Acta Cryst. (1969). B**25**, 805

Filtering in the Method of Least-Squares

BY R. DIAMOND

Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge, England

(*Received* 5 *August* 1968)

A recently published paper by Scheringer gives an account of filtering theory as applied to the least-squares refinement of crystal structures. This account is critical of some of the author's earlier work, and in this note some of these criticisms are answered and an alternative point of view is expressed.

Any least-squares operation starts with the observational equations

$$\mathbf{y}=\mathbf{D}\mathbf{x}+\boldsymbol{\varepsilon}$$

in which the column matrix $\mathbf{y}$ contains the observational quantities, $y_{\text{obs}}-y_{\text{calc}}$ where y_{calc} is based on the starting values of the parameters, $\mathbf{D}$ contains the derivatives $\frac{\partial y_{\text{calc}}}{\partial x}$, $\mathbf{x}$ contains the parametric shifts and $\boldsymbol{\varepsilon}$ contains the residuals. If we aim to minimize the quadratic $\tilde{\boldsymbol{\varepsilon}}\tilde{\mathbf{W}}\mathbf{W}\boldsymbol{\varepsilon}$ in which $\tilde{\mathbf{W}}\mathbf{W}$ is the weighting matrix for the observations (normally the inverse of the variance–covariance matrix for the observational errors) we may do so through weighted parameters $\mathbf{z}=\mathbf{U}\mathbf{x}$ by setting

$$\mathbf{x}=\mathbf{U}^{-1}(\tilde{\mathbf{U}}^{-1}\tilde{\mathbf{D}}\tilde{\mathbf{W}}\mathbf{W}\mathbf{D}\mathbf{U}^{-1})^{-1}\tilde{\mathbf{U}}^{-1}\tilde{\mathbf{D}}\tilde{\mathbf{W}}\mathbf{W}\mathbf{y}\,,$$

in which the transformed normal matrix, enclosed in brackets, now has the property that the ith eigenvalue, λ_i, is given by

$$\lambda_i=\frac{\text{decrement in } \tilde{\boldsymbol{\varepsilon}}\tilde{\mathbf{W}}\mathbf{W}\boldsymbol{\varepsilon} \text{ due to the } i\text{th eigenshift}}{\text{increment in } \tilde{\mathbf{x}}\tilde{\mathbf{U}}\mathbf{U}\mathbf{x} \text{ due to the } i\text{th eigenshift}}\simeq\frac{\text{improvement in fit}}{\text{disturbance to structure}}\,,$$

and filtering, based on these eigenvalues may be accomplished by writing

$$\mathbf{x}=\mathbf{U}^{-1}\mathbf{A}\mathbf{Z}(\tilde{\mathbf{A}}\tilde{\mathbf{U}}^{-1}\tilde{\mathbf{D}}\tilde{\mathbf{W}}\mathbf{W}\mathbf{D}\mathbf{U}^{-1}\mathbf{A})^{-1}\tilde{\mathbf{A}}\tilde{\mathbf{U}}^{-1}\tilde{\mathbf{D}}\tilde{\mathbf{W}}\mathbf{W}\mathbf{y}\,,$$

in which $\mathbf{A}$ is orthogonal such that the bracketed matrix is diagonal, and $\mathbf{Z}$ is the filter.

The theoretical basis for writing the inverse covariance matrix for $\tilde{\mathbf{W}}\mathbf{W}$ is well established. Hitherto, however, no comparable theory has existed for the proper assignment of $\mathbf{U}$, and it is to this problem that Scheringer has addressed himself. In this note we wish to draw attention to a number of considerations relevant to the assignment of $\mathbf{U}$ not considered by Scheringer. (Here $\mathbf{U}$ is equivalent to Scheringer's $\mathbf{H}^{-1}$ and to the matrix $\mathbf{W}$ used in paragraph 4·2 of Diamond (1966), where the idea of weighted parameters in connection with filtering was first introduced.)

The question of the assignment of $\mathbf{U}$ depends on whether one wishes to use $\mathbf{U}$ to control convergence in a non-linear problem or to use it to control the amount of disturbance to a trial structure that may arise in an ill-conditioned case, even if linear. These two quite separate functions have differing relevant considerations. Scheringer considers only one problem of the first type, but he makes his criticisms in quite general terms. Consider first the case of the non-linear problem, in particular the refinement of a crystal structure against X-ray data by use of rotational parameters, as Scheringer has done. For small displacements there are *two* assumptions of linearity; firstly, as Scheringer says, 'we treat the motion of the atoms on the arcs as linear', and secondly, we treat the observational quantities (F_c values) as linearly dependent on the positional coordinates of the atoms. The question of the proper choice of $\mathbf{U}$ depends on which of these two linear assumptions breaks down first in the event of the trial structure being far from the true structure. If the latter (dependence of F on position) breaks down first, then $\mathbf{U}$ should be constructed so that filtering conserves the sum of the squares of *distances* moved, weighted according to the Z values, as Scheringer has done, replacing angular parameters by their product with a radius of gyration. On the other hand, if the other linear assumption breaks down first (as it does in the Model Building procedure of Diamond, 1966) it does not make sense to conserve linear displacement by filtering when it is *angular* displacement which governs convergence. In the Model Building procedure, convergence is bound to occur over any distance, however great, provided angular shifts do not exceed ~ 1 radian. Scheringer is, therefore, quite wrong to state 'that Diamond's theory is not correctly established as his basic assumption of a metric in parameter space does not apply' because (i) the use of weighted parameters was introduced in the paper which Scheringer criticizes (and which he does not acknowledge), (ii) the correct choice of weighting scheme for the problem then under consideration was made, and (iii) nothing in the paper which Scheringer criticizes could be construed as meaning that the weighting schemes considered there were represented by Diamond as being appropriate to the problem which Scheringer has since applied filtering to.

