0l$) in space group $P2_1$ for which the factor $F_{\mathbf{H}}^{*2}$ is a positive real number. To investigate this point, the $H2$ summation for d = 0 Å has been computed without the centric reflections. The magnitude of $H2$ summed over the acentric reflections only is reduced to the values obtained for long Patterson vectors (Fig. 4). In this example, in space group $P2_1$, the contribution of the centric reflections, which is partially additive in absolute value, turns out to be predominant in the $H2$ summation for pairs of the same atom.

When the interatomic distances are increased from 0 to 5 Å, there is a decline in both $H1$ and $H2$. The decrease of $H1$ is however steeper. For the longest distance represented (d = 5 Å), the $H1$ and $H2$ summations are both non-additive in absolute value and have, on average, similar magnitudes.

4.4. The scorpion toxin

The sparsity of the matrix has also been assessed for a scorpion toxin, the diffraction data of which are available to atomic resolution (Housset *et al.*, 2000). The average level of the normalized matrix elements relative to the coordinates is shown in Fig. 5. The curves of A'_{xx} and A'_{xy} as a function of the Patterson vector length have a similar profile to that of crambin (Fig. 1). For interatomic distances longer than 3 Å, the r.m.s. values of the A'_{xx} and A'_{xy} elements plateau at 0.8 and 0.45%, respectively, at levels about twice high as those of crambin (Fig. 1). However, the magnitude of the matrix elements for a given interatomic distance varies greatly on the plateaus: the r.m.s. deviation represents 75% of the r.m.s.

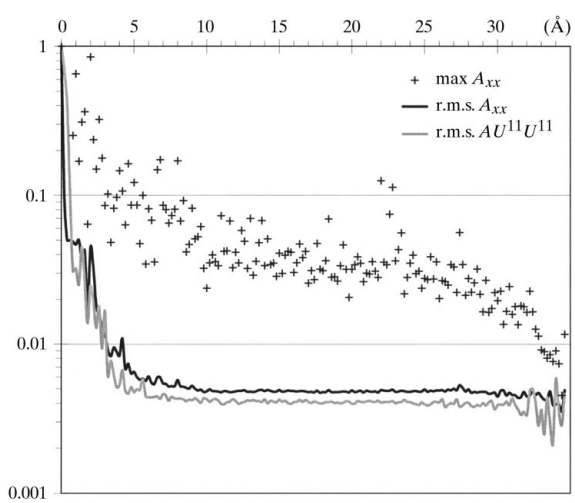

Figure 6
Evolution of the normalized-matrix elements $\mathbf{A}(x, x)$ and $\mathbf{A}(U^{11}, U^{11})$ as a function of the Patterson vector length in the case of aldose reductase at 0.65 Å resolution. The matrix elements have been grouped in shells of 0.2 Å for the calculation of the r.m.s. and maximum values.

electronic reprint

magnitude and the largest $|A'_{xx}|$ elements reach 0.05, which is about five times larger than the r.m.s. value (Fig. 5). For pairs of the same atom, the A'_{xy} elements have a r.m.s. value of 0.045 with a small r.m.s. deviation of 0.01. The A'_{xy} matrix elements related to bonded atoms, in the range $1.3 < d < 1.6$ Å, display a very variable magnitude: the r.m.s. value is 0.05 ± 0.04 and the maximum value is 0.24. The matrix elements show the highest variability, in relative value (r.m.s.d./r.m.s.), for bonded or neighbouring atoms ($1.3 < d < 2.5$ Å).

4.5. Aldose reductase

The magnitudes of the normal-matrix elements were investigated for a medium-size protein: aldose reductase (315 amino acids) at subatomic resolution 0.65 Å (Table 2). For atoms linked by a covalent bond ($d \simeq 1.5$ Å), the A'_{xx} and $A'_{U^{11},U^{11}}$ elements have r.m.s. values around 4% (Fig. 6). For longer distances, they plateau at values ten times smaller, which is comparable with the case of crambin at 0.54 Å (Fig. 1). The highest values of the A'_{xx} elements decrease globally with the Patterson vector lengths and, for distances longer than 10 Å, they are generally lower than 5%.

