electronic reprint

Journal of
Applied Crystallography
ISSN 0021-8898

Mathematical methods in small-angle scattering data analysis

D. I. Svergun

J. Appl. Cryst. (1991). **24**, 485–492

J. Appl. Cryst. (1991). **24**, 485–492

Mathematical Methods in Small-Angle Scattering Data Analysis

BY D. I. SVERGUN

*GKSS Research Center, GKSS-WS, 2054 Geesthacht, Germany**

(*Received* 7 *August* 1990; *accepted* 24 *January* 1991)

Abstract

Applications of modern mathematical methods to the problems of small-angle scattering data treatment and interpretation are considered. Special possibilities in data treatment, namely simultaneous treatment of data sets obtained with different experimental conditions and joint data processing for the anomalous-dispersion experiments, are presented. The methods are further developments of the indirect method based on the regularization technique [Svergun, Semenyuk & Feigin (1988). *Acta Cryst.* A**44**, 244–250]. Recent improvements in the shape-determination technique based on the multipole expansion theory [Stuhrmann (1970). *Z. Phys. Chem.* (*Frankfurt am Main*), **72**, 177–184, 185–198] are described. A new method is proposed for the determination of positioning and mutual orientation of subunits in complex particles using the spherical harmonics technique.

1. Introduction

Small-angle scattering (SAS) of X-rays and neutrons is widely used in the structure analysis of dispersed objects of various kinds. During the last decade, major improvements have occurred in the field of SAS instrumentation and data acquisition (powerful synchrotron and neutron sources, position-sensitive detectors and image plates *etc.*). Therefore, more and more complicated structural tasks can be put forward for the method to solve. However, extracting structural information from the SAS data is often a difficult problem owing to lack of ordering in the samples under investigation. As a result, only some general structural characteristics (invariants) can be evaluated directly from the SAS intensity curves. To obtain more detailed structural information one needs a *priori* knowledge about the object and sophisticated methods of data analysis.

The field of application of mathematical methods in SAS can be divided into two parts. First, there are methods of data treatment, allowing one to get rid of the instrumental distortions. Second, there are methods of data interpretation, which are aimed at the evaluation of structural characteristics of the object from the SAS intensity curves. Modern mathematical methods are widely used in both fields. In this paper new possibilities in data treatment and interpretation provided by regularization methods and the spherical harmonics technique combined with optimization algorithms are described.

2. Applications of the regularization technique

Application of Tikhonov's regularization method (Tikhonov & Arsenin, 1979) to the problems of SAS data treatment has been described by Svergun, Semenyuk & Feigin (1988). In particular, the regularization technique is implemented in the so-called indirect methods. The indirect approach in the SAS data treatment deals with the general operator equation

$$\mathbf{K}\,\mathbf{p} = \mathbf{J}. \tag{1}$$

Here vector **J** represents the experimental data set, **K** is an integral operator, describing the Fourier transformation and the experimental conditions, **p** is a real-space distribution function to be found. For example, in the case of a monodisperse system (1) can be written as

$$\int_0^D K(s,r)p(r)\,\mathrm{d}r = J(s), \tag{2}$$

where s is the modulus of the scattering vector **s** [$s = (4\pi/\lambda)\sin\theta$, λ is the wavelength, 2θ scattering angle], D is the maximum particle diameter, p is the distance distribution function [$p(r) \equiv 0$ for $r > D$] and the integral kernel has the form

$$K(s,r) = \int_{-\infty}^{\infty}\mathrm{d}u\int_{-\infty}^{\infty}\mathrm{d}v\int_0^{\infty}\mathrm{d}\lambda\, W_w(u)W_h(v)W_\lambda(\lambda) \times S\{[(s-u)^2+t^2]^{1/2}r/\lambda\}, \tag{3}$$

where $S(x) = \sin(x)/x$, $W_w(u)$ and $W_h(v)$ are the weighting functions describing the beam divergency, $W_\lambda(\lambda)$ describes its polychromaticity (see *e.g.* Feigin & Svergun, 1987). A similar equation holds for polydisperse systems.

