Application of a New Least-Squares Method to Structure Refinement

By John P. Chesick

Department of Chemistry

William C. Davidon

Department of Physics

and Chadmark Schoen

Department of Chemistry, Haverford College, Haverford, Pennsylvania 19041, *U.S.A.*

A new least-squares-type refinement algorithm which updates the parameter values after processing each reflection is tried in comparison with a standard block-diagonal least-squares refinement procedure. A ten-atom problem (C_9S) in space group *Fdd*2, and a 30-atom problem ($C_{26}N_4$) in space group $P2_1/c$ with varying-quality starting coordinate sets and choices of reflection/parameter ratios were used as test cases. With starting atomic coordinates off by at least ± 0.2 Å from the correct values, the new method gives rapid convergence with considerable saving in computation time. The method also gives rapid convergence for both the good and poor starting coordinate sets when the reflection data set for the 30-atom problem was restricted to $d > 2$ Å. For this restricted data set the traditional block-diagonal least-squares method diverged. Computer storage requirements are essentially the same for the new method as for the traditional least-squares methods.

Introduction

Traditional least-squares refinement minimizes the quantity

$$D = \sum_{m=1}^{M} \delta_m^2(a_1, \ldots, a_N) \tag{1}$$

with respect to a list of N parameters $a_1, \ldots, a_N$. Here the index m designates one of the M reflections for which $\delta_m = w_m^{1/2}(F_{m,\text{obs}} - k|F_{m,\text{calc}}|)$, where w_m is an appropriate weighting factor and the other terms have their usual significance. Since the δ_m are non-linear functions of the parameters, a method of successive approximations must be used to find the point in parameter space where D is minimized. Traditionally, the parameters are changed after processing *all M* data items and these new parameters provide the starting point for the next cycle. In the terminology of this paper, this conventional method can be obtained by writing (2), expressing the value of the residual $D(a+x)$ as a three-term Taylor's series expansion about the residual at the coordinate vector **a**:

$$D(\mathbf{a}+\mathbf{x}) = D(\mathbf{a}) + \mathbf{x}^T \nabla D(\mathbf{a}) + \tfrac{1}{2}\mathbf{x}^T[\nabla\nabla^T D(\mathbf{a})]\mathbf{x}$$
$$= \sum_m [\delta_m(\mathbf{a}) + \mathbf{x}^T \nabla \delta_m(\mathbf{a})]^2$$
$$+ \tfrac{1}{2}\mathbf{x}^T[\sum_m \delta_m(\mathbf{a})\nabla\nabla^T \delta_m(\mathbf{a})]\mathbf{x} . \tag{2}$$

Here the transpose of any matrix is denoted by the superscript T, ∇D is the gradient of D, and $\nabla\nabla^T D$ is the $N \times N$ matrix of second derivatives of D. Application of the minimization condition $\nabla D(\mathbf{a}+\mathbf{x}) = 0$ followed by discard of all second and higher-order derivatives

of δ_m gives the Gauss least-squares algorithm used in the conventional crystallographic least-squares refinement procedures to find the coordinate update **x**. In the current notation this is

$$\mathbf{x} = -[\sum_m \nabla\delta_m(\nabla\delta_m)^T]^{-1}(\sum_m \delta_m \nabla\delta_m) .$$

The new method described here makes a small update to the coordinate vector $\mathbf{a}_m$ after processing reflection m so as to minimize φ_{m+1}, a weighted sum of δ_m^2 and preceding square terms, as defined by the equation

$$\varphi_{m+1}(\mathbf{a}_m) \equiv \delta_m^2(\mathbf{a}_m) + \lambda\delta_{m-1}^2(\mathbf{a}_m) + \lambda^2\delta_{m-2}^2(\mathbf{a}_m) + \ldots$$
$$= \delta_m^2(\mathbf{a}_m) + \lambda\varphi_m(\mathbf{a}_m) . \tag{3}$$