4.6. A small molecule: NAD$^+$

The normal-matrix elements have also been analysed in the case of a small molecule at subatomic resolution: the oxidized form of nicotinamide adenosine dinucleotide (NAD$^+$), an important redox cofactor in biology. The X-ray diffraction experiment and the molecular structure have been described by Guillot *et al.* (2000) and a charge-density analysis is underway. An example of the behaviour of the normal-matrix elements relative to the anisotropic thermal displacement parameters is given in Fig. 7. The matrix elements relative to the homologous parameter pairs (β^{11}, β^{11}) and (β^{12}, β^{12}) decrease with the interatomic distance in both cases from unity to values around 2%. On the other hand, the diagonal elements of the pairs of the non-homologous parameters (β^{11}, β^{22}), (β^{11}, β^{12}) and (β^{11}, β^{13}) have average values 0.38, -0.45 and -0.25, respectively, with a small standard deviation of 0.01. For the non-diagonal elements, the r.m.s. values are generally smaller compared with pairs of the same parameter. The matrix elements related to coordinates have magnitudes that are comparable with those related to the thermal parameters (figure not shown).

A high-order refinement, in which only the reflections at high resolution are taken into account, is a commonly used procedure for the deconvolution of the thermal motion from the electron-density deformation (Coppens, 1967). To analyse the properties of the normal matrix in such a refinement, the thermal displacement parameters of NAD$^+$ were refined *versus* the high-resolution reflections only (Fig. 7). When the low resolution is truncated at 0.7 Å, the r.m.s. values of the normal-matrix elements are globally reduced by a factor of about four. The magnitudes of the non-diagonal normalized-matrix elements turn out to be generally larger by a factor of four compared with a refinement *versus* all the reflections, which may be related to the reduction of the number of reflections.

For this small molecule, the charge-density parameters of the individual atoms were refined; the magnitudes of the normalized matrix elements related to several pairs of multipoles are shown in Fig. 8. The elements decrease globally with the Patterson vector length and have r.m.s. values around 3% for distances above 2 Å. For zero Patterson vector lengths, the elements related to pairs of the same parameter, like $\mathbf{A}(Dx, Dx)$ and $\mathbf{A}(Qzx, Qzx)$, are equal to unity, while the elements related to different parameters are one order of magnitude smaller. It should be noted that the $\mathbf{A}(Dx, Dx)$ and $\mathbf{A}(Dx, Dy)$ elements have similar r.m.s. values for non-zero interatomic distances ($d > 1.3$ Å), as the orientation of the dipoles (defined in a local axis system) is different for each atom.

4.7. Number of reflections

To what extent does the number of reflections have an effect on the relative magnitudes of the diagonal and off-diagonal normal-matrix elements? For the diagonal elements, the partial summation $\sum W_{\mathbf{H}} U_{\mathbf{H}}$ is additive in absolute value. For off-diagonal elements, the terms tend to have arbitrary phases for sufficiently long Patterson vectors and for high-resolution reflections. The sum of N unitary complex numbers with an arbitrary phase has an expected modulus which grows only with $N^{1/2}$. On the other hand, the sum of the moduli is additive in absolute value and grows with N. The ratio of the two sums tends to zero as $N^{-1/2}$, when the number of terms becomes large. It can thus be expected that the larger the number of reflections, the larger the diagonal elements are likely to be compared with the off-diagonal elements.

