Acta Crystallographica Section A
Foundations and Advances
ISSN 2053-2733

Received 26 June 2014
Accepted 22 August 2014

From direct-space discrepancy functions to crystallographic least squares

Carmelo Giacovazzo

Istituto di Cristallografia - CNR, Via G. Amendola, 122/O 70126 Bari, Italy. Correspondence e-mail: carmelo.giacovazzo@ic.cnr.it

Crystallographic least squares are a fundamental tool for crystal structure analysis. In this paper their properties are derived from functions estimating the degree of similarity between two electron-density maps. The new approach leads also to modifications of the standard least-squares procedures, potentially able to improve their efficiency. The role of the scaling factor between observed and model amplitudes is analysed: the concept of *unlocated model* is discussed and its scattering contribution is combined with that arising from the *located model*. Also, the possible use of an ancillary parameter, to be associated with the classical weight related to the variance of the observed amplitudes, is studied. The crystallographic discrepancy factors, basic tools often combined with least-squares procedures in phasing approaches, are analysed. The mathematical approach here described includes, as a special case, the so-called vector refinement, used when accurate estimates of the target phases are available.

1. Notation

ρ, ρ_p, electron densities of the target and of the model structure, respectively.

$\rho_{\rm d} = \rho - \rho_p$, *ideal difference electron density*. Summed to ρ_p it exactly provides ρ, no matter the quality of ρ_p.

N, number of atoms in the unit cell for the target structure.

N_p, number of atoms in the unit cell for the model structure. Usually $N_p \le N$, but it may also be $N_p > N$.

f_j, $j = 1, \ldots, N$, atomic scattering factors for the target structure (thermal factor included).

$F_{\mathbf{h}} = |F_{\mathbf{h}}|\exp(i\varphi_{\mathbf{h}}) = A_{\mathbf{h}} + iB_{\mathbf{h}} = \sum_{j=1}^{N} f_j \exp(2\pi i\mathbf{hr}_j)$, structure factor of the target structure.

$F_{p\mathbf{h}} = |F_{p\mathbf{h}}|\exp(i\varphi_{p\mathbf{h}}) = A_{p\mathbf{h}} + iB_{p\mathbf{h}} = \sum_{j=1}^{p} f_j \exp(2\pi i\mathbf{hr}'_j)$, where $\mathbf{r}'_j = \mathbf{r}_j + \Delta\mathbf{r}_j$. Structure factor of the model structure.

$F_{\rm d\mathbf{h}} = F_{\mathbf{h}} - F_{p\mathbf{h}} = |F_{\rm d\mathbf{h}}|\exp(i\varphi_{\rm d\mathbf{h}})$, structure factor of the *ideal difference structure*.

$\sum_N = \sum_{j=1}^{N} f_j^2$, $\sum_{N_p} = \sum_{j=1}^{N_p} f_j^2$.

$D = \langle\cos(2\pi\mathbf{h}\Delta\mathbf{r})\rangle$, the average is performed per resolution shell.

$\sigma_A = D(\Sigma_p/\Sigma_N)^{1/2}$.

$\sigma_R^2 = \langle|\mu|^2\rangle / \sum_N$, $\langle|\mu|^2\rangle$ is the measurement error.

$e = 1 + \sigma_R^2$.

$I_i(x)$, modified Bessel function of order i.

$m = \langle\cos(\varphi - \varphi_p)\rangle = I_1(X)/I_0(X)$, where $X = 2\sigma_A|EE_p|/(e - \sigma_A^2)$.

$|E|$ and $|E_p|$ are the normalized structure-factor moduli corresponding to F and F_p, respectively.

EDM, electron-density modification.

2. Introduction

Crystallographic least squares minimize a discrepancy function between observed and calculated structure-factor amplitudes to optimize the structural parameters of the model given the target diffraction amplitudes. They are usually based on the Gauss–Newton approach: more recently maximum-likelihood least-squares procedures (Bricogne, 1997; Pannu & Read, 1996; Murshudov *et al.*, 1997) have been introduced. All such methods work in reciprocal space and rely on the statistical properties of the structure factors. Their application to observed diffraction amplitudes led to hundreds of thousands of unbiased crystal structure models, the heritage of modern crystallography.

Functions estimating the degree of similarity of two electron-density maps were established a long time ago and are used not only in the latest stages of the crystal structure analysis but also in various steps of the phasing process. This paper is mainly concerned with establishing in direct space some functions estimating the degree of similarity between the target and the model electron densities and, from them, deriving least-squares procedures. As we will see, this approach is potentially able to modify some aspects of the standard crystallographic least squares. We will only consider the Gauss–Newton method: our approach has only a speculative nature and provides the theoretical basis for future applications.

Let us suppose that the target structure $\rho(\mathbf{r})$ has been perfectly identified, and that $|F_{\mathbf{h}}|$ and $\varphi_{\mathbf{h}}$ are the corresponding amplitudes and phases, respectively. Let $\rho_p(\mathbf{r})$ be a structural model available at a certain step of the phasing approach, to which amplitudes $|F_{p\mathbf{h}}|$ and phases $\varphi_{p\mathbf{h}}$ correspond. If we want to assess the similarity between target and model, the integral relationship

$$I = \int_V \rho_{\rm d}^2(\mathbf{r})\,{\rm d}\mathbf{r} \equiv \int_V [\rho(\mathbf{r}) - \rho_p(\mathbf{r})]^2\,{\rm d}\mathbf{r} \qquad (1)$$

should be calculated, no matter if in direct or in reciprocal space. In this last case

$$\int_V \rho^2(\mathbf{r})\,{\rm d}\mathbf{r} = \frac{1}{V}\sum_{\mathbf{h}} |F_{\mathbf{h}}|^2,$$

$$\int_V \rho_p^2(\mathbf{r})\,{\rm d}\mathbf{r} = \frac{1}{V}\sum_{\mathbf{h}} |F_{p\mathbf{h}}|^2,$$

$$\int_V \rho(\mathbf{r})\rho_p(\mathbf{r})\,{\rm d}\mathbf{r} = \frac{1}{V}\sum_{\mathbf{h}} |F_{\mathbf{h}}||F_{p\mathbf{h}}|\cos(\varphi_{\mathbf{h}} - \varphi_{p\mathbf{h}}),$$

and then

$$I = \frac{1}{V}\sum_{\mathbf{h}} |F_{{\rm d}\mathbf{h}}|^2 = \frac{1}{V}\sum_{\mathbf{h}} [|F_{\mathbf{h}}|^2 + |F_{p\mathbf{h}}|^2 - 2|F_{\mathbf{h}}||F_{p\mathbf{h}}|\cos(\varphi_{\mathbf{h}} - \varphi_{p\mathbf{h}})]. \qquad (2)$$

The integral I, in reciprocal space, is therefore equal (but for the scaling factor V^{-1}) to the sum of the squared structure factor of the ideal difference electron density $\rho_{\rm d}$. Such an integral enjoys various good features: it vanishes when the model coincides with the target and gets its maximum value when the model is uncorrelated with the target. In this last condition $\langle\cos(\varphi_{\mathbf{h}} - \varphi_{p\mathbf{h}})\rangle$ is expected to vanish, and equation (2) reduces to

$$\frac{1}{V}\sum_{\mathbf{h}} (|F_{\mathbf{h}}|^2 + |F_{p\mathbf{h}}|^2) = \int_V [\rho^2(\mathbf{r}) + \rho_p^2(\mathbf{r})]\,{\rm d}\mathbf{r}.$$