Solving (1) with respect to the function $p(r)$ is an ill-posed problem, *i.e.* the solution is unstable with

* Permanent address: Institute of Crystallography, Academy of Sciences of the USSR, 117333 Moscow, Leninsky pr. 59, USSR.

0021-8898/91/050485-08$03.00

respect to the experimental errors in **J**. The regularization method enables one to find a stable solution by minimization of the Tikhonov functional

$$\mathbf{T}_\alpha(\mathbf{p}) = ||\mathbf{J} - \mathbf{K}\mathbf{p}||^2 + \alpha\Omega(\mathbf{p}), \tag{4}$$

where $||\mathbf{J} - \mathbf{K}\mathbf{p}||$ denotes the function norm, $\Omega(\mathbf{p})$ is a stabilizer containing *a priori* information about the solution, α is a (non-negative) regularization parameter. In practical applications the functional is normally taken in the form.

$$\mathbf{T}_\alpha[\mathbf{p}] = \sum_{i=1}^{N}(1/\sigma_i^2)\left[J(s_i) - \int_0^D p(r)K(s_i,r)\,\mathrm{d}r\right]^2 + \alpha\int_0^D[\mathrm{d}p(r)/\mathrm{d}r]^2\mathrm{d}r \tag{5}$$

[here N is the number of experimental points, $J(s_i)$ and σ_i denote the intensity and standard deviation at the ith point, respectively, whereas the stabilizer demands smoothness of the solution]. The solution thus obtained depends on the regularization parameter α. The value of α can be estimated, for example, by the so-called discrepancy method (Tikhonov & Arsenin, 1979), in which the condition $||\mathbf{J} - \mathbf{K}\mathbf{p}|| = 1$ is to be satisfied.

Svergun, Semenyuk & Feigin (1988) realized the approach described above in a program package *GNOM* which allows for automatic estimation and refinement of the α value. In principle, the package has features similar to those of other indirect methods (Glatter, 1977; Moore, 1980; Provencher, 1982; Mangani, Puliti & Stefanon, 1988). It allows reliable solutions to be obtained, being stable with respect to the experimental errors and termination effects. On the other hand, the value of D should be estimated *a priori*, which is a general demand for all indirect approaches.

Apart from providing stable results, indirect methods allow some extra possibilities. In particular, with *GNOM* [as well as with Provencher's (1982) algorithm] one can impose the condition of non-negativity of the solution. Here some additional possibilities offered by the indirect method based on the regularization technique are presented.

2.1. *Simultaneous treatment of different data sets*

In experimental practice one often faces the situation where one and the same object is to be measured with two (or sometimes more) configurations of the experimental set-up in order to cover a wide range of momentum transfer. For example, the innermost portion of the scattering curve needs to be measured with nearly perfect collimation conditions in order to reach the lowest s values, whereas the outer part can be recorded with higher beam divergency to increase the scattering intensity. We shall denote the two measuring regions as $[s_{1b}, s_{1e}]$ and $[s_{2b}, s_{2e}]$. In order to obtain the full scattering curve, one should merge the two data sets using the overlap interval $[s_{1e}, s_{2b}]$.

The merging procedure can be carried out more easily when the two sets have been recorded under the same conditions. If the conditions are different, one should in principle desmear the sets before merging. However, separate desmearing could become a complicated task because of strong termination effects (in particular, the data sets may not be sufficiently representative to be treated by indirect methods).