In the summation over the reciprocal space of $\sum W_{\mathbf{H}}(U_{\mathbf{H}} + V_{\mathbf{H}})$, the terms have variable moduli and the evolution of their r.m.s. values is not uniform with the resolution. Globally, the factors $|I_{\mathbf{H}} G_i G_j|$ decrease with the resolution. This is concomitantly compensated by the weighting scheme $[W_{\mathbf{H}} = 1/\sigma^2(I_{\mathbf{H}})]$ and other factors depending on the reflection indices (for example h^2 in the case of A_{xx} elements) which increase with the resolution.

The number of reflections augments with the resolution limit of the diffraction data and with the unit-cell dimensions. The number of reflections is largest in the case of macromolecules at atomic or sub-atomic resolution compared with standard protein crystallography or charge-density analyses of small molecules. As the normal matrix is then extremely sparse, it is a very favourable case for neglecting *a priori* most of the non-diagonal elements according to interatomic distance criteria. This is also illustrated by Figs. 1, 3, 4, 5, 6 and 7, which demonstrate that the sparsity of the normal matrix clearly increases with the resolution of the diffraction data and the size of the system.

4.8. Matrix truncation and parameter shifts

Block-diagonal approximations, by which a large number of the non-diagonal elements are set to zero, in conjunction with

electronic reprint

stereochemical restraints, can be very effective in refinement procedures (Scheringer, 1968). The block-diagonal method, with fast Fourier algorithms inspired by Agarwal (1978), has been applied successfully to refine protein structures (Dodson, 1981).

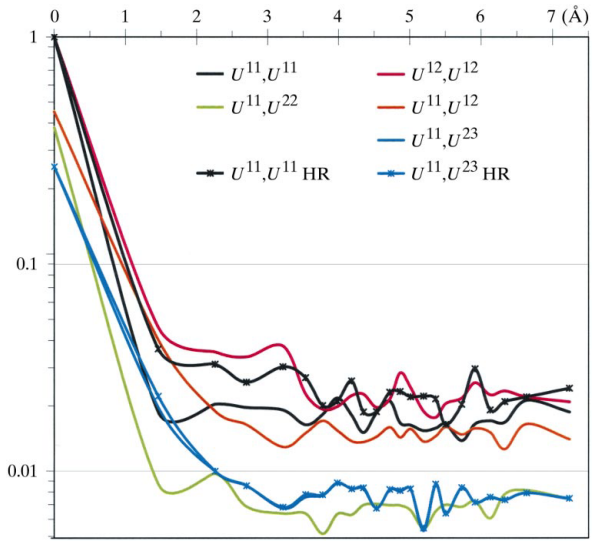

Figure 7
Evolution of the normalized-matrix elements related to the thermal factors U^{ij} as a function of the Patterson vector length in the case of NAD^+. The anisotropic thermal displacement parameters of the 44 NAD^+ non-hydrogen atoms were refined *versus* the structure-factor moduli. The r.m.s. and maximum values are computed over samples of 44 elements. The 'HR' curves are obtained from a high-resolution refinement using only the 8096 reflections with $s = 1/2d > 0.7$ Å.