Here λ is a number between zero and one, usually very close to unity. As λ is decreased from unity, the speed of convergence to the region about the minimum is increased, but the fluctuations in **a** about a mean value also are larger. In the limit as λ approaches unity, and $\varphi_0 = 0$, the value of φ approaches the value of D given by (1). The update procedure then calculates the new coordinate vector $\mathbf{a}_{m+1}$ to be used to obtain δ_{m+1}^2 for the next reflection. As a result, successive values of the coordinate vector **a** do not converge to a single value, but fluctuate about it with an amplitude which depends on the parameter λ. The approach to constancy of φ replaces constancy of D as a criterion of convergence. On completion of one cycle through the reflection data, one may go back to the first reflection in the data list and continue the update procedure if desired convergence has not been achieved. The optimum set of parameters is obtained by computing the average of **a** over a number of reflections after satisfactory con-

vergence of φ has been reached, or alternatively after some arbitrary point is reached in the list of reflections. This will be discussed in detail later. Operating parameters for the algorithm are then λ and the number of reflections to be considered when finding the average (best) coordinate vector. A rough estimate of probable errors in the components of the starting coordinate vector is also needed as input. No matrix inversions are required in the new method.

Derivation of the algorithm and more detailed justification of the new method will be published elsewhere (Davidon, 1976). We here attempt only to present the algorithm and to describe its use in the crystallographic refinement application.

Algorithm

Inputs

M Number of reflections in data set to be included.
L Convergence weighting parameter

NX Number of reflections in data set to be skipped before start of coordinate averaging for one cycle through the data set.

a N-dimensional starting parameter vector, including scale factor k, and all variable atomic coordinates and temperature factors.

s N-dimensional vector giving estimates of standard deviations for the input components of a. The set of reflection weighting factors, w, is presumed to have been scaled so that the final φ equals $M-N$, to within an order of magnitude.

A schematic flow diagram of the algorithm is shown in Fig 1.

In forming the $N \times N$ symmetric A matrix, the updates create the off-diagonal elements. All such elements may be formed, analogous to full-matrix least squares, or a block or band-diagonal approximation may be used by retaining only the elements of
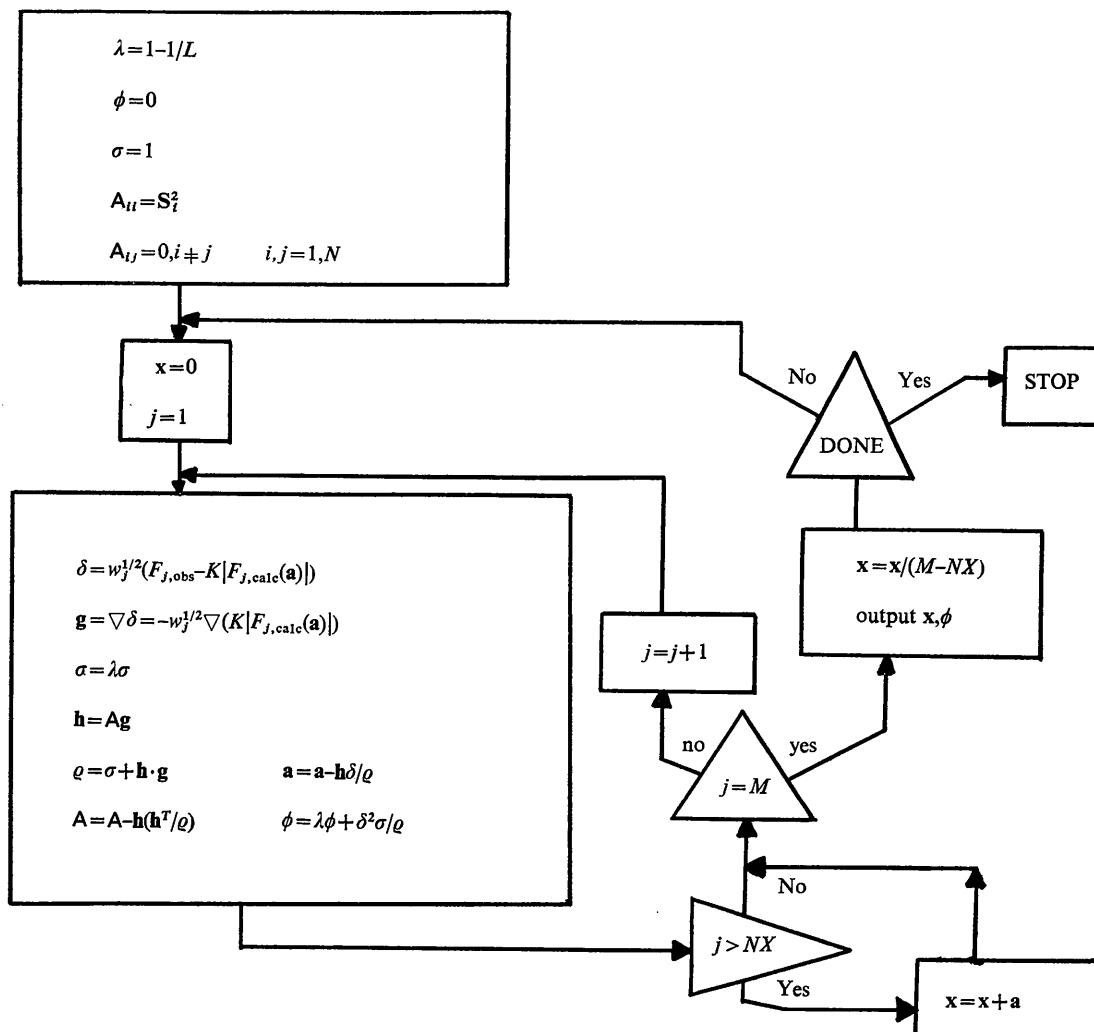