I is strictly connected with crystallographic least squares. Indeed minimizing equation (2) (for simplicity we omit the constant term V^{-1}) according to

$$I \simeq \sum_{\mathbf{h}} [|F_{\mathbf{h}}|^2 + |F_{p\mathbf{h}}|^2 - 2|F_{\mathbf{h}}||F_{p\mathbf{h}}|\cos(\varphi_{\mathbf{h}} - \varphi_{p\mathbf{h}})] = \min \qquad (3)$$

allows us to modify the model structural parameters (atomic coordinates, vibrational parameters, site occupancy *etc.*) provided $|F_{\mathbf{h}}|$, $|F_{p\mathbf{h}}|$, $\varphi_{\mathbf{h}}$ and $\varphi_{p\mathbf{h}}$ are known. If the $\varphi_{\mathbf{h}}$ values are unknown (as occurs when a phasing attempt is started) equation (3) cannot be applied: a simple way of overcoming the difficulty is to associate to $\varphi_{\mathbf{h}}$ its best current estimate of the model phase, say

$$\varphi_{\mathbf{h}} \simeq \varphi_{p\mathbf{h}}. \qquad (4)$$

Then equation (3) reduces to

$$I_{\rm s}(|F|) = \sum_{\mathbf{h}} (|F_{\mathbf{h}}| - |F_{p\mathbf{h}}|)^2 = \min, \qquad (5)$$

which is only based on amplitudes and does not take phases into account. Equation (5), however, is a special case of equation (3), working only under the condition (4): therefore equation (5) should work well or badly according to the quality of the model. The condition (4) is not explicitly considered in the standard Gauss–Newton least-squares approach, but it agrees well with a well known limit of the crystallographic least squares. Indeed the minimization of the function $I_{\rm s}$ for non-linear problems may be very complicated, in particular it usually possesses numerous local extrema, so that the least-squares procedure will converge to the correct solution only if the model is sufficiently close to the target [that is, only if equation (4) approximately holds], otherwise the procedure will be trapped in a local minimum.

The discussion of equation (5) requires the preliminary description of the basic relations used in the crystallographic Gauss–Newton approaches. For this purpose we introduce into equation (5) a scale factor g_{1t} between observed and calculated amplitudes, and the weight $w_{\mathbf{h}}$, which is expected to be inversely proportional to the variance of the observation [say $w_{\mathbf{h}} \simeq 1/\sigma^2(|F|)$] but would, in practice, also take other types of error into account. The standard form of $I_{\rm s}$ will then be

$$I_{\rm s}(|F|) = \sum_{\mathbf{h}} w_{\mathbf{h}}(|F_{\mathbf{h}}| - g_{1t}|F_{p\mathbf{h}}|)^2 = \min, \qquad (6)$$

which is the function minimized by canonical least-squares approaches (Busing *et al.*, 1962; Rollett *et al.*, 1976; Prince, 1994; Watkin, 2008).

We now explicitly recall the basics of the crystallographic Gauss–Newton least-squares procedures because they will be useful when modified forms for (6) are suggested. The minimization of $I_{\rm s}(|F|)$ may be achieved by requiring that

$$\sum_{\mathbf{h}} w_{\mathbf{h}}(|F_{\mathbf{h}}| - g_{1t}|F_{p\mathbf{h}}|)\frac{\delta|F_{p\mathbf{h}}|}{\delta x_j} = 0, \quad \text{for } j = 1, \ldots, \mu, \qquad (7)$$

where μ is the number of structural parameters on which $F_{p\mathbf{h}}$ depends. If $|F_{p\mathbf{h}}|$ is expanded about the current model value $|F_{p\mathbf{h}}|_{\rm M}$ as a function of the μ parameters x_k, say

$$|F_{p\mathbf{h}}| = |F_{p\mathbf{h}}|_{\rm M} + \sum_k \frac{\delta|F_{p\mathbf{h}}|}{\delta x_k}\Delta x_k, \qquad (8)$$

and if equation (8) is introduced into equation (7), we get

$$\sum_{\mathbf{h}} w_{\mathbf{h}}\left[|F_{\mathbf{h}}| - g_{1t}|F_{p\mathbf{h}}|_{\rm M} - g_{1t}\sum_k \frac{\delta|F_{p\mathbf{h}}|}{\delta x_k}\Delta x_k\right] \times \frac{\delta|F_{p\mathbf{h}}|}{\delta x_j} = 0, \quad \text{for } j = 1, \ldots, \mu, \qquad (9)$$

from which the canonical normal equations are obtained. In matrix form

$$\mathbf{B}\Delta\mathbf{X} = \mathbf{D}, \qquad (10)$$

where

$$\mathbf{B} = \left\{g_{1t}\sum_{\mathbf{h}} w_{\mathbf{h}}\sum_k \frac{\delta|F_{p\mathbf{h}}|}{\delta x_j}\frac{\delta|F_{p\mathbf{h}}|}{\delta x_k}\right\}, \qquad (11)$$

$$\mathbf{D} = \left\{ \sum_{\mathbf{h}} w_{\mathbf{h}} [|F_{\mathbf{h}}|_{\mathrm{M}} - g_{1t}|F_{p\mathbf{h}}|_{\mathrm{M}}] \frac{\delta |F_{p\mathbf{h}}|}{\delta x_j} \right\} \qquad (12)$$

and $\Delta\mathbf{X}$ is the parameter shift vector.

We will see in the next sections how equations (10)–(12) may be modified. We only recall here that the derivative of $|F_{p\mathbf{h}}|$ also depends on the phase $\varphi_{p\mathbf{h}}$.

It is well known that crystallographic structure refinement is also performed *via*

$$I_{\mathrm{s}}(|F|^2) = \sum_{\mathbf{h}} w_{\mathbf{h}}(|F_{\mathbf{h}}|^2 - g_{2t}|F_{p\mathbf{h}}|^2)^2 = \min . \qquad (13)$$

The derivation of least-squares procedures from equation (13) and their possible modifications are described in Appendix *A*.

Throughout this paper, the concept of missing structure (that is, the part of the target structure not contained in the model) will be discussed. It has been recently reconsidered by Blanc *et al.* (2004), in combination with maximum-likelihood techniques, for improving the efficiency of protein structure refinement. The missing atoms were assumed to be statistically distributed, and the calculated overall electron density was assumed to be the sum of three components:

$$\rho_{\mathrm{ov}}(\mathbf{r}) = \rho_p(\mathbf{r}) + \rho_u(\mathbf{r}) + \rho_{\mathrm{solv}}(\mathbf{r})$$

representing, in order, the atomic positions of the located fragment, the missing atom model and the bulk solvent. Correspondingly, the calculated structure factor was assumed to be the sum of three components:

$$F_{\mathrm{ov}\mathbf{h}} = F_{p\mathbf{h}} + F_{u\mathbf{h}} + F_{\mathrm{solv}\mathbf{h}}.$$

It was shown that, when the model is very incomplete, the bias affecting model refinement is reduced.

In this paper, mostly directed at small- and medium-sized structures, the solvent contribution will be neglected and more simple hypotheses will be made.

3. About the classical scale factor in least-squares procedures

Usually the $|F_{p\mathbf{h}}|$'s are on the absolute scale while the $|F_{\mathbf{h}}|$'s are on a relative scale. Minimizing the function $\sum_{\mathbf{h}} w_{\mathbf{h}}(g_{1t}|F_{\mathbf{h}}| - |F_{p\mathbf{h}}|)^2$ is, however, not used because the minimum could be found at $g_{1t} = 0$ and at extremely large atomic vibrational motion. During the various least-squares refinement cycles it is the structural model which must be refined and not *vice versa*. This is the reason why g_{1t} is associated to $|F_{p\mathbf{h}}|$ rather than to $|F_{\mathbf{h}}|$, exactly as reported in equation (6). From this equation g_{1t} is usually estimated *via* the relationship

$$g_{1t} = \frac{\sum_{\mathbf{h}} w_{\mathbf{h}} |F_{p\mathbf{h}}||F_{\mathbf{h}}|}{\sum_{\mathbf{h}} w_{\mathbf{h}} |F_{p\mathbf{h}}|^2}; \qquad (14)$$

sometimes the simpler relation

$$g_{1t} = \frac{\sum_{\mathbf{h}} |F_{\mathbf{h}}|}{\sum_{\mathbf{h}} |F_{p\mathbf{h}}|} \qquad (15)$$

is used. The $|F_{\mathbf{h}}|$'s, rescaled by the factor $1/g_{1t}$, may work as observations in subsequent cycles of least squares.