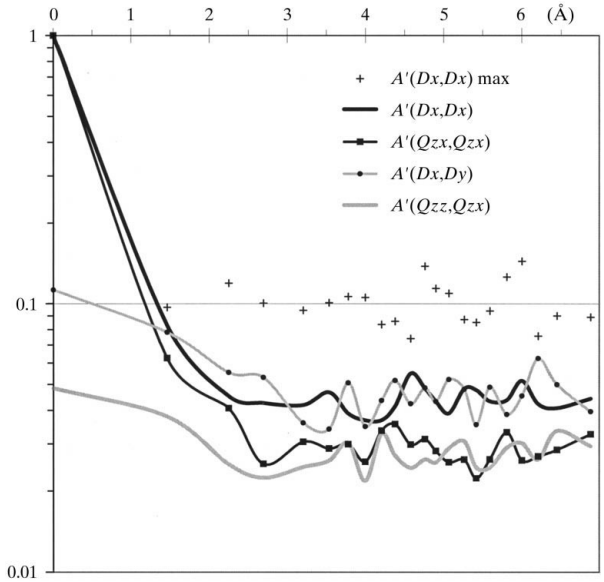

Figure 8
Evolution of the normalized-matrix elements related to charge-density parameters as a function of Patterson vector length in the case of the NAD^+ molecule at 0.50 Å resolution. The dipoles Dx and Dy, and the quadrupoles Qzz and Qzx of the 44 non-hydrogen atoms of NAD^+ were refined *versus* the structure-factor amplitudes. The r.m.s. and maximum values were computed over samples of size 44.

The extent to which the *a priori* small elements of the normal matrix can be set to zero without perturbing the solution of the normal equations has been tested for crambin at different resolutions in a refinement cycle of the x, y, z coordinates of the non-disordered carbon, oxygen and nitrogen atoms of the protein. The matrix elements corresponding to Patterson vectors longer than a given cutoff were systematically set to zero. The shifts of the x coordinates obtained with different cutoffs were compared with those obtained when using the full matrix (Fig. 9). The alteration of the parameter shifts generally decreases as the resolution is varied from 1.5 to 0.54 Å resolution. The solution of the system of normal equations [equation (2)] is significantly altered at the lowest resolutions (1.5 and 1 Å). When all the reflections are used, the discrepancy is modest, with plateaus from 7 to 5% and distance cutoffs in the range 1–10 Å. The solution of the normal equations is more sensitive to the matrix truncation when the resolution of the diffraction data is low. This instability is presumably related to the tendency of the normal matrix to become singular at lower resolution. Cowtan & Ten Eyck (2000), in their analysis of a small protein refinement, observed that the eigenvalue spectrum becomes wider at lower resolutions, with an increasing number of eigenvalues close to zero.

At the highest resolutions of the diffraction data set (0.54, 0.75 and 1. Å), the r.m.s. magnitude of the shifts is very stable (0.649 ± 0.001, 0.71 ± 0.01 and 1.08 ± 0.01, respectively), whatever the distance cutoff applied for the matrix truncation. At 1.5 Å resolution, the r.m.s. magnitude is less stable and

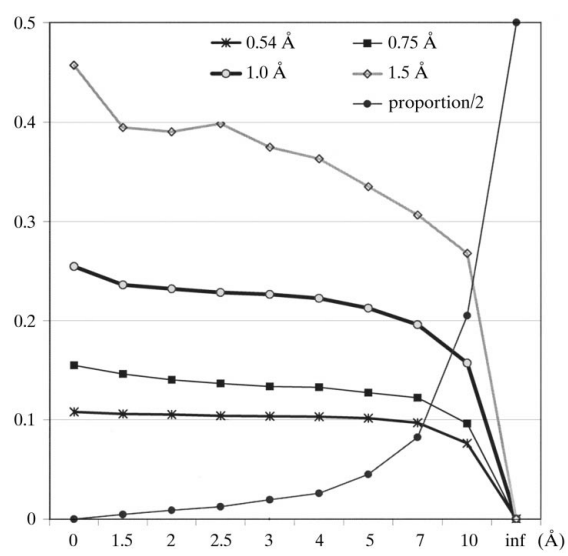

Figure 9
Alteration of the x-coordinate shifts Δx when truncating the normal matrix. The x, y, z coordinates of the non-hydrogen non-disordered atoms of the protein crambin were refined at several high-resolution cutoffs. No restraints were applied. The shifts Δx have been normalized by dividing them by $A_{xx}^{1/2}$, representing estimates of the standard uncertainties of x. The x axis corresponds to the long-distance cutoff (Å) applied in the sparse-matrix truncation. The relative discrepancy R between the normalized shifts $\Delta x'$ and Δx when using the sparse or the full normal matrix are shown at different resolutions. The proportion (divided by 2) of the normal matrix that is used after truncation is also represented. $R = [\sum_{\rm atoms}(\Delta x - \Delta x')^2 / \sum_{\rm atoms} \Delta x^2]^{1/2}$.

electronic reprint

increases gradually from 1.33 to 1.63 with the matrix truncation. There is also a gradual increase of the parameter shifts (normalized with respect to the diagonal elements) when the resolution is lowered; this may be attributed to the fact that the structure used is close to convergence in the refinement at 0.54 Å.