Let us consider the possibility of simultaneous treatment of two runs by the indirect approach. If one has two data sets $\mathbf{J}_1$ and $\mathbf{J}_2$ corresponding to the experimental conditions described by the operators $\mathbf{K}_1$ and $\mathbf{K}_2$, then (1) can be generalized as

$$\begin{pmatrix}\mathbf{K}_1\\ \mathbf{K}_2\end{pmatrix}p = \begin{pmatrix}\mathbf{J}_1\\ \chi\mathbf{J}_2\end{pmatrix}. \tag{6}$$

Here the second run is multiplied by a constant χ, showing that the data sets may be available only on a relative scale (which is often the case in practice). If the constant was known, (6) could be solved by the regularization method according to (5). In order to find its value (6) can be treated as a system of equations with respect to the vector $\mathbf{p}_1 = (\mathbf{p},\chi)$. The system can be solved by a conventional least-squares technique by minimizing the functional

$$\Phi(\mathbf{p},\chi) = \left|\left|\begin{pmatrix}\mathbf{K}_1\\ \mathbf{K}_2\end{pmatrix}p - \begin{pmatrix}\mathbf{J}_1\\ \chi\mathbf{J}_2\end{pmatrix}\right|\right|^2, \tag{7}$$

which leads to the system of normal equations

$$\partial\Phi/\partial\mathbf{p} = 0, \tag{8}$$

$$\partial\Phi/\partial\chi = 0. \tag{9}$$

When using the norm definition according to (5), χ can be expressed from (9) as a function of **p** and then substituted into (8). It is readily shown that the obtained linear system is symmetric. After solving (6) with the least-squares method and backward substitution the value of χ is obtained. This value determines the self-consistency of the system (6) [that is, with a wrong χ value the two data sets cannot correspond to the same function $p(r)$]. When writing (9) explicitly, it is seen that χ is obtained from $p(r)$ by an integration operation. Therefore, although the vector **p** obtained by the least-squares method is unstable within the experimental errors and cannot be regarded as a solution of (6), the estimation of χ can be used for solving this equation by the regularization technique.

An example of the application of the this approach is given by a contrast-variation study of the bacterial virus $T7$. The measurements were carried out on the $D11$ instrument (Institut Laue–Langevin, Grenoble, France). The average wavelength of the neutron

beam was $\lambda = 10$ Å, the beam polychromaticity $\Delta\lambda/\lambda = 8\%$. The data were registered by a two-dimensional position-sensitive detector and subsequently circular averages were taken. To cover a larger angular range, different sample-to-detector distances were used, namely the distance 10 m corresponding to the momentum-transfer region $(s_{1b}, s_{1e}) = (0{\cdot}003, 0{\cdot}023)$ Å^{-1} and the distance 3 m $[(s_{2b}, s_{2e}) = (0{\cdot}011, 0{\cdot}075)$ Å$^{-1}]$. Fig. 1 shows two typical data sets obtained for the phage solution containing 65% D_2O. Taken separately, they are not sufficiently representative to produce reliable results. The result of their simultaneous treatment is shown in Fig. 2. Evaluated distance-distribution functions for different contrasts are presented in Fig. 3. One can see the maxima shift towards larger dimensions with the increase of the protein contribution to the total scattering indicating that the protein component occupies the periphery of the phage particle (Ronto, Toth, Timmins, Feigin & Svergun, 1991).

It should be noted that the described approach is of rather general value. In particular, the two measured regions may not overlap. Data sets obtained with different devices or even with different radiations (say, X-rays and neutrons) can be treated together in this way. The approach can be easily estended for more than two data sets.

2.2. *Anomalous-dispersion data treatment*

The use of anomalous dispersion in SAS is now a powerful tool in structural studies (Stuhrmann, 1985). Anomalous-scattering experiments with synchrotron radiation allow one to obtain unique information about the inner structure of the sample. If one considers X-ray scattering at a wavelength λ near the absorption edge of a particular type of atom in the scattering object, then the intensity can be written as

$$J(s,\lambda) = J_0(s) + A(\lambda)J_{a0}(s) + B(\lambda)J_a(s), \quad (10)$$

where $J_0(s)$ is the scattering intensity far from the absorption edge of the resonant atoms, $J_a(s)$ is the extra scattering intensity due to these atoms, $J_{a0}(s)$ the cross term. Here $A(\lambda) = 2f'$, $B(\lambda) = (f')^2 + (f'')^2$, f' and f'' being real and imaginary parts of the anomalous-dispersion corrections. Normally, a series of measurements is made for different $\lambda = \lambda_j$ $(j = 1,\ldots,M)$ in the vicinity of the absorption edge $(\lambda = \lambda_{abs})$, where the corrections are significant. Then the recorded set of intensities is developed into the basic functions (10), which gives the information about the distribution of the resonant atoms.