A question arises: is it really advisable to put the calculated structure-factor amplitudes on the scale of the observed ones? Such a practice is equivalent, in direct space, to making the total electronic charge of the model equal to the total electronic charge of the target, even if the scattering powers of the model and of the target densities are quite different. Or might it be better in this situation if the two sets of amplitudes remain on different scales? For example, if (Ne) and $(\mathrm{Ne})_{\mathrm{M}}$ denote the number of electrons belonging to the target and to the model structure, respectively, might it be better if the two scales are related by the factor $(\mathrm{Ne})/(\mathrm{Ne})_{\mathrm{M}}$?

The question is not negligible. Indeed, if we consider equation (11), we see that the use of g only scales the amplitudes of the elements of the matrix **B**, but, if we consider equation (12), we see that the scale factor deeply modifies (in magnitude and sign) the elements of the matrix **D**. Changing the scales changes the elements of **D**, and therefore signs and magnitudes of the parameter shifts may change too, so modifying the efficiency of the least-squares procedure.

A valid reason for putting observed and calculated amplitudes on the same scale is the following. If the scattering power of the model is smaller than that of the target, but the observed and the calculated amplitudes are obliged to stay on their respective absolute scales, the minimization of I_{s}, as defined by equation (6), may be obtained by reducing the vibrational motion of the model. *Vice versa*, if the scattering power of the model is larger than that of the target, the minimization may be obtained by increasing the vibrational motion of the model. In both cases, undesired systematic errors would be obtained.

On the other hand, using a scale factor which puts the observed and the calculated amplitudes on the same scale is numerically equivalent to pouring the scattering contribution of the atoms not included in the model but present in the target, into the scattering power of the model atoms. Indeed, $|F_{p\mathbf{h}}|$ is replaced by

$$g_{1t}|F_{p\mathbf{h}}| = |F_{p\mathbf{h}}| + (g_{1t} - 1)|F_{p\mathbf{h}}|. \qquad (16)$$

Using $g_{1t}|F_{p\mathbf{h}}|$ in equation (6) or in any discrepancy factor between observed and scaled calculated amplitudes is equivalent to accepting the following assumptions:

(i) In the case where the set of target atoms includes the model atoms ($N > N_p$), then it is automatically implied that the target atoms not belonging to the model provide (to the overall calculated structure factor) the contribution $(g_{1t} - 1)|F_{p\mathbf{h}}|\exp(i\varphi_{p\mathbf{h}})$. That is, for any reflections (*hkl*), the structure factors of the located and of the unlocated fragments have the same phase and proportional amplitude.

(ii) In the case where the set of model atoms includes the target atoms ($N < N_p$), then the corresponding calculated amplitudes are equally scaled to fit the target amplitudes.

In none of the two cases an ideal assumption is made.

4. A modified scaling in least-squares procedures

Let us suppose that the $|F_{\mathbf{h}}|$ amplitudes have been set on the absolute scale by some statistical technique (*e.g.*, by the Wilson method) and that the scattering power of the model is smaller than that of the target structure. Let us consider under the above conditions a logic paradigm, condensed in a few items:

(i) The prior knowledge of the chemical content of the target unit cell is usually available. It is a valuable piece of information because it allows us to estimate the total scattering power $\sum_N$ at any $s = \sin^2\theta/\lambda^2$.

(ii) If a model structure (from now on denoted as *located model*), constituted by $N_p < N$ atomic positions with their isotropic vibrational parameters, is available, the prior information in (i) allows us to estimate the lack of the model scattering power $(\sum_{N_p} - \sum_N)$ with respect to that of the target structure.

(iii) In the absence of any information on the positions of the missing atoms, the located model may be integrated by an *unlocated model component* to which the structure factor $F_{u\mathbf{h}}$ may be associated. $F_{u\mathbf{h}}$ has to be calculated *via* statistical approaches, given the ignorance of the missing atomic positions (see below). In simple terms, if the located model is constituted by $N_p < N$ atomic positions, the scattering power $\sum_N - \sum_{N_p}$ of the $N - N_p$ atoms with undefined positions is added to that of the located model.

(iv) Once $F_{u\mathbf{h}}$ has been determined for each $\mathbf{h}$, the least-squares procedure aiming at fixing the best values of the structural parameters may be started. But now the calculated structure factor is no longer given by $F_{p\mathbf{h}}$ but by

$$\begin{aligned}F_{\mathrm{ov}\mathbf{h}} &= (F_{p\mathbf{h}} + F_{u\mathbf{h}}) = |F_{\mathrm{ov}\mathbf{h}}|\exp(i\varphi_{\mathrm{ov}\mathbf{h}})\\ &= |F_{p\mathbf{h}}|\exp(i\varphi_{p\mathbf{h}}) + |F_{u\mathbf{h}}|\exp(i\varphi_{u\mathbf{h}}).\end{aligned}$$

When $N_p < N$, in the absence of any prior information, we assume $F_{u\mathbf{h}} = |F_{u\mathbf{h}}|\exp(i\varphi_{p\mathbf{h}})$ so that $|F_{\mathrm{ov}\mathbf{h}}| = |F_{p\mathbf{h}}| + |F_{u\mathbf{h}}|$ and $\varphi_{\mathrm{ov}\mathbf{h}} = \varphi_{p\mathbf{h}}$. That is, the phase of the statistical structure factor of the unlocated fragment is assumed to be that of the located fragment, in full analogy with one of the implicit choices in the usual scaling procedure [see equation (16)].

The value of $|F_{u\mathbf{h}}|$ which brings the overall scattering amplitudes of the integrated model on the observed (in our hypothesis, absolute) amplitude scale may be refined by least squares. For example, the observed reflections are subdivided in resolution shells and, for each shell, $|F_{u\mathbf{h}}|$ is found by minimizing

$$I_{1n}(|F|) = \sum_{\mathbf{h}} w_{\mathbf{h}}[|F_{\mathbf{h}}| - |F_{p\mathbf{h}}| - |F_{u\mathbf{h}}|]^2. \qquad (17)$$

We then obtain

$$|F_{u\mathbf{h}}| = \frac{\sum_{\mathbf{h}} w_{\mathbf{h}}[|F_{\mathbf{h}}| - |F_{p\mathbf{h}}|]}{\sum_{\mathbf{h}} w_{\mathbf{h}}} = \langle|F_{\mathbf{h}}|\rangle - \langle|F_{p\mathbf{h}}|\rangle, \qquad (18)$$

where the averages are calculated over the resolution shell to which $\mathbf{h}$ belongs.

A more general way of evaluating $|F_{u\mathbf{h}}|$, even applicable when the observed amplitudes are on an arbitrary scale, implies replacing $\langle|F_{\mathbf{h}}|\rangle - \langle|F_{p\mathbf{h}}|\rangle$ by a statistical estimate, such as that provided by Wilson statistics. Then

$$|F_{u\mathbf{h}}| \simeq q\left\{\left[\sum_N (s_{\mathbf{h}})\right] - \left[\sum_p (s_{\mathbf{h}})\right]\right\}^{1/2}, \qquad (19)$$

where $q = (2/\pi)^{1/2}$ for centric space groups and $q = (\pi^{1/2}/2)$ for acentric space groups. Consequently,

$$g_{1n} = \frac{\sum_{\mathbf{h}} w_{\mathbf{h}}|F_{\mathrm{ov}\mathbf{h}}||F_{\mathbf{h}}|}{\sum_{\mathbf{h}} w_{\mathbf{h}}|F_{\mathrm{ov}\mathbf{h}}|^2} \qquad (20)$$

or, more simply,

$$g_{1n} = \frac{\sum_{\mathbf{h}} |F_{\mathbf{h}}|}{\sum_{\mathbf{h}} |F_{\mathrm{ov}\mathbf{h}}|} \qquad (21)$$

may be used for scaling. The potential advantage of introducing the concept of unlocated fragment is the following. Usual scaling is equivalent to assuming that the target atoms not included in the model provide a structure factor with phase $\varphi_{u\mathbf{h}} = \varphi_{p\mathbf{h}}$ and amplitude proportional to the amplitude of the model atoms [say, equal to $(g_{1t} - 1)|F_{p\mathbf{h}}|$]. According to the second assumption, amplitudes calculated as large from the located model are even more emphasized by the scaling factor, while amplitudes calculated as small will remain small.