For small molecules like NAD^+ at subatomic resolution, the normal matrix is relatively sparse, as two parameters related to distinct atoms are generally poorly correlated (Figs. 7 and 8). As the calculation times are generally modest, computing only the essential elements of the sparse normal matrix is not crucial for small molecules. For systems of larger size, the matrix elements clearly decrease more rapidly with the interatomic distance, and to lower levels, while the computer times become larger. In such cases, truncating the normal matrix according to distance criteria is a fully valid approximation in the least-squares method and saves considerable computing time in the case of large structures at very high resolution.

5. Conclusions

The expected magnitudes of the normal-matrix elements are highly dependent on the distance in the crystal between the two atoms concerned. The importance of the shortest Patterson vector has been demonstrated analytically for the triclinic $P1$ case and, in Appendix A, for all space groups. For a macromolecule at (sub)atomic resolution, the majority of the positional and thermal motion parameters are uncorrelated, as the corresponding normal-matrix elements are small. This is systematically the case for a pair of refined parameters when the distance between the two atoms concerned is typically larger than three times the resolution. The same conclusions can be drawn for parameters describing the charge density of a molecule.

This *a priori* knowledge that a large number of off-diagonal terms in the least-squares matrix are small and can be neglected means that the full normal matrix does not have to be computed. With the conjugate-gradient algorithm, which has been implemented in the program *MOPRO* (Guillot *et al.*, 2001), it is not necessary to compute and to store the full normal matrix and its inverse, thus saving computer time and memory in the case of large systems. An automated routine in *MOPRO* is able to select, according to a distance criterion, the matrix elements for computation. The application of this sparse-matrix principle is an approximation that is particularly efficient in the case of macromolecules at very high resolution.

APPENDIX *A*
Generalization to other space groups

The expression of the A_{xx} elements in the triclinic $P1$ case can be generalized to other space groups, where there are several symmetry operators σ_m in the unit cell. The operators σ_m are isometries which can be expressed as

$$\sigma_m \mathbf{X} = \mathbf{S}_m \mathbf{X} + \mathbf{T}_m, \tag{35}$$

where $\mathbf{S}_m$ is the isometric matrix and $\mathbf{T}_m$ is the translation vector related to the symmetry operation. The expression of the structure factor is then the sum over all the symmetry operators and all the atoms a in the asymmetric unit:

$$F_{\mathbf{H}} = \sum_a \sum_m G_a \exp(2i\pi\, \mathbf{H} \cdot \sigma_m \mathbf{X}_a). \tag{36}$$

The derivative of the structure factor with respect to the coordinate p_i is then

$$\partial F_{\mathbf{H}}/\partial p_i = 2i\pi\, G_i \sum_m \langle \mathbf{H} \cdot \mathbf{S}_m \mathbf{u} \rangle \exp(2i\pi\, \mathbf{H} \cdot \sigma_m \mathbf{X}_j), \tag{37}$$

where $\mathbf{u}$ is the unit-cell vector (among $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$) associated with the coordinate p_i. For an A_{xx} element, the values of the previously defined $U_{\mathbf{H}}$ and $V_{\mathbf{H}}$ quantities [equations (20) and (21)] are then obtained by double summation over the symmetry operators:

$$U_{\mathbf{H}} = \sum_m \sum_n U_{\mathbf{H}}^{m,n} \quad \text{and} \quad V_{\mathbf{H}} = \sum_m \sum_n V_{\mathbf{H}}^{m,n}, \tag{38}$$