Fig. 1. Flow diagram of the algorithm.

the matrix product $\mathbf{h}(\mathbf{h}^T/\varrho)$ which relate to single atoms or to small numbers of highly correlated atoms. In the trials reported here, the atomic block-diagonal approximation was used for A in which no interatomic terms were included. Only the upper triangle of each symmetric block was actually created and stored.

The matrix A and the vectors $\mathbf{g}$, $\mathbf{h}$, $\mathbf{a}$, and $\mathbf{x}$ require $(K^2+K)/2+4K$, or $(K^2+9K)/2$ words of memory for each block of K parameters, and the update procedure with this algorithm then requires approximately $(3K^2+7K)/2$ multiplications and divisions per block for each reflection. The conventional least-squares refinement method requires about $(K^2+7K)/2$ words of memory and $(K^2+3K)/2$ multiplications and divisions per block for each reflection in addition to approximately $K^3/2+K^2$ multiplications and divisions per block at the end of each data cycle to calculate a matrix inverse or otherwise solve simultaneous linear equations to find the coordinate update. The calculation of δ and the gradient vector $\mathbf{g}$ are common to all methods. The ratio of computation time per data cycle, *exclusive* of time for computation of δ and $\mathbf{g}$, for the new and standard methods is then

$$(3+7/K)/(1+3/K+K/M) .$$

If $K=4$, corresponding to the simplest block-diagonal approximation, this ratio of computation times, comparing our method to the standard method, is then 2·7. If full-matrix methods are used, K becomes large and this time ratio approaches three for a reasonable reflection/parameter ratio. The common overhead time for calculation of δ and $\mathbf{g}$ will reduce these ratios to give values nearer to unity; the new algorithm is seen to incur the least time penalty *per data cycle* for the block-diagonal approximation. Cases with many equivalent positions and/or atoms included in the structure factor calculation which are not refined will have times for calculation of δ and $\mathbf{g}$ which become a large fraction of the total computation time and therefore also give small time penalties per cycle for the new method. Savings for the new method then must result from some significant reduction in number of cycles required for convergence.

Because of small fluctuations in the parameter $(\mathbf{a})$ values computed for each reflection, even after constancy in φ is achieved, the best or final parameter set is *not* the $\mathbf{a}$ calculated for the reflection processed immediately prior to stopping. For reasonably large L values the last $\mathbf{a}$ computed will not be greatly different from an average of $\mathbf{a}$ values computed for a large number of reflections: however, this average $\mathbf{a}$ will give a lower residual than that computed using the terminal $\mathbf{a}$. Taking the parameter NX, as defined in the algorithm flow chart, equal to zero then results in taking an average over coordinates computed for all reflections in a single cycle through the data set. If the coordinates at the start of the cycle are poor, this average will be unduly weighted by the early poor values for a. If the value of NX is large, so that $(M-NX)/M$

is a near unity, the last reflections in the data list may have weight in determining the 'best' coordinates reported for the cycle. If the reflections are in random order in the data list, this would not be a concern. If there is systematic error associated with some index or scattering angle in an ordered data set, then some bias would be included in the 'best' coordinate set obtained from the average over the last $(M-NX)$ values of $\mathbf{a}$ if the ratio $(M-NX)/M$ is too small. Actual experience with varying the operating parameters L, NX, and $\mathbf{s}$ will be described later.