In our scaling approach the first assumption (say, $\varphi_{u\mathbf{h}} = \varphi_{p\mathbf{h}}$) is maintained, because we do not have any better estimate. On the contrary, the second assumption (say, the structure-factor amplitude of the unlocated fragment proportional to that of the located fragment) is not supported. Indeed, this second assumption is against any sensible statistical expectation, because the atoms of the unlocated fragment are in unknown positions: this lack of information allows us to estimate their contribution *via* the more sound Wilson statistics.

Using equation (19) as the scattering amplitude of the unlocated fragment implies that the contribution of the atoms present in the target and not included in the model is assumed to be proportional (in the absence of any information on their positions) to their expected average amplitude. As a practical consequence, if the scattering power of the model fragment is negligible with respect to that of the target, the contribution of the unlocated atoms is dominant, and the square amplitudes corresponding to the integrated model will be close to $\sum_N$. $|F_{p\mathbf{h}}|$ will progressively become dominant for increasing values of the scattering power of the located fragment. Accordingly, when $\sum_N = \sum_{N_p}$, $|F_{p\mathbf{h}}|$ will represent the total structure factor and $F_{u\mathbf{h}}$ will vanish. New and traditional scalings will converge in the last steps of the crystal structure refinement.

We have now to adapt the standard least-squares equations to the new paradigm, under the hypothesis that we have already estimated $F_{u\mathbf{h}}$ for each resolution shell. Then

$$\mathbf{B} = \left\{g_{1n}\sum_{\mathbf{h}} w_{\mathbf{h}} \sum_k \frac{\delta(|F_{\mathrm{ov}\mathbf{h}}|)}{\delta x_j}\frac{\delta(|F_{\mathrm{ov}\mathbf{h}}|)}{\delta x_k}\right\}, \qquad (22)$$

$$\mathbf{D} = \sum_{\mathbf{h}} w_{\mathbf{h}}[|F_{\mathbf{h}}|_{\mathrm{M}} - g_{1n}|F_{\mathrm{ov}\mathbf{h}}|_{\mathrm{M}}]\frac{\delta|F_{\mathrm{ov}\mathbf{h}}|}{\delta x_j}. \qquad (23)$$

The reader should notice that the only non-vanishing derivative of $|F_{u\mathbf{h}}|$ is that with respect to the Wilson overall B factor.

It is not rare that the size of the model exceeds that of the target structure (*e.g.*, in *ab initio* phasing procedures, when the phases are still far from the correct values). In this case $\sum_N - \sum_{N_p}$ is expected to be negative for all the resolution shells, and it may be assumed that

$$\begin{aligned}|F_{\mathrm{ov}\mathbf{h}}|\exp(i\varphi_{\mathrm{ov}\mathbf{h}}) &= |F_{p\mathbf{h}}|\exp(i\varphi_{p\mathbf{h}}) + |F_{u\mathbf{h}}|\exp[i(\varphi_{p\mathbf{h}}+\pi)]\\ &= (|F_{p\mathbf{h}}| - |F_{u\mathbf{h}}|)\exp(i\varphi_{p\mathbf{h}}).\end{aligned}$$

$|F_{u\mathbf{h}}|$ is found by minimizing

$$I_{1n}(|F|) = \sum_{\mathbf{h}} w_{\mathbf{h}}[|F_{\mathbf{h}}| - |F_{p\mathbf{h}}| + |F_{u\mathbf{h}}|]^2,$$

which leads to

$$|F_{u\mathbf{h}}| = \langle|F_{p\mathbf{h}}|\rangle - \langle|F_{\mathbf{h}}|\rangle.$$

Accordingly, if the $|F_{p\mathbf{h}}|$'s are on the absolute scale, then

$$F_{\mathrm{ov}\mathbf{h}} = |F_{p\mathbf{h}}|\exp(i\varphi_{p\mathbf{h}}) + q\left(\sum_{N_p} - \sum_{N}\right)^{1/2}\exp i(\varphi_{p\mathbf{h}}+\pi), \quad (24)$$

which is equivalent to the following relations:

$$F_{\mathrm{ov}\mathbf{h}} = \left[|F_{p\mathbf{h}}| - q\left(\sum_{N_p} - \sum_{N}\right)^{1/2}\right]\exp i(\varphi_{p\mathbf{h}})$$

$$\text{if } |F_{p\mathbf{h}}| > q\left(\sum_{N_p} - \sum_{N}\right)^{1/2},$$

otherwise

$$F_{\mathrm{ov}\mathbf{h}} = \left[q\left(\sum_{N_p} - \sum_{N}\right)^{1/2} - |F_{p\mathbf{h}}|\right]\exp i(\varphi_{p\mathbf{h}}+\pi).$$

In the first case $F_{\mathrm{ov}\mathbf{h}}$ and $F_{p\mathbf{h}}$ share the same phase $\varphi_{p\mathbf{h}}$, in the second case they have opposite phases: that may be important when $F_{\mathrm{ov}\mathbf{h}}$ is involved in least squares because the derivative of $|F_{\mathrm{ov}\mathbf{h}}|$ depends on the phase of $F_{\mathrm{ov}\mathbf{h}}$ (indeed weak reflections for which $F_{\mathrm{ov}\mathbf{h}}$ and $F_{p\mathbf{h}}$ have opposite phases would have a larger leverage). In both the cases

$$|F_{\mathrm{ov}\mathbf{h}}| = \left||F_{p\mathbf{h}}| - q\left(\sum_{N_p} - \sum_{N}\right)^{1/2}\right|. \quad (25)$$

5. About the crystallographic discrepancy factors

We analyse here the main properties of some statistical indices used in crystallography to estimate the fitting between two known structures. Let us still call the first structure target and the second model, even if they may be completely uncorrelated, and let us suppose that both $|F_{\mathbf{h}}|$ and $|F_{p\mathbf{h}}|$ are on their absolute scales. At a given step of the phasing process the misfit between model and target structure may be estimated *via* the criterion

$$\begin{aligned}R_{\mathrm{phas}} &= \frac{\sum_{\mathbf{h}}|F_{\mathrm{d}\mathbf{h}}|^2}{\sum_{\mathbf{h}}|F_{\mathbf{h}}|^2}\\ &= \frac{\sum_{\mathbf{h}}[|F_{\mathbf{h}}|^2 + |F_{p\mathbf{h}}|^2 - 2|F_{\mathbf{h}}||F_{p\mathbf{h}}|\cos(\varphi_{\mathbf{h}} - \varphi_{p\mathbf{h}})]}{\sum_{\mathbf{h}}|F_{\mathbf{h}}|^2}. \quad (26)\end{aligned}$$

In direct space R_{phas} corresponds to

$$\frac{\int_V[\rho(\mathbf{r}) - \rho_p(\mathbf{r})]^2\,\mathrm{d}\mathbf{r}}{\int_V\rho^2(\mathbf{r})\,\mathrm{d}\mathbf{r}},$$

which is an optimum criterion to measure the similarity of two structures. R_{phas} vanishes when the model coincides with the target. At the opposite extreme, when the model is uncorrelated with the target, it becomes

$$1 + \frac{\sum_{\mathbf{h}}|F_{p\mathbf{h}}|^2}{\sum_{\mathbf{h}}|F_{\mathbf{h}}|^2}, \quad (27)$$

which is expected to be close to 1 when the scattering power of the model is a small fraction of the target, close to 2 if model and target have the same scattering power. R_{phas} shows some interesting properties:

(i) According to equation (26), R_{phas} is sensitive both to the completeness of the model and to its quality. This property is shared by the σ_A parameter (Srinivasan & Ramachandran, 1965), which represents the correlation between the normalized square amplitudes of the model and of the target structures (Carrozzini, Cascarano, Giacovazzo & Mazzone, 2013).