In principle, the problem seems to be rather trivial. If $M > 3$ scattering curves for different λ_j are available, one has for each angular point a system of linear equations with three unknowns, which can be

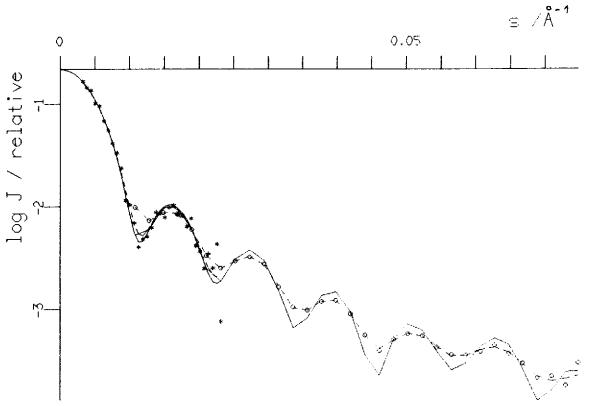

Fig. 2. Simultaneous treatment of the two sets shown in Fig. 1. Asterisks and open circles indicate experimental points, the solid line shows the desmeared curve, dashed lines represent the smeared curves.

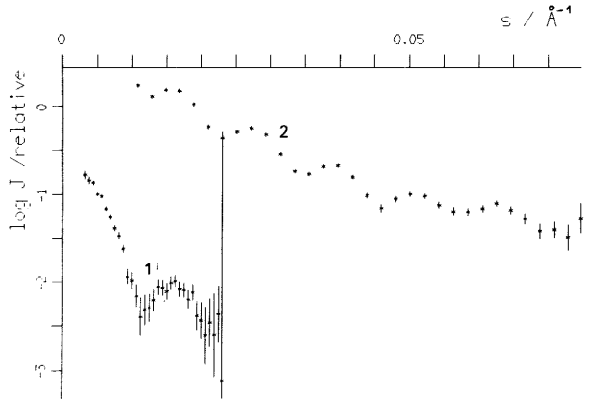

Fig. 1. Experimental neutron SAS intensities from the solution of phage $T7$ with 65% D_2O, corresponding to sample–detector distances 10 m (1) and 3 m (2).

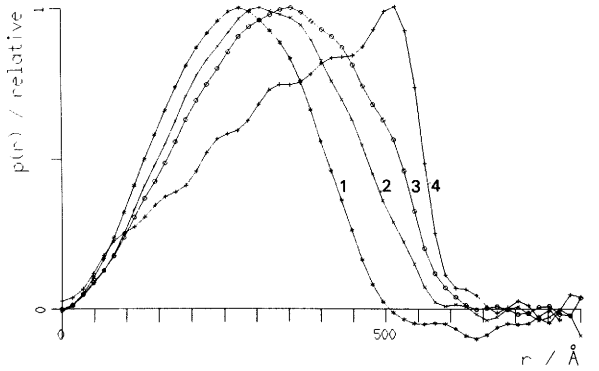

Fig. 3. Distance distribution functions of phage $T7$ obtained by simultaneous treatment procedure. (1) corresponds to a D_2O content of 42% (protein is matched); (2) to 0% D_2O (relative contribution of protein scattering equals 0·31); (3) to 100% D_2O (0·5); (4) to 65% D_2O (1·0).

solved by the least-squares technique. This is true, for example, in classical contrast-variation studies, where the intensity at a given angle is just a second-degree polynomial of contrast (Stuhrmann & Kirste, 1965). In anomalous-scattering studies, however, the wavelength dependencies of $A(\lambda)$ and $B(\lambda)$ can be very similar in the region where the corrections are most important. This means that the matrix $\mathbf{U}_\lambda = (1,\mathbf{A},\mathbf{B})$ of such a system is poorly conditioned and the solution will be unstable. Moreover, in many cases (*e.g.* in the investigations of biological macromolecules in solution) the anomalous term is essentially smaller than the non-resonant one. As a result, one cannot expect this straightforward approach to give reliable results.