Convergence comparisons

Data for two crystal structures previously solved in this laboratory were used in trials of the new algorithm in comparison with a standard block-diagonal least-squares program. No comparisons were made with full-matrix methods, since it was felt that the greatest value in application of the new method would probably be for the largest problems for which the block-diagonal least-squares techniques are currently required. In all comparisons which follow, BDLS refers to a reference calculation done with the standard block-diagonal least-squares techniques, and DAVM refers to calculations performed with the new algorithm. Data sets and initial crystal-model parameter values are identical in all such comparisons. The sensitivity of the new algorithm to the required input parameters was explored in some detail and the results will be summarized in the last section. The comparisons between BDLS and DAVM were run using 'reasonable' estimates for these operating parameters in DAVM. To compare convergence properties of BDLS and DAVM we present $R=\sum|F_{\mathrm{obs}}-|kF_{\mathrm{calc}}||/\sum F_{\mathrm{obs}}$ and $D=\sum(F_{\mathrm{obs}}-|kF_{\mathrm{calc}}|)^2$ after each cycle (one trip through the reflection data set) for both methods. These quantities were separately computed for the DAVM runs using the average of 'best' coordinates provided as output from each cycle, since the DAVM method does not permit calculation of exactly these criteria of convergence in the normal course of its operation. The quantity φ computed by DAVM provides a monitor for convergence, but is not identically comparable to the residual D which is provided by BDLS and all other standard crystallographic least-squares programs.

(a) Dicinnamyl disulfide

This trial used the reflection data and corrected structure for the compound dicinnamyl disulfide, $(C_9H_9S)_2$, space group $Fdd2$, as recently redetermined by Donohue & Chesick (1975). The asymmetric unit is C_9H_9S, and restriction of $\sin\theta/\lambda \leq 0.55$ and $F_{\mathrm{obs}} \geq 2.5\sigma_{F_{\mathrm{obs}}}$ selected 660 reflections for the calculations out of a total of 722 unique reflections observed. The 'good' starting coordinate set for the ten C and S atoms used in the least squares calculation was derived from the sulfur-phased Fourier map; this coordinate set showed mean deviations of ±0.04 Å in x, ±0.03 Å

in y, and ± 0.08 Å in z from the correct or refined coordinate set. A 'poor' starting set was obtained by randomly adding or subtracting (signs chosen by coin tosses) 0.27 Å to each coordinate of the good starting set. In both cases the scale factor and the isotropic thermal parameter used as input for all atoms were obtained from a Wilson plot. Both starting sets of coordinates were used in block-diagonal isotropic refinement with one block per atom, and the better coordinate set was used as the starting point for aniso-tropic refinement. The convergence parameters obtained for these trials are shown in Table 1 along with the ratio of computation time/cycle for the two methods, DAVM and BDLS.

Table 1. *Dicinnamyl disulfide comparison tests of BDLS and DAVM operation*

		Good start‡		Poor start§	
Isotropic refinement		BDLS	DAVM	BDLS	DAVM
Input	R^*	0.152	0.152	0.509	0.509
	D†	35 532	35 532	451 202	451 202
After cycle 1	R	0.106	0.107	0.411	0.121
	D	18 431	19 464	381 133	23 862
After cycle 2	R	0.102	0.100	0.3026	0.100
	D	16 595	16 372	177 071	16 410
After cycle 3	R	0.100	0.100	0.195	0.100
	D	16 369	16 196	62 835	16 367
After cycle 4	R			0.125	0.100
	D			24 847	16 375

Cycle time ratio, DAVM/BDLS = 1.23

Anisotropic refinement					
Input	R	0.152	0.152		
	D	35 532	35 532		
After cycle 1	R	0.089	0.076		
	D	14 209	14 044		
After cycle 2	R	0.074	0.072		
	D	11 018	11 431		

Cycle ratio time, DAVM/BDLS = 1.49

* $R = \sum |F_o - |kF_c|| / \sum F_o$.
† $D = \sum (F_o - |kF_c|)^2$.
‡ ± 0.04 Å off in x, ± 0.03 Å off in y, and ± 0.08 Å off in z at start.
§ ± 0.27 Å off in each atomic coordinate at start.