(ii) The criterion

$$R(|F|^2) = \sum_{\mathbf{h}}[|F_{\mathbf{h}}| - |F_{p\mathbf{h}}|]^2\Big/\left(\sum_{\mathbf{h}}|F_{\mathbf{h}}|^2\right) \quad (28)$$

is a special case of R_{phas}, obtained when equation (4) is satisfied. If the model is poor, the approximation (4) does not hold and equation (28) becomes biased with respect to equation (26). In other words, only for high-quality models [that is, when the approximation (4) is satisfied] is $R(|F|^2)$ a good approximation of R_{phas}, but it numerically diverges for poor models.

(iii) The misfit between $\rho(\mathbf{r})$ and $\rho_p(\mathbf{r})$, once Fourier transformed, implies the misfit between the vectors $F_{\mathbf{h}}$ and $F_{p\mathbf{h}}$. Taking only their moduli into consideration, like equation (28) does, is an approximation not always acceptable when the target structure is known. Indeed it may occur (more frequently than we think today) that, because of some pseudo-translational symmetry, two small values of equation (28) or similar discrepancy indices may be calculated for two models, one of which is definitively wrong, even if the corresponding phases are remarkably different [see Cascarano *et al.* (2013) for an example].

(iv) If $\rho(\mathbf{r})$ and $\rho_p(\mathbf{r})$ are referred by an allowed origin shift, R_{phas} will reveal the misfit of the two maps, while discrepancy indices based only on amplitudes are completely insensitive to it. According to the purpose of the crystallographer, sometimes the insensitivity is preferred, sometimes it should be avoided.

Let us now consider the case in which the target structure is unknown but a model is available. Then a very popular criterion is

$$R_t(|F|) = \sum_{\mathbf{h}} [||F_{\mathbf{h}}| - g_{1t}|F_{p\mathbf{h}}||] / \left(\sum_{\mathbf{h}} |F_{\mathbf{h}}|\right), \qquad (29)$$

where the scale factor g_{1t}, calculated *via* equations (16) or (17), sets the calculated amplitudes on the scale of the observed amplitudes. $R_t(|F|)$ usually underestimates the misfit, because the phase misfit is neglected; furthermore, the undesirable statistical effects described in §4 occur.

If the unlocated fragment is involved in the calculations, then the discrepancy factor may be written down as

$$R_n(|F|) = \left\{\sum_{\mathbf{h}} |[|F_{\mathbf{h}}| - g_{1n}|F_{\mathrm{ov}\mathbf{h}}|]|\right\} / \left(\sum_{\mathbf{h}} |F_{\mathbf{h}}|\right) \qquad (30)$$

with

$$g_{1n} = \frac{\sum_{\mathbf{h}} |F_{\mathbf{h}}|}{\sum_{\mathbf{h}} |F_{\mathrm{ov}\mathbf{h}}|}$$

and

$$|F_{\mathrm{ov}\mathbf{h}}| = ||F_{p\mathbf{h}}| \pm |F_{u\mathbf{h}}||.$$

The plus sign occurs when $\sum_{N_p} < \sum_N$ and the minus sign when $\sum_{N_p} > \sum_N$.

Frequently the misfit between normalized structure-factor amplitudes $|E_{\mathbf{h}}|$ (where $E_{\mathbf{h}} = F_{\mathbf{h}} / \sum_N^{1/2}$) is calculated as

$$R_t(|E|) = \sum_{\mathbf{h}} [||E_{\mathbf{h}}| - |E_{p\mathbf{h}}||] / \left(\sum_{\mathbf{h}} |E_{\mathbf{h}}|\right).$$

Both $|E_{\mathbf{h}}|$ and $|E_{p\mathbf{h}}|$ are by definition on their absolute scales (indeed $\langle |E_{\mathbf{h}}|^2 \rangle = \langle |E_{p\mathbf{h}}|^2 \rangle = 1$). $R_t(|E|)$ may also be rewritten as

$$R_t(|E|) = \frac{\sum_{\mathbf{h}} ||F_{\mathbf{h}}| - |F_{p\mathbf{h}}|(\sum_N / \sum_{N_p})^{1/2}|}{(\sum_{\mathbf{h}} |F_{\mathbf{h}}|)}. \qquad (31)$$

Equation (31) suggests that $R_t(|E|)$ is equivalent to $R_t(|F|)$ as given by equation (29) provided $|F_{\mathbf{h}}|$ and $|F_{p\mathbf{h}}|$ are on the same scale. The use of the unlocated fragment modifies $R_t(|E|)$ into

$$R_n(|E|) = \sum_{\mathbf{h}} [||E_{\mathbf{h}}| - |E_{\mathrm{ov}\mathbf{h}}||] / \left(\sum_{\mathbf{h}} |E_{\mathbf{h}}|\right).$$

Let us now discuss the effects of the unlocated fragment on the Pearson correlation coefficient:

$$\mathrm{CORR} = \frac{\sum_{\mathbf{h}} |(|F_{\mathbf{h}}| - \langle |F_{\mathbf{h}}| \rangle)(|F_{p\mathbf{h}}| - \langle |F_{p\mathbf{h}}| \rangle)|}{\left\{[\sum_{\mathbf{h}} (|F_{\mathbf{h}}| - \langle |F_{\mathbf{h}}| \rangle)^2][\sum_{\mathbf{h}} (|F_{p\mathbf{h}}| - \langle |F_{p\mathbf{h}}| \rangle)^2]\right\}^{1/2}}.$$

Since CORR is expressed in terms of uncentred moments, it is invariant under scaling and is equal to the value assumed when both model and target amplitudes are on the same scale. CORR is therefore intrinsically insensitive to the incompleteness of the model.

6. An additional statistical parameter in least-squares procedures

We have emphasized in §2 that the weight $w_{\mathbf{h}}$, a necessary ingredient for obtaining an unbiased least-squares estimate of the structural parameters, is expected to be inversely proportional to the variance of the observed amplitude (in practical applications additional criteria are used, but their nature is not of interest for this paper). In the same §2 we also reported, in equation (2), the explicit expression of the integral I, which, under the condition (4) and suitably minimized, leads to the standard crystallographic least-squares formula (6). The necessary intermediate step is the use of the relation (4), the validity of which varies with the quality of the model. If we consider equation (2) as the source of least squares, we should accept the idea that least squares may benefit by a supplementary parameter, say $m_{\mathbf{h}}$, to be coupled with the weight $w_{\mathbf{h}}$, and taking into account the reliability of the relation (4). The problem may be solved by rewriting equation (2) in the form

$$I = \frac{1}{V} \sum_{\mathbf{h}} ||F_{\mathbf{h}}| \exp(i\varphi_{\mathbf{h}}) - |F_{p\mathbf{h}}| \exp(i\varphi_{p\mathbf{h}})|^2 \qquad (32)$$

and by replacing the unknown phase factor $\exp(i\varphi_{\mathbf{h}}) = \cos\varphi_{\mathbf{h}} + i\sin\varphi_{\mathbf{h}}$ by its expected value (see Sim, 1959; Srinivasan & Ramachandran, 1965; Read, 1986; Carrozzini, Cascarano, Giacovazzo & Mazzone, 2013): *i.e.*,

$$\langle \exp(i\varphi_{\mathbf{h}}) \rangle = \langle \cos\varphi_{\mathbf{h}} \rangle \exp(i\varphi_{p\mathbf{h}}) = m_{\mathbf{h}} \exp(i\varphi_{p\mathbf{h}}). \qquad (33)$$

In this case the standard least-squares equation (6) is modified into

$$I_{\mathrm{s}} = \sum_{\mathbf{h}} w_{\mathbf{h}} (m_{\mathbf{h}} |F_{\mathbf{h}}| - g|F_{p\mathbf{h}}|)^2 = \min. \qquad (34)$$

In its turn, equation (34) is related to the standard least-squares equation (7) as follows: both derive from equation (3), but in the first case we used the constraint $\varphi_{\mathbf{h}} \simeq \varphi_{p\mathbf{h}}$, while in the second we applied the condition $\cos(\varphi_{\mathbf{h}} - \varphi_{p\mathbf{h}}) \simeq m_{\mathbf{h}}$ [or, equivalently, we imposed the condition (33)]. This second constraint is less severe than the first.