The problem, however, can be formulated in terms of the indirect approach. For the distribution function one can write an equation similar to (10), namely

$$p(r) = p_0(r) + A(\lambda)p_{a0}(r) + B(\lambda)p_a(r). \qquad (11)$$

Since each basic distribution function is connected to the corresponding basic intensity by (1), it is possible to write

$$\mathbf{K}\,\mathbf{U}_\lambda\mathbf{p}_\lambda = \mathbf{J}_\lambda, \qquad (12)$$

where $\mathbf{p}_\lambda = (\mathbf{p}_0,\mathbf{p}_{a0},\mathbf{p}_a)$, $J_\lambda = (J_1,J_2,\ldots,J_M)$. This equation can be solved with respect to the three vectors $\mathbf{p}_0$, $\mathbf{p}_{a0}$ and $\mathbf{p}_a$ by the described regularization technique. In order to take into account differences in magnitudes between the resonant and non-resonant terms, the stabilizer should be taken in the form

$$\Omega(\mathbf{p}_\lambda) = \Omega(\mathbf{p}_0) + \beta^{-1}\Omega(\mathbf{p}_{a0}) + \beta^{-2}\Omega(\mathbf{p}_a). \qquad (13)$$

Here β is the expected fraction of the resonant scattering, which can be predicted from the known chemical composition of the particle.

In comparison to the conventional case [(1)], the system (12) has three times as many unknowns and M times more experimental values. Therefore it is a more time-consuming procedure, which also requires handling matrices of high dimensions. However it generates reliable solutions for the cases when conventional methods fail to produce reasonable results. Examples of the practical application of this approach to monodisperse systems can be found in Kühnholz (1991) and Hütsch (1991). Similar treatment can also be performed for polydisperse systems.

3. Data interpretation methods using multipole expansion

Possibilities of structure interpretation of SAS data depend considerably on the type of system under investigation. Monodisperse systems represent one of the most important cases when it is possible to search for the particle structure, as the SAS intensity from such systems is proportional to the single-particle scattering averaged with respect to all orientations. This average, however, leads to a significant loss of information contained in the data. With classical methods of data interpretation based on Fourier transforms, only general characteristics of particles (such as radius of gyration, volume *etc.*) can be determined directly. For obtaining more detailed information about the particle structure, modelling methods are normally used.

Multipole theory, introduced in SAS by Stuhrmann (1970*a*), gives a very elegant description of the particle scattering. Here, the particle density distribution function $\rho(\mathbf{r})$ is represented as a series

$$\rho(\mathbf{r}) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\rho_{lm}(r)\,Y_{lm}(\omega), \qquad (14)$$

where $(r,\omega) = (r,\theta,\varphi)$ are spherical coordinates,

$$\rho_{lm}(r) = \int \rho(\mathbf{r})\,Y^*_{lm}(\omega)\,\mathrm{d}\omega \qquad (15)$$

are the radial functions and $Y_{lm}(\omega)$ spherical harmonics. With this expansion the particle SAS intensity is expressed as

$$I(s) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}|A_{lm}(s)|^2, \qquad (16)$$

where

$$A_{lm}(r) = i^l(2/\pi)^{1/2}\int_0^\infty \rho_{lm}(r)j_l(sr)r^2\mathrm{d}r \qquad (17)$$

are the Hankel transforms of the radial functions [$j_l(sr)$ are spherical Bessel functions]. Here the notation $I(s)$ is used to indicate the ideal (not smeared) intensity curve.