where

$$U_{\mathbf{H}}^{m,n} = 8\pi^2 \langle \mathbf{H} \cdot \mathbf{S}_m \mathbf{u} \rangle \langle \mathbf{H} \cdot \mathbf{S}_n \mathbf{u} \rangle I_{\mathbf{H}} \times \mathrm{Re}\{G_i G_j^* \exp[2i\pi\, \mathbf{H} \cdot (\sigma_m \mathbf{X}_i - \sigma_n \mathbf{X}_j)]\} \tag{39}$$

and

$$V_{\mathbf{H}}^{m,n} = -8\pi^2 \langle \mathbf{H} \cdot \mathbf{S}_m \mathbf{u} \rangle \langle \mathbf{H} \cdot \mathbf{S}_n \mathbf{u} \rangle \times \mathrm{Re}\{F_{\mathbf{H}}^{*2} G_i G_j \exp[2i\pi\, \mathbf{H} \cdot (\sigma_m \mathbf{X}_i + \sigma_n \mathbf{X}_j)]\}. \tag{40}$$

The summation over the reflections of the terms $U_{\mathbf{H}}^{m,n}$ and $V_{\mathbf{H}}^{m,n}$ is carried over an independent fraction of the reciprocal space. The summation of the second terms $V_{\mathbf{H}}^{m,n}$ is, in general, non-additive in absolute value, as in the case of the triclinic space group $P1$. It should be noted that, for centrosymmetric space groups, $F_{\mathbf{H}}^{*2}$ is a positive real number and the $U_{\mathbf{H}}$ and $V_{\mathbf{H}}$ terms are actually equal.

Let (a_i, a_j) be a pair of atoms in the asymmetric unit, close to each other (*i.e.* $\mathbf{X}_i - \mathbf{X}_j$ is small). All the same symmetry-related atom pairs $(\sigma_p a_i, \sigma_p a_j)$ are then equally close to each other since the operators σ_p are isometric. As the factor $\langle \mathbf{H} \cdot \mathbf{S}_p \mathbf{u} \rangle^2$ [equation (39)] is always positive, the corresponding $U_{\mathbf{H}}^{p,p}$ terms have a positive sign for many reflections. In the summation $\sum_{\mathbf{H}} \sum_m \sum_n U_{\mathbf{H}}^{m,n}$, the portion due to the $U_{\mathbf{H}}^{p,p}$ terms is partly additive in absolute value and the normal-matrix element consequently takes large values. The same conclusion can be drawn for two atoms $(a_i, \tau a_j)$ that are close to each other after application of a unit-cell vector translation τ to the atom a_j, as the matrix $\mathbf{S}_\tau$ [equation (39)] of the symmetry operator is equal to the identity I.

More generally, when one of the Patterson vectors $\sigma_m \mathbf{X}_i - \mathbf{X}_j$ is small, the corresponding exponential factor in $U_{\mathbf{H}}^{m,I}$ is close to unity, or has a positive real part for a large number of reflections, and the partial summation $\sum_{\mathbf{H}} U_{\mathbf{H}}^{m,I}$ takes large values. In addition, the Patterson vectors $\sigma_p \sigma_m \mathbf{X}_i - \sigma_p \mathbf{X}_j$ between the symmetry-related atoms are equally short. The sign of the corresponding $U_{\mathbf{H}}^{pm,p}$ terms depends, however, on the factors $\langle \mathbf{H} \cdot \mathbf{S}_p \mathbf{S}_m \mathbf{u} \rangle \langle \mathbf{H} \cdot \mathbf{S}_p \mathbf{u} \rangle$ [equa-