The relation (34) has however a handicap: $m_{\mathbf{h}}$, and therefore $m_{\mathbf{h}}|F_{\mathbf{h}}|$, is continuously refined during the least-squares cycles, against the canonical rule according to which the observations should remain unmodified during the refinement. If equation (34) is applied, a minimum may be attained at $g = 0$, even with a model uncorrelated with the target: then $m_{\mathbf{h}} \simeq 0$ for most of the reflections, and extremely large thermal motion might arise from least-squares cycles. The handicap may be overcome by using the relation

$$\langle \exp(i\varphi_{p\mathbf{h}}) \rangle = m_{\mathbf{h}} \exp(i\varphi_{\mathbf{h}}) \qquad (35)$$

instead of equation (33): now $i\varphi_{\mathbf{h}}$ is considered fixed, even if unknown, while $\varphi_{p\mathbf{h}}$ is distributed around it.

Equation (35) is symmetric to equation (33) but is equally legitimate. Indeed, if the unknown $\varphi_{\mathbf{h}}$ value is expected to be close to the known $\varphi_{p\mathbf{h}}$ value, in an equivalent way we can state that, according to equation (35), $\varphi_{p\mathbf{h}}$ is expected to be close to

$\varphi_{\mathbf{h}}$ no matter if $\varphi_{\mathbf{h}}$ is unknown. Then the standard least-squares equation becomes

$$\sum_{\mathbf{h}} w_{\mathbf{h}}(|F_{\mathbf{h}}| - g_{1t}m_{\mathbf{h}}|F_{p\mathbf{h}}|)^2 = \min. \tag{36}$$

The use of equation (36) leads again to equation (10) but now

$$\mathbf{B} = \left\{ g_{1t} \sum_{\mathbf{h}} w_{\mathbf{h}} m_{\mathbf{h}} \sum_{k} \frac{\delta|F_{p\mathbf{h}}|}{\delta x_j} \frac{\delta|F_{p\mathbf{h}}|}{\delta x_k} \right\}, \tag{37}$$

$$\mathbf{D} = \left\{ \sum_{\mathbf{h}} w_{\mathbf{h}}(|F_{\mathbf{h}}| - g_{1t}m_{\mathbf{h}}|F_{p\mathbf{h}}|_{\mathrm{M}}) \frac{\delta|F_{p\mathbf{h}}|}{\delta x_j} \right\} \tag{38}$$

and

$$g_{1t} = \frac{\sum_{\mathbf{h}} w_{\mathbf{h}} m_{\mathbf{h}} |F_{p\mathbf{h}}||F_{\mathbf{h}}|}{\sum_{\mathbf{h}} w_{\mathbf{h}} m_{\mathbf{h}}^2 |F_{p\mathbf{h}}|^2}$$

or

$$g_{1t} = \frac{\sum_{\mathbf{h}} |F_{\mathbf{h}}|}{\sum_{\mathbf{h}} m_{\mathbf{h}} |F_{p\mathbf{h}}|}.$$

For high-quality models, equations (37) and (38) are equivalent to equations (11) and (12), respectively (where $m = 1$ for each structure factor). But, if the model is weakly correlated with the target, the parameter $m_{\mathbf{h}}$ may vary strongly from one reflection to another, and, for bad models, may frequently fall close to zero. As a consequence, the elements of the matrix **B** and of the vector **D**, respectively, are expected to be very different from the corresponding elements in equations (11) and (12). Indeed, the introduction of the parameter $m_{\mathbf{h}}$ may change not only the moduli but also the signs of the elements of the **D** vector (so changing the directions of some structural parameter shifts), and may also strongly change the scale factor.

From the above observations it is evident that the use of the statistical parameter m does not aim at obtaining lower values of the crystallographic discrepancy indices, but mainly aims at increasing the least-square convergence when poor models are available.

The use of equations (37) and (38) has the same handicaps described for equation (6) in §3, probably made more critical by the use of the reliability parameter $m_{\mathbf{h}}$. It may therefore be useful to introduce in the least-squares treatment the scattering of the unlocated model, as described in §4. The function to minimize is again

$$\sum_{\mathbf{h}} w_{\mathbf{h}}[|F_{\mathbf{h}}| - g_{1n}|F_{\mathrm{ov}\mathbf{h}}|]^2.$$

If the model scattering power is smaller than that of the target, then

$$\begin{aligned} |F_{\mathrm{ov}\mathbf{h}}| \exp(i\varphi_{\mathrm{ov}\mathbf{h}}) &= |F_{p\mathbf{h}}| \exp(i\varphi_{p\mathbf{h}}) + |F_{u\mathbf{h}}| \exp(i\varphi_{p\mathbf{h}}) \\ &= m_{\mathbf{h}}(|F_{p\mathbf{h}}| + |F_{u\mathbf{h}}|) \exp(i\varphi_{\mathbf{h}}): \end{aligned}$$

that is, in the absence of prior information, the same reliability value $m_{\mathbf{h}}$ may be associated to the probability that both $\varphi_{p\mathbf{h}}$ and $\varphi_{u\mathbf{h}}$ are close to $\varphi_{\mathbf{h}}$. Then

$$|F_{\mathrm{ov}\mathbf{h}}| = m_{\mathbf{h}}(|F_{p\mathbf{h}}| + |F_{u\mathbf{h}}|), \ \varphi_{\mathrm{ov}\mathbf{h}} = \varphi_{\mathbf{h}}.$$

Furthermore

$$|F_{u\mathbf{h}}| = \langle|F_{\mathbf{h}}|\rangle - \langle|F_{p\mathbf{h}}|\rangle = q\left(\sum_{N} - \sum_{N_p}\right)^{1/2}.$$

If the model scattering power is larger than that of the target, then

$$\begin{aligned} |F_{\mathrm{ov}\mathbf{h}}| \exp(i\varphi_{\mathrm{ov}\mathbf{h}}) &= |F_{p\mathbf{h}}| \exp(i\varphi_{p\mathbf{h}}) + |F_{u\mathbf{h}}| \exp[i(\varphi_{p\mathbf{h}} + \pi)] \\ &= m_{\mathbf{h}}(|F_{p\mathbf{h}}| - |F_{u\mathbf{h}}|) \exp(i\varphi_{\mathbf{h}}). \end{aligned}$$

In the absence of prior information, the same reliability value $m_{\mathbf{h}}$ may be associated to the probability that $\varphi_{p\mathbf{h}}$ is close to $\varphi_{\mathbf{h}}$ and that $\varphi_{u\mathbf{h}}$ is close to $\varphi_{\mathbf{h}} + \pi$. Since

$$|F_{u\mathbf{h}}| = \langle|F_{p\mathbf{h}}|\rangle - \langle|F_{\mathbf{h}}|\rangle = q\left(\sum_{N_p} - \sum_{N}\right)^{1/2}$$

we have

$$F_{\mathrm{ov}\mathbf{h}} = m_{\mathbf{h}}\left[|F_{p\mathbf{h}}| \exp(i\varphi_{p\mathbf{h}}) + q\left(\sum_{N_p} - \sum_{N}\right)^{1/2} \exp i(\varphi_{p\mathbf{h}} + \pi)\right],$$

which is equivalent to the following relations:

$$F_{\mathrm{ov}\mathbf{h}} = m_{\mathbf{h}}\left[|F_{p\mathbf{h}}| - q\left(\sum_{N_p} - \sum_{N}\right)^{1/2}\right] \exp i(\varphi_{p\mathbf{h}})$$

$$\text{if } |F_{p\mathbf{h}}| > q\left(\sum_{N_p} - \sum_{N}\right)^{1/2},$$

otherwise

$$F_{\mathrm{ov}\mathbf{h}} = m_{\mathbf{h}}\left[q\left(\sum_{N_p} - \sum_{N}\right)^{1/2} - |F_{p\mathbf{h}}|\right] \exp i(\varphi_{p\mathbf{h}} + \pi).$$

It may be important to distinguish between the two alternatives owing to the fact that, when $F_{\mathrm{ov}\mathbf{h}}$ is involved in least squares, the derivative of $|F_{\mathrm{ov}\mathbf{h}}|$ depends on the phase of $\varphi_{\mathrm{ov}\mathbf{h}}$. In both cases

$$|F_{\mathrm{ov}\mathbf{h}}| = ||F_{p\mathbf{h}}| - |F_{u\mathbf{h}}|| = \left| |F_{p\mathbf{h}}| - q\left(\sum_{N_p} - \sum_{N}\right)^{1/2} \right|. \tag{39}$$

Then equations (10), (22) and (23) are again obtained, but now $m_{\mathbf{h}}$ has been incorporated into $F_{\mathrm{ov}\mathbf{h}}$. That changes the absolute values of the derivatives in equation (22) and also may change the modulus and sign of the elements of the **D** vector in equation (23).