From Table 1 we see that the performances per cycle for DAVM and BDLS are virtually the same when a starting set of coordinates is used which gives D within 20% of the convergence limit after one cycle of BDLS. However, if each of the starting atomic coordinates is off by ± 0.27 Å, we see than *one* cycle of DAVM gives results which are better than *four* cycles of BDLS, and a second cycle of DAVM gives essentially the conver-gence values of R and D. The reduction in number of cycles needed for convergence by DAVM greatly offsets the 23% increase in time per cycle required. Small discrepancies between the R and D values for BDLS and DAVM as convergence is reached may be explained by the difference in minimization procedures employed. Coordinate values at convergence do differ on the average by $\frac{1}{2}$ to $\frac{1}{5}$ of the average standard devia-

tions when comparing the BDLS and DAVM results at convergence. The two optimization procedures thus give essentially the same results. Operation of the reference BDLS calculation to limit coordinate shifts to not more than 0.15 Å on each cycle and likewise limiting isotropic B changes to 2.0 Å^2 on each cycle caused no improvement in BDLS convergence for the poor starting set.

(b) Bisphenylazostilbene

This compound, $C_{26}H_{20}N_4$, space group $P2_1/c$ (Chesick, 1973) provides an example with 30 C and N atoms for refinement with a total data set of 3113 unique reflections. A 'good' starting coordinate set was obtained from an E map using phases from a direct-method program. This coordinate set showed mean deviations of ± 0.05 Å in x, y, and z from the correct (refined) values. A 'poor' start was again obtained by randomly adding ± 0.2 Å to each atomic coordinate of the 'good' starting set. Limitations of $\sin \theta / \lambda \leq 0.50$ and $F_{obs} \geq 3\sigma F_{obs}$ selected the data set of 2059 reflections included in the calculations for a reflection/parameter ratio of 17 for isotropic refinement. Interest in refine-ment of very large molecules with a much less favorable reflection/parameter ratio, *e.g.* the protein rubredoxin at 1.5 Å resolution with a ratio close to two (Waten-paugh, Sieker, Herriott & Jensen, 1973) suggested repeating the BDLS *vs* DAVM comparison for both

Table 2. *Bisphenylazostilbene comparison tests of BDLS and DAVM operation*

		Good start‡		Poor start§	
$\sin \theta / \lambda < 0.5$		BDLS	DAVM	BDLS	DAVM
Input	R^*	0.281	0.281	0.669	0.669
	D†	39 573	39 573	244 463	244 463
After cycle 1	R	0.228	0.169	0.592	0.208
	D	22 571	13 560	171 185	21 108
After cycle 2	R	0.172	0.159	0.458	0.162
	D	14 034	11 791	92 969	12 228
After cycle 3	R	0.163	0.159	0.362	0.158
	D	12 446	11 677	57 376	11 741
After cycle 4	R	0.159	0.158	0.292	
	D	11 805	11 630	38 200	

Cycle time ratio, DAVM/BDLS = 1.55

$\sin \theta / \lambda < 0.25$					
Input	R	0.203	0.203	0.537	0.537
	D	7 334	7 334	74 065	74 065
After cycle 1	R	0.183	0.118	0.544	0.230
	D	5 105	2 435	40 543	11 710
After cycle 2	R	0.172	0.102	0.454	0.118
	D	5 014	1 720	38 503	2 673
After cycle 3	R	0.178	0.096	0.488	0.103
	D	6 438	1 462	55 296	1 880
After cycle 4	R	0.176	0.094	0.60	0.098
	D	7 390	1 319	63 723	1 593

Cycle time ratio, DAVM/BDLS = 1.22.

* $R = \sum |F_o - |kF_c|| / \sum F_o$.
† $D = \sum (F_o - |kF_c|)^2$.
‡ ± 0.05 Å off in each atomic coordinate at start.
§ ± 0.2 Å off in each atomic coordinate at start.

starting coordinate sets with sin $\theta/\lambda \leq 0.25$, or $d \geq 2$ Å, which includes only 272 reflections (all of $F_{obs} \geq 3\sigma_{F_{obs}}$) in the refinement. This gives a reflection/parameter ratio of 2·2. Only isotropic refinement, one atom/block, was done in the comparisons of the two methods for bisphenylazostilbene. The results of these comparisons are shown in Table 2.