electronic reprint

tion (39)]. When $\mathbf{S}_m\mathbf{u}$ and $\mathbf{u}$ happen to be equal, these factors are then equal to the square of $\langle \mathbf{H}\cdot\mathbf{S}_p\mathbf{u}\rangle$ and are therefore positive. Similarly, when $\mathbf{S}_m\mathbf{u}$ and $\mathbf{u}$ are opposite, these factors are always negative, and the normal-matrix element takes large negative values since the summation is partially additive in absolute value with respect to the $U_{\mathbf{H}}^{pm,p}$ terms. In triclinic, monoclinic or orthorhombic space groups, the linear operators $\mathbf{S}_m$ are necessarily a twofold axis, a mirror plane or the inversion, and the vectors $\mathbf{S}_m\mathbf{u}$ and $\mathbf{u}$ are consequently always collinear.

In space groups possessing threefold or fourfold axes, these two vectors can be non-collinear and the factors $\langle \mathbf{H}\cdot\mathbf{S}_p\mathbf{S}_m\mathbf{u}\rangle\langle \mathbf{H}\cdot\mathbf{S}_p\mathbf{u}\rangle$ are not necessarily always of the same sign. Depending on the space-group symmetry operators and on the direction $\mathbf{u}$, the number of factors of one sign might exceed the number with the other sign. Alternatively, in some cases, the number of positive and negative factors might be globally equivalent and the summation of the $U_{\mathbf{H}}$ terms might not take large values. This can for instance occur for A_{xx} elements in space group $P4$ with the fourfold rotation axis $\sigma = 4_c$. Let a_i and a_j be two atoms in the asymmetric unit with a short interatomic distance $d(a_i, \sigma a_j)$. The factor $\langle \mathbf{H}\cdot\sigma\mathbf{a}\rangle\langle \mathbf{H}\cdot\mathbf{a}\rangle$ is here equal to hk, which is reminiscent of the expression of A_{xy} in space group $P1$ [equation (20)]. The four atom pairs $(\sigma^{p+1}a_i, \sigma^p a_j)$ form short Patterson vectors and the four factors $\langle \mathbf{H}\cdot\mathbf{S}_p\mathbf{S}_m\mathbf{u}\rangle\langle \mathbf{H}\cdot\mathbf{S}_p\mathbf{u}\rangle = \langle \mathbf{H}\cdot\sigma^{p+1}\mathbf{a}\rangle\langle \mathbf{H}\cdot\sigma^p\mathbf{a}\rangle$ are equal to hk for $p = 0$ or 2 and to $-hk$ for $p = 1$ or 3. As a result, the signs of the four $U_{\mathbf{H}}^{\sigma^{p+1},\sigma^p}$ terms compensate one another globally in the A_{xx} summation. In this example, the magnitudes of the A_{xx} elements for pairs of neighbouring atoms, like $(a_i, \sigma a_j)$, are expected to follow a similar trend to those described for A_{xy} in space group $P1$ (§3.4). This is related to the fact that a shift applied in the x direction on atom a_j results in a movement of the symmetry-related atom σa_j in the y direction. It can also be shown that the A_{xy} and A_{yx} elements corresponding to atoms a_i and σa_j are large and tend respectively to -1 and 1 for short $(a_i, \sigma a_j)$ distances.

Similarly to the triclinic $P1$ case, the A_{xx} elements of the normal matrix in other space groups are expected to be large when the two atoms concerned are close to each other in the asymmetric unit or after application of unit-cell vector translations. The normalized A_{xx} elements tend to unity when the Patterson vector length tends to zero and it can be shown that A_{xy} take significantly smaller values.

When the two atoms are close to each other after application of a symmetry operation, the magnitudes of the A_{xx}- and A_{xy}-type elements are also likely to be large for short Patterson vectors. However, if the symmetry operator changes the nature of the coordinates, the A_{xx} and A_{xy} profiles may be consequently exchanged or modified.