7. Phase-driven least squares: the vector refinement

We noticed in §2 that minimizing equation (2) leads to equation (6) provided the condition $\varphi_{\mathbf{h}} = \varphi_{p\mathbf{h}}$ holds. If some previous information on $\varphi_{\mathbf{h}}$ is available, then the more general expression (3) may be used, which may also be rewritten in the form

$$I = \sum_{\mathbf{h}} ||F_{\mathbf{h}}| \exp(i\varphi_{\mathbf{h}}) - |F_{p\mathbf{h}}| \exp(i\varphi_{p\mathbf{h}})|^2 = \sum_{\mathbf{h}} |F_{q\mathbf{h}}|^2 = \min . \tag{40}$$

The expression (40) corresponds to the most complete application of equation (3). Indeed:

(i) It is equivalent to minimizing the $|F_{q\mathbf{h}}|$ amplitudes, and is referred to in the literature (Arnold & Rossmann, 1988) as *vector refinement*.

(ii) It doubles [with respect to equation (6)] the number of observational equations, because it minimizes functions of amplitudes and of phases. Indeed equation (40) may also be written as

$$\sum_{\mathbf{h}} w_{\mathbf{h}}[(A_{\mathbf{h}} - A_{p\mathbf{h}})^2 + (B_{\mathbf{h}} - B_{p\mathbf{h}})^2] = \min, \tag{41}$$

which directly leads, in accordance with Arnold & Rossmann, to the elements of the least-squares normal matrix

$$\sum_{\mathbf{h}} w_{\mathbf{h}} \left[\frac{\partial A_{p\mathbf{h}}}{\partial x_j} \frac{\partial A_{p\mathbf{h}}}{\partial x_k} + \frac{\partial B_{p\mathbf{h}}}{\partial x_j} \frac{\partial B_{p\mathbf{h}}}{\partial x_k} \right], \quad \text{for parameters } j, k,$$

and to the gradient vector elements

$$\sum_{\mathbf{h}} w_{\mathbf{h}} \left[\frac{\partial A_{p\mathbf{h}}}{\partial x_k} \Delta A_{\mathbf{h}} + \frac{\partial B_{p\mathbf{h}}}{\partial x_k} \Delta B_{\mathbf{h}} \right],$$

where

$$\Delta A_{\mathbf{h}} = A_{\mathbf{h}} - (A_{p\mathbf{h}})_{\rm M}, \qquad \Delta B_{\mathbf{h}} = B_{\mathbf{h}} - (B_{p\mathbf{h}})_{\rm M}.$$

(iii) It optimizes the fit of the molecular model to the electron density calculated by using observed amplitudes and $\varphi_{\mathbf{h}}$ phases.

It was suggested by Arnold & Rossmann that the pure vector refinement is appropriate when exceedingly good phase information is available, as may occur in virus crystallography where non-crystallographic symmetry may be employed to improve the molecular-replacement phases. The combination of vector and scalar refinement was suggested in less favourable cases like isomorphous replacement. More recently, restrained vector refinement has been implemented in *REFMAC* (Murshudov *et al.*, 1997) *via* maximum-likelihood techniques to refine protein model structures.

We describe here some potential applications of unconstrained and constrained least-squares vector refinement in small and large molecules, in situations that were not considered by previous authors.

Let us suppose that the model structure available at the end of a least-squares procedure is a very rough approximation of the target, and that additional least-squares cycles are unable to improve it. That frequently occurs when the model is severely incomplete and/or when there is a large misfit between the model and the target atomic coordinates. In this case, even if least squares do not converge to the target structure, other techniques less sensitive to model errors may reduce the phase bias: *e.g.*, EDM techniques may reduce by 5–20° the average phase error calculated for the best least-squares model. This point seems today to be of particular interest because EDM techniques have been recently powered by a new entry like the *VLD* (*Vive la Difference*) approach (Burla, Carrozzini *et al.*, 2011; Burla, Giacovazzo & Polidori, 2011; Burla *et al.*, 2012), originally designed for *ab initio* phasing but extremely efficient for improving phases in non-*ab initio* methods (Carrozzini, Cascarano, Comunale *et al.*, 2013).

The above considerations suggest a new opportunity, of interest when the model is weakly correlated with the target: to design an interaction between least squares and EDM techniques, deeper than in current procedures. Let us consider two examples.

In *ab initio* phasing techniques for small- and medium-sized molecules, at a certain step of the phasing approach, a molecular model may be obtained by a simple (even if guided by basic crystal chemical rules) peak search procedure applied to the current electron-density map. If the model is of low quality it will hardly converge to the target structure when submitted to least-squares cycles. However the best set of phases $\{\varphi_{p\mathbf{h}}\}$ (*i.e.*, that corresponding to the best least-squares model) may be used as a starting point for the application of EDM techniques. These may end with a set of phases (say $\{\varphi_{\mathbf{h}}\}$) remarkably better than the set $\{\varphi_{p\mathbf{h}}t\}$. That opens two alternatives:

(i) The electron density calculated *via* the set $\{\varphi_{\mathbf{h}}\}$ may be chemically interpreted and used to construct a new molecular model which may again be submitted to EDM procedures and after to standard least squares, and so on cyclically. That corresponds to the usual refinement procedure.

(ii) Least squares may be modified in such a way to profit not only by the amplitudes $\{|F_{p\mathbf{h}}|\}$ and $\{|F_{\mathbf{h}}|\}$, but also by the sets $\{\varphi_{p\mathbf{h}}\}$ and $\{\varphi_{\mathbf{h}}\}$. In this case, the modified least-squares procedure would lead to a model automatically fitting the best electron density corresponding to the phases $\{\varphi_{\mathbf{h}}\}$, without passing through the model-building step.

As a second and more interesting example, let us consider a molecular-replacement case, where a low-quality molecular model (*e.g.*, with a degree of similarity with the target of less than 30%) has been submitted to molecular-replacement techniques and has been correctly oriented and located. The standard methods for phasing the target structure very likely fail because of the model bias. The use of vector refinement like that implemented in *REFMAC* may be able to automatically fit the model to the new electron density without passing through the model-building step, which may fail if the electron density is still of poor quality.

Additional criteria like the introduction of the reliability parameter $m_{\mathbf{h}}$ into the vector-refinement procedure may be easily accomplished if the minimization of

$$\sum_{\mathbf{h}} w_{\mathbf{h}}[(A_{\mathbf{h}} - m_{\mathbf{h}}A_{p\mathbf{h}})^2 + (B_{\mathbf{h}} - m_{\mathbf{h}}B_{p\mathbf{h}})^2]$$

is performed.

8. Conclusions

This paper shows that crystallographic least-squares properties may be derived from direct-space functions, establishing the similarity between two electron-density maps. The approach is also able to suggest modified procedures potentially able to improve the least-squares efficiency. The concept of the unlocated fragment is introduced to analyse the scaling procedures more frequently used to scale diffraction amplitudes, the crystallographic discrepancy criteria are discussed and the vector-refinement method is obtained as a special case of the new paradigm. The approach is purely speculative: the applications are not part of this paper, which only provides the theoretical basis for future applications.

