



Soviet-era science, translated into English

Mathematics

Corresponding Member of the Academy of Sciences of the USSR I.
M. Gelfand, I. I. Pyatetskii-Shapiro,

1963

SovietRxiv

View the original and related papers at <https://sovietrxiv.org/items/ru-196301.11350>

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.

Abstract

Full Text

Mathematics

Corresponding Member of the Academy of Sciences of the USSR I. M. Gelfand,
I. I. Pyatetskii-Shapiro,
Yu. G. Fedorov

FINDING THE STRUCTURE OF CRYSTALS BY MEANS OF THE METHOD OF NONLOCAL SEARCH

The basic problem of X-ray structural analysis of crystals is the determination of the coordinates of the atoms in the elementary cell of a crystal from the results of measuring the intensities of scattering of X-rays. Up to the present time there has been no regular method for solving this problem. In our note we shall show that it is possible to formulate the problem as that of finding the global minimum of a certain function of many variables, knowledge of which makes it possible to determine the coordinates of the atoms in the crystal cell.

The problem of finding the extremum of a function of many variables is in itself a very difficult computational problem. The usual methods of local search (the gradient method, steepest descent, etc.) do not lead to the goal. However, ideas about the “good organization” of the corresponding function and the method of nonlocal search based on these ideas (the “ravine method”) ⁽¹⁾ make it possible to hope for the effectiveness of the computational method proposed here, which has previously been successfully applied in problems of phase analysis ^(2,3).

1°. The arrangement of the atoms of substances in crystals is characterized by the electron density $\rho(x, y, z)$, whose maxima correspond to the positions of the atoms; the magnitudes of these maxima are proportional to the number of electrons of the corresponding atoms. The function $\rho(x, y, z)$ is periodic in the arguments x, y, z with periods a, b, c , respectively. The region $0 \leq x \leq a$, $0 \leq y \leq b$, $0 \leq z \leq c$ is called the elementary cell. For simplicity we shall henceforth assume that $a = b = c = 1$.

Represent the function $\rho(x, y, z)$ by the Fourier series

$$\rho(x, y, z) = \sum_{h,k,l} F_{hkl} \exp[-2\pi i(hx + ky + lz)]. \quad (1)$$

From experiment one obtains only the quantities $F_{hkl}^{\text{expt}} = |F_{hkl}|$ (the moduli of the structural amplitudes of the corresponding reflections), which, of course, is insufficient for determining the function $\rho(x, y, z)$. Additional information

about the structure of this function is necessary. It is customary to assume ^(4,5) that the electron density of the substance contained in the elementary cell can be represented in the form of a sum of the electron densities of separate atoms or ions. Thus,

$$\rho(x, y, z) = \sum_{i=1}^N \rho_i(x - x_i, y - y_i, z - z_i), \quad (2)$$

where the functions $\rho_i(\xi, \eta, \zeta)$ are known functions of their arguments (the electron densities of separate atoms or ions), and N is the total number of atoms in the elementary cell. The coordinates x_i, y_i, z_i of the atoms are unknown and are to be determined. Generally speaking, the number of atoms to be determined in the elementary cell can be reduced if the requirements of group invariance are taken into account. This, however, is not essential for our purposes, and in what follows we shall not make use of it.

From (1) and (2) it follows that

$$F_{hkl} = \sum_{j=1}^N f_j(h, k, l) \exp[2\pi i(hx_j + ky_j + lz_j)]. \quad (3)$$

where

$$f_j(h, k, l) = \int \rho_j(\xi, \eta, \zeta) \exp[-2\pi i(h\xi + k\eta + l\zeta)] d\xi d\eta d\zeta. \quad (4)$$

The functions $f_j(h, k, l)$ are called atomic factors. Detailed tables are available for them. Let us consider the function $\Phi_{\alpha\beta}$ of $3N$ variables—the coordinates of the atoms x_i, y_i, z_i ; $i = 1, 2, \dots, N$:

$$\Phi_{\alpha\beta}(x_1, y_1, z_1, x_2, \dots, y_N, z_N) = \sum_{h,k,l} \left| |F_{hkl}|^\alpha - (F_{hkl}^{\text{exp}})^\alpha \right|^\beta, \quad (5)$$

where F_{hkl} —the structure amplitudes—are defined in (3), F_{hkl}^{exp} is some set of experimental values of the moduli of the structure amplitudes, and the summation is carried out over the corresponding set of h, k, l . Thus, the function $\Phi_{\alpha\beta}$ can be formed for different sets of h, k, l . The parameters α and β may be chosen differently, depending on the nature of the particular problem.

This function has a minimum at the point $(x_1, y_1, z_1, x_2, \dots, y_N, z_N)$ corresponding to the desired position of the atoms in the cell. Generally speaking, the value of this minimum is different from zero, and the point of the minimum itself corresponds to the desired position of the atoms only with a certain error determined by the accuracy of the experiment.