This approach is very useful for structure interpretation. By using *a priori* information about the particle structure (*e.g.* constitution, symmetry, linear dimensions), one can impose restrictions on the multipole components. In this way additive contributions can be separated from the sum (16), and then the particle structure can be reconstructed (see *e.g.* Svergun, Feigin & Schedrin, 1982).

In particular, the relations derived from multipole theory can be used for an 'indirect' structure determination. One sees that the theory allows convenient relationships between the particle structure and SAS intensity to be established. The structure can, in principle, be parameterized in some way, followed by the fit of these parameters to the scattering intensity. Of course, the obtained relations will be non-linear, therefore general minimization methods must be used for this purpose. In the following, some applications of the multipole theory combined with optimization techniques are presented.

3.1. *Shape determination*

A general approach to shape determination has been proposed by Stuhrmann (1970*b*). Here the

structure of a homogeneous particle is described with the help of the two-dimensional angular shape function $F(\omega)$ as

$$\rho(\mathbf{r}) = \begin{cases} 1, & 0 \le r < F(\omega) \\ 0, & r \ge F(\omega). \end{cases} \quad (18)$$

This function can also be developed into the series

$$F(\omega) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} f_{lm} Y_{lm}(\omega), \quad (19)$$

where the multipole coefficients are just complex numbers

$$f_{lm} = \int F(\omega) Y^*_{lm}(\omega) \mathrm{d}\omega. \quad (20)$$

Using representation of the spherical Bessel function as a power series

$$j_l(sr) = \sum_{p=0}^{\infty} \frac{(-1)^p}{2^p p![2(l+p)+1]!!} (sr)^{l+2p} = \sum_{p=0}^{\infty} d_{lp}(sr)^{l+2p} \quad (21)$$

and substituting this series into (17), one gets (Stuhrmann, 1970*b*)

$$A_{lm}(s) = i^l (2/\pi)^{1/2} \sum_{p=0}^{\infty} \frac{d_{lp} f_{lm}^{(l+2p+3)}}{l+2p+3} s^{l+2p}, \quad (22)$$

where

$$f_{lm}^{(q)} = \int [F(\omega)]^{(q)} Y^*_{lm}(\omega) \mathrm{d}\omega. \quad (23)$$

Therefore, the SAS intensity [which is easily evaluated from the amplitudes (22) using (16)] is represented *via* a non-linear combination of the multipole coefficients $f_{lm}^{(q)}$ of the shape function. These relationships can be used to find the shape by minimizing the functional of type

$$\Phi = [\{f_{lm}\}] = \int_{s_{\min}}^{s_{\max}} [I_{\exp}(s) - I_{\mathrm{mod}}(s)]^2 W^2(s) \mathrm{d}s, \quad (24)$$

where the fitting region and the choice of the weighting function $W(s)$ depend on the resolution which one can obtain, on the experiment errors *etc.*

This method proved to be useful for a low-resolution shape determination. With only a few first terms in the series (19) (up to $l_{\max} = 3$), the number of unknowns does not exceed half a dozen (some f_{lm} can be fixed in order to define the particle orientation). Under these conditions unambiguous shape determination is possible, that is, minimization of functional (24) starting with any (reasonable) first approximation results in the same low-resolution shape function (Stuhrmann, 1970*b*).

When improving the resolution of the method, one faces the problem of evaluation of the coefficients $f_{lm}^{(q)}$ for higher l and q values. For example, taking into account harmonics up to $l_{\max} = 6$, the maximum value of q should be about 25–30. It is clear that, for a non-spherical $F(\omega)$, the function $F^{(30)}(\omega)$ will show such sharp features that no reliable integration in (23) will be possible.