Table 2 shows that for the good start with the large data set, the smaller number of cycles for convergence of DAVM offsets the 55% larger time/cycle required. The superiority of DAVM is rather striking when considering the large data set and the set of starting atomic coordinates which are ± 0.2 Å from the final values. DAVM has essentially converged after two cycles, and BDLS will require seven or eight cycles for convergence. With the restricted data set, DAVM converges in two or three cycles, and BDLS diverges. The use of the set of starting coordinates with the ± 0.2 Å offsets only retarded the convergence by one cycle when using the DAVM calculation.

Restriction of coordinate shifts to less than 0·15 Å and isotropic B shifts to less than 2·0 Å^2 on each cycle for BDLS trials with sin $\theta/\lambda \leq 0.25$ still gave divergence with R and D values similar to those given in Table 2. This restriction of BDLS parameter shifts for the poor start with sin $\theta/\lambda \geq 0.5$ gave values of R and D of 0·271 and 29 617 after the fourth cycle, in contrast to the values of 0·292 and 38 200 shown in Table 2 for unconstrained shifts. Thus significant but not spectacular improvement was made in that particular BDLS reference calculation by limiting the parameter shifts. The number of cycles for BDLS convergence will be essentially the same.

Operating parameters

(a) Initial A matrix

As the algorithm flow chart indicates, the initial A matrix is taken to be diagonal with the square of the estimated standard deviations or probable errors of the starting coordinates as the diagonal elements. These uncertainties in the initial atomic coordinates and thermal parameters are usually readily estimated or known from the nature of the source for the starting coordinate set. Trials using both the good and poor initial coordinate sets and also the limited data set indicated that the input s matrix used to initialize A is uncritical. Increasing or decreasing the best estimate of s by a factor of four made little difference in convergence in the most extreme case, the poor initial coordinate set using data for sin $\theta/\lambda > 0.25$ in the bisphenylazostilbene trials. With somewhat better starting coordinate sets an increase or decrease in the estimate of s by a factor of 10 seemed to have little effect on the speed of convergence.

(b) Choice of λ

The parameter λ is generally close to unity in this crystallographic refinement application and is therefore most conveniently defined and discussed in terms of the variable L in the expression $\lambda = 1 - 1/L$. The value of L also seemed to be relatively uncritical in the trials with the two data sets. The general experience seems to indicate that L should be at least two or three times the number of parameters and if possible in the range of $M/3$ to M, where M is the number of reflections included in the least-squares. In the bisphenylazostilbene trials with $M = 272$, $L = 500$ gave slightly better results than $L = 250$ when comparing the results after two cycles. $L = 121$ gave significantly poorer results. With the full data set for this compound ($M = 2059$) $L = 1000$ gave slightly faster convergence than $L = 480$. Again, the results indicate that L is not a critical operating parameter.

(c) Coordinate averaging parameter NX

Since most of the coordinate adjustment seems to take place in the first part of the first cycle through the data set, the selection of $NX \neq 0$ is of considerable importance only if the optimum coordinate set is to be provided from only one cycle. In this case trials suggest that a value of NX giving $(M - NX)/M$ between $\frac{1}{3}$ and $\frac{3}{4}$ is a reasonable compromise of the concerns mentioned earlier. If a coordinate set is desired after more than one computation cycle, then selection of a smaller NX value to give the nearer to unity is suggested. The trials reported here used a fixed value of NX for each cycle in a calculation of more than one cycle. A simple program change would permit variation of NX with the cycle number if one wished to follow the changes lin the coordinate average from cycle to cycle.

References

CHESICK, J. P. (1973). *Acta Cryst.* B29, 2304–2311.
DAVIDON, W. C. (1976). *J. Optimization Theory Appl.* 18, No. 2 (in the press).
DONOHUE, J. & CHESICK, J. P. (1975). *Acta Cryst.* B31, 1068–1072.
WATENPAUGH, K. D., SIEKER, L. C., HERRIOTT, J. R. & JENSEN, L. H. (1973). *Acta Cryst.* B29, 943–956.