In non-linear problems in which filtering is employed to maximize convergence, it is usual, on the final cycle of refinement, to include all eigenshifts, by which stage the form of **U** is unimportant, as any non-singular $\tilde{\mathbf{U}}\mathbf{U}$ will yield the same **x**. Thus, in this context, the choice of **U** is only important in the early cycles, where it may determine which of several minima is eventually found.

In the second class of problem, in which filtering is used to control behaviour in an ill-conditioned problem, the circumstance is quite different in that the final **x** does depend on **U** because, in such cases, the filter does not admit all eigenshifts, even on the final cycle. Here, therefore, the choice of **U** is more important than previously and should be based on quite another set of considerations. Consider a case in which a trial structure has been arrived at by calculations involving the minimization of conformational energy, as is being done in a number of laboratories, then in refining such a structure against X-ray data in which the resolution limit may be too low to determine all structural details, it would be sensible to arrange for the quadratic form $\mathbf{x}\tilde{\mathbf{U}}\mathbf{U}\mathbf{x}$ to represent an increase in the conformational energy. The eigenvalues would then directly express the increment in strain energy per unit decrement in X-ray residual, and filtering could be employed to ensure that disturbances to the trial structure involving large increases in energy for little gain in agreement do not occur. In my current work involving the refinement of flexible chain models of proteins against electron density at less than atomic resolution, the matrix **U** is being employed to make dihedral angles in single bonds elastically softer than inter-bond angles, which is a step in this direction. In these circumstances the radius of gyration idea is not relevant.

To sum up this discussion, it is agreed that Scheringer's choice of the matrix **U** is appropriate to his problem, it is not accepted that his 'orthonormal condition' has universal applicability, nor that the 'basic assumption of a metric in parameter space' invalidates the theory as originally presented. What matters is that the quadratic $\tilde{\mathbf{x}}\tilde{\mathbf{U}}\mathbf{U}\mathbf{x}$ should be related to the properties of the problem.

Scheringer, in addition, criticizes Diamond's derivation of convergence behaviour, a derivation which was offered more as an explanation of why it generally does work, than as a proof that it must. Scheringer points out quite rightly that the eigenshifts are not strictly proportional to $\lambda^{-1/2}$, but that they contain $\lambda^{-1/2}$ as a systematic factor. Scheringer gives examples in which the eigenshifts are not arranged in order of decreasing eigenvalue (as did Diamond), but fails to point out why the trend which he concedes may exist as 'a slight tendency' is unimportant in his case when it is important in the context of model building. The reason is simply that the ratio $\lambda_{\max}/\lambda_{\min}$ is very much smaller in his case than in Diamond's, so that the trend which is undoubtedly present, is largely masked in his case by the essentially random nature of the shifts required, depending, as they do, on the trial structure. The trend may be clearly seen in Fig. 5 of Diamond (1966), which is a typical case. Furthermore, Scheringer's presentation of the reasons why the dominant eigenshifts have the largest range of convergence is essentially equivalent to those of Diamond, which he rejects. He states that 'the geometrical part of the derivatives with respect to the dominant "eigenparameters" will be large in the mean over all structure factors, and this means that the computed dominant eigenshift has a high probability of being correct with respect to convergence' which is formally equivalent to the argument which Diamond adduced and which Scheringer rejects.

Finally, Scheringer states that Diamond's criterion (i) 'is based on a claimed inverse proportionality between the magnitude of the eigenvalues and the eigenshifts which does not exist'. In fact, it is based on a relation between the decrement in $\tilde{\boldsymbol{\varepsilon}}\tilde{\mathbf{W}}\mathbf{W}\boldsymbol{\varepsilon}$ and the increment in $\tilde{\mathbf{x}}\tilde{\mathbf{U}}\mathbf{U}\mathbf{x}$ which does exist, although it is true that one knows only the total $\tilde{\boldsymbol{\varepsilon}}\tilde{\mathbf{W}}\mathbf{W}\boldsymbol{\varepsilon}$ and the upper limit to the total $\tilde{\mathbf{x}}\tilde{\mathbf{U}}\mathbf{U}\mathbf{x}$. Accordingly if any eigenvalue is so small that $\tilde{\boldsymbol{\varepsilon}}\tilde{\mathbf{W}}\mathbf{W}\boldsymbol{\varepsilon}/n\lambda$ (n = order of normal matrix) exceeds the maximum tolerable total $\mathbf{x}\tilde{\mathbf{U}}\mathbf{U}\mathbf{x}$ for convergence, then this eigenshift is contributing to $\mathbf{x}\tilde{\mathbf{U}}\mathbf{U}\mathbf{x}$ to an extent that non-convergent behaviour is to be expected if it is included, but is not bound to occur. With a broad eigenvalue spectrum this provides the best available means of anicipating non-convergence. With a narrow eigenvalue spectrum it is true this criterion could include non-convergent eigenshifts and exclude others with smaller λ which may yet converge, but this is true of Scheringer's criterion also.

References

DIAMOND, R. (1958). *Acta Cryst.* **11**, 129.
DIAMOND, R. (1966). *Acta Cryst.* **21**, 253.
SCHERINGER, C. (1968). *Acta Cryst.* B**24**, 947.