This problem can be overcome if in (23) one puts $F^{(q)}(\omega) = F(\omega)F^{(q-1)}(\omega)$ and uses the following formula for the integral over three spherical harmonics (Edmonds, 1957)

$$\int Y_{lm}(\Omega) Y_{pq}(\Omega) Y_{kt}(\Omega) \mathrm{d}\Omega = [(2l+1)(2p+1)(2k+1)/4\pi]^{1/2} \times \begin{pmatrix} l & p & k \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l & p & k \\ m & q & t \end{pmatrix}, \quad (25)$$

where

$$\begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix}$$

are $3j$ Wigner symbols. This leads to the following recurrent formula

$$f_{l_1 m_1}^{(q)} = (-1)^{m_1} \sum_{l_2=0}^{\infty} \sum_{l_3=|l_2-l_1|}^{l_2+l_1} [(2l_1+1)(2l_2+1) \times (2l_3+1)/4\pi]^{1/2} \begin{pmatrix} l_1 & l_2 & l_3 \\ 0 & 0 & 0 \end{pmatrix} \times \sum_{m_2=-l_2}^{l_2} f_{l_2 m_1} f_{l_3, m_1-m_2}^{(q-1)} \begin{pmatrix} l_1 & l_2 & l_3 \\ -m_1 & m_2 & m_1-m_2 \end{pmatrix}. \quad (26)$$

This formula allows one to evaluate the coefficients $f_{lm}^{(q)}$ for $q > 1$ without any integration.

To show the possibilities of the improved algorithm, some results on the shape determination of the 50*S E.coli* ribosome subparticle are presented. The desmeared shape-scattering curve (extrapolated to zero angle and zero concentration) was obtained from the neutron contrast-variation experiments (Fig. 4). With this curve, a shape function with

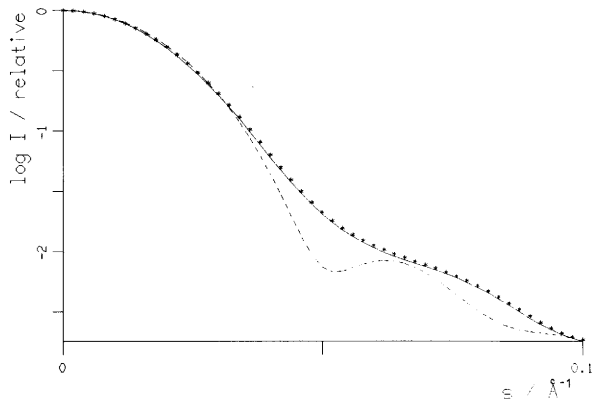

Fig. 4. Scattering curves from the 50*S* ribosome subunit and its models: desmeared neutron SAS curve (solid line); scattering from the model shown in Fig. 5(*a*) (dashed line); scattering from the model shown in Fig. 5(*b*) (asterisks).

resolution up to $l_{max} = 3$ was evaluated by Stuhrmann, Koch, Parfait, Haas, Ibel & Crichton (1977); however, the outer part of the curve ($0{\cdot}05 < s < 0{\cdot}1$ Å^{-1}) could not be fitted well at this resolution. The electron microscopic model proposed by Yonath, Leonard & Wittmann (1987) was used as an initial approximation, harmonics up to $l_{max} = 6$ being taken into account. This model is shown in Fig. 5(*a*); its scattering curve differs significantly from the neutron data (Fig. 4). For minimizing the functional (24) an interactive optimization program package *OPTIS*, developed in the Institute of Crystallography (Moscow, USSR), was used. With subsequent application of the variable matrix and downhill simplex methods of optimization (Gill, Murray & Wright, 1981), the experimental curve is neatly fitted. The resulting model, shown in Fig. 5(*b*), differs from the electron microscopy one mainly by some linear enlargement factors along axes in Cartesian coordinates. This agrees with the recently obtained results (Yonath, 1990).