The function $\Phi_{\alpha\beta}(x_1, y_1, z_1, x_2, \dots, y_N, z_N)$, as a rule, has a complicated, “poorly organized” structure with an enormous number of minima, most of which do not correspond to the desired arrangement of the atoms. Therefore solving the problem by means of $\Phi_{\alpha\beta}(x_1, y_1, z_1, x_2, \dots, y_N, z_N)$ is, generally speaking, very difficult.

2°. For a large class of crystal structures (for example, for crystals of many organic compounds) the problem of finding the required minimum can be substantially simplified. In such crystals the totality of all atoms in the unit cell breaks up into a small number of parts, in each of which the atoms are bound to one another by rigid or semirigid bonds (molecules). In other words, each of these parts may be regarded as a rigid body, possibly with additional degrees of freedom. In this case the positions of the atoms in the unit cell are specified by a comparatively small number of parameters (the degrees of freedom of the rigid bodies and the additional degrees of freedom). Denote the new variables by x_1, x_2, \dots, x_p , and express through them the coordinates of the atoms. Put

$$\tilde{\Phi}_{\alpha\beta}(x_1, x_2, \dots, x_p) = \Phi_{\alpha\beta}[x_1(x_1, x_2, \dots, x_p), \dots, z_N(x_1, x_2, \dots, x_p)]. \quad (6)$$

Thanks to this “natural organization,” the function $\tilde{\Phi}_{\alpha\beta}$ is constructed much more simply than the originally taken function $\Phi_{\alpha\beta}$.

We propose reducing the problem of X-ray structural analysis to the search for the global minimum of the function $\tilde{\Phi}_{\alpha\beta}$ over the full domain of variation of the variables x_1, x_2, \dots, x_p .

Let us note that the function $\Phi_{\alpha\beta}$ (normalized in the appropriate way) is often used in X-ray structural analysis [5], for example, in solving the problem of refining atomic coordinates and checking the correctness of a structure. In this case one usually takes the values $\alpha = \beta = 2$ or $\alpha = \beta = 1$. We also note that in these cases, to construct the function $\Phi_{\alpha\beta}$, all measured reflections are usually used (100-1000 of them). In contrast to this generally accepted way of using the function $\Phi_{\alpha\beta}$, we propose: first, to seek the minimum of the function $\tilde{\Phi}_{\alpha\beta}$, not $\Phi_{\alpha\beta}$; second, to seek it globally, not locally; and third, when constructing the function $\tilde{\Phi}_{\alpha\beta}$ for a preliminary search of the region of the minimum in the variables x_1, x_2, \dots, x_p , to use a relatively small number of reflections (significantly fewer than the total number usually measured).

The latter is due to the fact that the object (molecule) whose position in the unit cell we determine has dimensions considerably larger than those of individual atoms.

It may turn out that the function $\tilde{\Phi}_{\alpha\beta}$, formed from a small set of reflections, has several “equivalent” minima. Selecting the minimum corresponding to the desired arrangement of atoms is associated with a significant increase in this set of reflections. Along with this, one can use additional (geometrical) conditions on the admissible arrangements of molecules relative to one another. A

more detailed use of these conditions is intended to be described in subsequent publications.

3°. In searching for minima of the function $\tilde{\Phi}_{\alpha\beta}$, we propose using the method of nonlocal search (the method of “ravines”), first applied to problems of phase analysis ^(2,3). The basic concepts of the method of “ravines” (“essential” and “inessential” variables, a well-organized function, a step along a ravine, a gradient trial, etc.) are described in ⁽¹⁾. In this note we shall briefly describe only some essential additions to the search procedure used earlier.

The total amount of computation in solving the problem of finding the minimum of a function $f(x_1, x_2, \dots, x_n)$ is, roughly speaking, proportional to the number of evaluations of the function f . To find the gradient of the function f it is necessary to compute its value $n + 1$ times, which for large n amounts to a significant number of operations. Finding the minimum of the function f on a straight line is much “cheaper.” For this purpose it is necessary to compute the values of f at 3-4 points of the line.

The methods proposed here make it possible to reduce the total number of computations of the function f and, what is essential, make it possible to judge the character of the relief of the function f in a large region of variation of the variables x_1, x_2, \dots, x_n , by comparing the local structure of the function f at various points of this region.

Recall that by ravine points, or descent points, we mean those points at which the gradient descent ends (points A_i , Fig. 1 ⁽¹⁾). Points of departure, or non-descent points, in motion along a ravine are the points from which the gradient descent begins (points X_i , *ibid.*).