In the triclinic case, the summation over the reflections in equation (1) is carried out over half of the reciprocal space. In space groups of higher symmetry, the summation is performed over an independent fraction of the reciprocal space and the equivalent reflections in equation (37) are generated by the symmetry operations *via* the $\langle \mathbf{H}\cdot\mathbf{S}_m\mathbf{u}\rangle$ factors.

This research was supported by CNRS. The author gratefully acknowledges N. H. Hansen and A. G. Urzhumtsev for helpful discussions and C. Lecomte for his support and interest in this work.

References

Agarwal, R. C. (1978). *Acta Cryst.* A**34**, 791–809.
Coppens, P. (1967). *Science*, **158**, 1577.
Coppens, P., Guru Row, T. N., Leung, P., Stevens, E. D., Becker, P. J. & Yang, Y. W. (1979). *Acta Cryst.* A**35**, 63–72.
Cowtan, K. & Ten Eyck, L. F. (2000). *Acta Cryst.* D**56**, 842–856
Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1997). *Curr. Opin. Struct. Biol.* **7**, 681–688.
Dodson, E. J. (1981). *Refinement of Protein Structures*, pp. 29–39. Warrington: Daresbury Laboratory.
Franklin, J. N. & Marsh, R. E. (1982). *Acta Cryst.* B**38**, 1681–1682.
Guillot, B., Jelsch, C. & Lecomte, C. (2000). *Acta Cryst.* C**56**, 726–728.
Guillot, B., Viry, L., Guillot, R., Lecomte, C. & Jelsch, C. (2001). *J. Appl. Cryst.* **34**, 214–223.
Hamilton, W. C. (1964). In *Statistics in Physical Science*. New York: Ronald Press.
Hansen, N. K. & Coppens, P. (1978). *Acta Cryst.* A**34**, 909–921.
Hestenes, M. R. & Stiefel, E. (1952). *J. Natl Bur. Stand.* **49**, 409–436.
Housset, D., Pichon-Pesme, V., Jelsch, C., Benabicha, F., Maierhofer, A., David, S., Fontecilla-Camps, J. C. & Lecomte, C. (2000). *Acta Cryst.* D**56**, 151–160.
Jelsch, C., Pichon-Pesme, V., Lecomte, C. & Aubry, A. (1998). *Acta Cryst.* D**54**, 1306–1318.
Jelsch, C., Teeter, M. M., Lamzin, V., Pichon-Pesme, V., Blessing, R. H. & Lecomte, C. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 3171–3176.
Konnert, J. H. (1976). *Acta Cryst.* A**32**, 614–617.
Longhi, S., Czjzek, M. & Cambillau, C. (1998). *Curr. Opin. Struct. Biol.* **8**, 730–737.
Lunin, V. Y. & Urzhumtsev, A. G. (1985). *Acta Cryst.* A**41**, 327–333.
Mitschler, A., Sanishvili, R., Joachimiak, A., Howard, E., Barth, P., Lamour, V., Guillot, B., Van Zandt, M., Sibley, E., Moras, D. & Podjarny, A. (2000). 19th European Crystallographic Meeting, 25–31 August 2000, Nancy, France. Poster s7.m0.p1.
Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. (1999). *Acta Cryst.* D**55**, 247–255.
Rae, A. D. (1978). *Acta Cryst.* A**34**, 578–582.
Scheringer, C. (1968). *Acta Cryst*. B**24**, 662–666.
Sheldrick, G. M. & Schneider, T. (1997). *SHELXL: High-Resolution Refinement*, in *Methods in Enzymology*, Vol. 276, *Macromolecular Crystallography*, *Part B*, edited by C. W. Carter Jr & R. M. Sweet, pp. 319–343. New York: Academic Press.
Templeton, D. H. (1999). *Acta Cryst.* A**55**, 695–699.
Tronrud, D. E. (1999). *Acta Cryst.* A**55**, 700–703.
Yamano, A., Heo, N. H. & Teeter, M. M. (1997). *J. Biol. Chem.* **272**, 9597–9600.

electronic reprint