APPENDIX *A*

When the crystallographic structure refinement is performed *via* $|F|^2$ refinement, the discrepancy integral may be described by equation (13). Minimizing such $I_s(|F|^2)$ requires that

$$\sum_{\mathbf{h}} w_{\mathbf{h}}(|F_{\mathbf{h}}|^2 - g_{2t}|F_{p\mathbf{h}}|^2)\frac{\delta|F_{p\mathbf{h}}|^2}{\delta x_j} = 0, \text{ for } j = 1, \ldots, \mu.$$

If $|F_{p\mathbf{h}}|^2$ is expanded about the current model value $|F_{p\mathbf{h}}|^2_{\mathrm{M}}$ as a function of the μ parameters x_k, say

$$|F_{p\mathbf{h}}|^2 = |F_{p\mathbf{h}}|^2_{\mathrm{M}} + \sum_k \frac{\delta|F_{p\mathbf{h}}|^2}{\delta x_k}\Delta x_k,$$

and if the relation $\delta|F_{p\mathbf{h}}|^2/\delta x_k = 2|F_{p\mathbf{h}}|\delta|F_{p\mathbf{h}}|/\delta x_k$ is taken into account, then equation (10) is again obtained but

$$\mathbf{B} = \left\{2g_{2t}\sum_{\mathbf{h}} w_{\mathbf{h}}|F_{p\mathbf{h}}|^2_{\mathrm{M}}\sum_k \frac{\delta|F_{p\mathbf{h}}|}{\delta x_j}\frac{\delta|F_{p\mathbf{h}}|}{\delta x_k}\right\}, \tag{42}$$

$$\mathbf{D} = \left\{\sum_{\mathbf{h}} w_{\mathbf{h}}|F_{p\mathbf{h}}|_{\mathrm{M}}[|F_{\mathbf{h}}|^2_{\mathrm{M}} - g_{2t}|F_{p\mathbf{h}}|^2_{\mathrm{M}}]\frac{\delta|F_{p\mathbf{h}}|}{\delta x_j}\right\}. \tag{43}$$

Scaling between observed and calculated square amplitudes is obtained *via*

$$g_{2t} = \frac{\sum_{\mathbf{h}} w_{\mathbf{h}}|F_{p\mathbf{h}}|^2|F_{\mathbf{h}}|^2}{\sum_{\mathbf{h}} w_{\mathbf{h}}|F_{p\mathbf{h}}|^4} \tag{44}$$

or *via* the simpler relation

$$g_{2t} = \frac{\sum_{\mathbf{h}} |F_{\mathbf{h}}|^2}{\sum_{\mathbf{h}} |F_{p\mathbf{h}}|^2}. \tag{45}$$

Hold for such scaling the same considerations described in §3 for the standard scaling procedure in the case of $|F|$ refinement: that is, if the set of target atoms includes the model atoms, the structure factors of the located and of the unlocated fragments are assumed to have the same phase and proportional amplitude, an improper statistical behaviour. To overcome this problem the paradigm of the unlocated fragment may be introduced: *e.g.*, when $\sum_N > \sum_p$ then

$$|F_{u\mathbf{h}}|^2 = \langle|F_{\mathbf{h}}|^2\rangle - \langle|F_{p\mathbf{h}}|^2\rangle = \left[\sum_N(s_{\mathbf{h}}) - \sum_p(s_{\mathbf{h}})\right].$$

In this case $|F_{\mathrm{ov}\mathbf{h}}|^2 = |F_{p\mathbf{h}}|^2 + |F_{u\mathbf{h}}|^2$ and scaling may be performed *via*

$$g_{2n} = \frac{\sum_{\mathbf{h}} w_{\mathbf{h}}|F_{\mathbf{h}}|^2|F_{\mathrm{ov}\mathbf{h}}|^2}{\sum_{\mathbf{h}} w_{\mathbf{h}}|F_{\mathrm{ov}\mathbf{h}}|^4} \tag{46}$$

or, more simply, *via*

$$g_{2n} = \frac{\sum_{\mathbf{h}} |F_{\mathbf{h}}|^2}{\sum_{\mathbf{h}} |F_{\mathrm{ov}}|^2}. \tag{47}$$

The use of the unlocated fragment paradigm transforms equations (42) and (43) into

$$\mathbf{B} = \left\{2g_{2n}\sum_{\mathbf{h}} w_{\mathbf{h}}|F_{\mathrm{ov}\mathbf{h}}|^2_{\mathrm{M}}\sum_k \frac{\delta|F_{\mathrm{ov}\mathbf{h}}|}{\delta x_j}\frac{\delta|F_{\mathrm{ov}\mathbf{h}}|}{\delta x_k}\right\} \tag{48}$$

and

$$\mathbf{D} = \left\{\sum_{\mathbf{h}} w_{\mathbf{h}}|F_{\mathrm{ov}\mathbf{h}}|_{\mathrm{M}}[|F_{\mathbf{h}}|^2_{\mathrm{M}} - g_{2n}|F_{\mathrm{ov}\mathbf{h}}|^2_{\mathrm{M}}]\frac{\delta|F_{\mathrm{ov}\mathbf{h}}|}{\delta x_j}\right\}, \tag{49}$$

respectively.

Corresponding modifications should be applied to the usual discrepancy criterion between two structures, say to

$$R_t(|F|^2) = \sum_{\mathbf{h}} ||F_{\mathbf{h}}|^2 - g_{2t}|F_{p\mathbf{h}}|^2| / \left(\sum_{\mathbf{h}} |F_{\mathbf{h}}|^2\right), \tag{50}$$

where g_{2t} is calculated *via* equation (44) or equation (45). If the concept of unlocated model described in §4 is used, then $R_t(|F|^2)$ is replaced by

$$R_n(|F|^2) = \left\{\sum_{\mathbf{h}} [||F_{\mathbf{h}}|^2 - g_{2n}(|F_{p\mathbf{h}}|^2 \pm |F_{u\mathbf{h}}|^2)|]\right\} / \left(\sum_{\mathbf{h}} |F_{\mathbf{h}}|^2\right), \tag{51}$$

where g_{2n} is calculated *via* equations (46) or (47).

I heartily thank Giovanni Luca Cascarano for very useful discussions.

References

Arnold, E. & Rossmann, M. G. (1988). *Acta Cryst.* A**44**, 270–283.

Blanc, E., Roversi, P., Vonrhein, C., Flensburg, C., Lea, S. M. & Bricogne, G. (2004). *Acta Cryst.* D**60**, 2210–2221.

Bricogne, G. (1997). *Methods Enzymol.* **276**, 361–423.

Burla, M. C., Carrozzini, B., Cascarano, G. L., Giacovazzo, C. & Polidori, G. (2011). *J. Appl. Cryst.* **44**, 1143–1151.

Burla, M. C., Carrozzini, B., Cascarano, G. L., Giacovazzo, C. & Polidori, G. (2012). *J. Appl. Cryst.* **45**, 1287–1294.

Burla, M. C., Giacovazzo, C. & Polidori, G. (2011). *J. Appl. Cryst.* **44**, 193–199.

Busing, W. R., Martin, K. O. & Levy, H. A. (1962). ORFLS Report ORNL-TM-305. Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA.

Carrozzini, B., Cascarano, G. L., Comunale, G., Giacovazzo, C. & Mazzone, A. (2013). *Acta Cryst.* D**69**, 1038–1044.

Carrozzini, B., Cascarano, G. L., Giacovazzo, C. & Mazzone, A. (2013). *Acta Cryst.* A**69**, 408–412.

Cascarano, G. L., Ferguson, G., Giacovazzo, C., Glidewell, C. & Spek, A. L. (2013). *Acta Cryst.* C**69**, 774–777.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.

Pannu, N. S. & Read, R. J. (1996). *Acta Cryst.* A**52**, 659–668.

Prince, E. (1994). *Mathematical Techniques in Crystallography and Materials Science*, 2nd ed., ch. 9. Berlin: Springer-Verlag.

Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.

Rollett, J. S., McKinlay, T. G. & Haigh, P. N. (1976). *Crystallographic Computing Techniques*, edited by F. R. Ahmed, pp. 413–419. Copenhagen: Munksgaard.

Sim, G. A. (1959). *Acta Cryst.* **12**, 813–815.

Srinivasan, R. & Ramachandran, G. N. (1965). *Acta Cryst.* **19**, 1008–1014.

Watkin, D. (2008). *J. Appl. Cryst.* **41**, 491–522.