I. Let $A_i(\bar{x}_1^i, \bar{x}_2^i, \dots, \bar{x}_n^i)$ be the last of the available ravine points; $g_1^i, g_2^i, \dots, g_k^i$ a sequence of descent directions by means of which this point was obtained from its own non-descent point $X_i(x_1^i, x_2^i, \dots, x_n^i)$. Let $X_{i+1}(x_1^{i+1}, x_2^{i+1}, \dots, x_n^{i+1})$ be a new point of departure. For the descent from the point X_{i+1} we first use the vectors $g_1^i, g_2^i, \dots, g_k^i$, namely, we shall decrease the value of the function $f(x_1, x_2, \dots, x_n)$ successively along the straight lines defined by the vectors $g_1^i, g_2^i, \dots, g_k^i$. From the point obtained as a result $(\tilde{x}_1^{i+1}, \tilde{x}_2^{i+1}, \dots, \tilde{x}_n^{i+1})$ we additionally carry out an ordinary gradient descent and obtain a new ravine point $A_{i+1}(\bar{x}_1^{i+1}, \bar{x}_2^{i+1}, \dots, \bar{x}_n^{i+1})$. Let h_1, h_2, \dots, h_m be the sequence of vectors of the additional gradient descent. From the totality of vectors $g_1^i, g_2^i, \dots, g_k^i; h_1, h_2, \dots, h_m$, according to some “reasonable” principle, the most “useful” vectors are selected, which we denote by $g_1^{i+1}, g_2^{i+1}, \dots, g_l^{i+1}$. We use these vectors for the computation at the next point, and so on.

One may, for example, use the following selection principles:

1. From the $k + m$ descent directions, the k “best” ones (according to the magnitude Δ , see ⁽¹⁾) are selected and used at the next point.
2. From the $k + m$ directions, all “ineffective” ones are excluded; the remaining ones are used at the next point and are supplemented, when necessary, by

one new direction.

In the selection, one may also take into account the effectiveness of a given direction at several preceding points (“memory”).

Thus the number of vectors k can be made variable in different—

points of the ravine. Important nonlocal information about the function f is, for example, the “invariance” of certain vectors from the sets $g_1^i, g_2^i, \dots, g_k^i$.

- II. Span, on the vectors $g_1^i, g_2^i, \dots, g_k^i$, a subspace G_k^i (we shall assume that $g_1^i, g_2^i, \dots, g_k^i$ form an orthonormal basis in G_k^i). To compute the directions of descent from the new starting point $X_{i+1}(x_1^{i+1}, x_2^{i+1}, \dots, x_n^{i+1})$, we first use only vectors from G_k^i (i.e., we shall find only the projections of the gradient in the full space onto the subspace G_k^i). In this case gradient descent (let us call it descent in G_k^i) is carried out with its own gradient test. From the point thus obtained $(\tilde{x}_1^{i+1}, \tilde{x}_2^{i+1}, \dots, \tilde{x}_n^{i+1})$, we additionally perform ordinary gradient descent and obtain a new point of the ravine $A_{i+1}(\bar{x}_1^{i+1}, \bar{x}_2^{i+1}, \dots, \bar{x}_n^{i+1})$. Let $\tilde{g}_1^i, \tilde{g}_2^i, \dots, \tilde{g}_s^i$ be the descent vectors in the subspace G_k^i ; h_1, h_2, \dots, h_m the vectors of the additional gradient descent. Span, on the vectors $\tilde{g}_1^i, \tilde{g}_2^i, \dots, \tilde{g}_s^i, h_1, h_2, \dots, h_m$, a subspace G_l^{i+1} and take in it an orthonormal basis $g_1^{i+1}, g_2^{i+1}, \dots, g_l^{i+1}$ (using, as far as possible, the basis $g_1^i, g_2^i, \dots, g_k^i$). The new subspace G_l^{i+1} is used for the computation at the next point, and so on. Generally speaking, the dimensions of the subspaces G_k^i and G_l^{i+1} almost always coincide and are small compared with n , except at places where the relief of the function f changes sharply. Information about the bases of the subspaces G_k^i constitutes essential information about the behavior of the function f in a large region.

As a rule, in the additional gradient descents in these methods it is necessary to compute only 1-2 gradients, which substantially reduces the total number of computations.

In conclusion, we note that the proposed method for finding crystal structures was preliminarily tested on a simple known structure (naphthalene).

We express our gratitude to Acad. N. N. Semenov, who drew our attention to the desirability of creating new direct methods in X-ray structural analysis. N. S. Andreeva, B. K. Vainshtein, A. I. Kitaigorodskii, M. A. Porai-Koshits, and A. A. Levin took part in the discussion of the basic provisions of this article. Conversations with M. L. Tsetlin were also useful to us. L. N. Ivanova, S. L. Ginzburg, E. I. Dinaburg, and M. M. Voronovitskii took an active part in carrying out the computations. We express our deep gratitude to all of them.

Received
13 VII 1963

REFERENCES

1. I. M. Gelfand, M. L. Tsetlin, *UMN*, 17, no. 1 (103) (1962).
2. I. M. Gelfand, A. F. Grashin, I. Ya. Pomeranchuk, V. A. Borovikov, *ZhETF*, 40, no. 4 (1961).
3. I. M. Gelfand, A. F. Grashin, L. N. Ivanova, *ZhETF*, 40, no. 5 (1961).
4. A. I. Kitaigorodskii, *X-ray Structural Analysis*, 1950.
5. A. I. Kitaigorodskii, *Theory of Structural Analysis*, Publishing House of the Academy of Sciences of the USSR, 1957.

Note: Figure translations are in progress. See original paper for figures.

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.