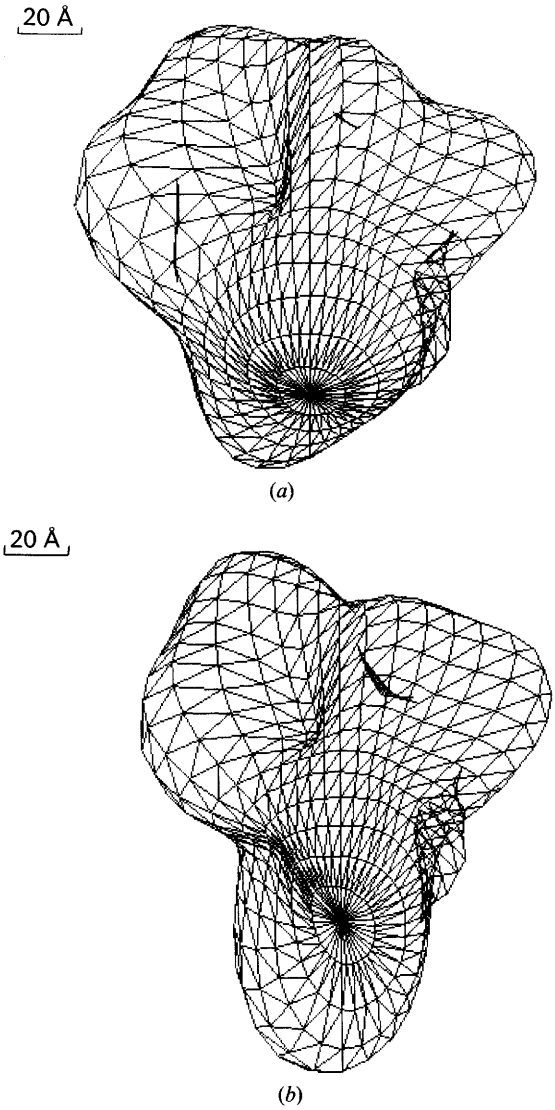

Fig. 5. Models of the shape of the 50*S E.coli* ribosome subparticle: (*a*) model proposed by Yonath *et al.* (1987) with the resolution up to the maximum order of harmonics $l_{max} = 6$; (*b*) obtained from it by the shape determination.

It should be noted that at higher resolution the shape determination is no longer unambiguous. In fact, in the example presented above it is ambiguous for the range of momentum transfer used. In order to select the unique shape function at this resolution, additional information should be used. The example given above only shows the principal possibility of handling higher harmonics allowing precise fitting of the scattering curves. More detailed descriptions of the shape-determination procedures, including model examples and discussions, will be presented elsewhere (Stuhrmann & Svergun, 1991).

3.2. *Positioning and mutual orientation of subparticles*

The spherical harmonics technique can also be applied to describe scattering from complex particles, consisting of subunits. Let us consider a system consisting of two subparticles A and B. We assume that the scattering amplitudes $A(\mathbf{s})$ and $B(\mathbf{s})$ for both particles in a given (reference) orientation are known:

$$A_0(\mathbf{s}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} A_{lm}(s)\, Y_{lm}(\Omega), \tag{27}$$

$$B_0(\mathbf{s}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} B^{(0)}_{lm}(s)\, Y_{lm}(\Omega). \tag{28}$$

Here the partial amplitudes $A_{lm}(s)$ and $B^{(0)}_{lm}(s)$ are specified in such a way that the centres of mass coincide with the origin of coordinates in real space. Arbitrary positioning of the second body with respect to the first one can be described *via* its rotation by the Euler angles α, β, γ followed by a shift of the centre of mass by a vector $\mathbf{r} = (r, \omega) = (r, \theta, \varphi)$. In terms of the scattering amplitude it will lead to

$$B(\mathbf{s}) = \exp(i\mathbf{s}.\mathbf{r})\, \Pi_{\alpha\beta\gamma}[B_0(\mathbf{s})], \tag{29}$$

where $\Pi_{\alpha\beta\gamma}$ is the rotational operator.

The SAS intensity from such a system is written as

$$I(s) = A^2(s) + B^2(s) + \langle |A(\mathbf{s})\, B(\mathbf{s})| \rangle, \tag{30}$$

where the cross term

$$\langle |A(\mathbf{s})B(\mathbf{s})| \rangle = \langle A(\mathbf{s})B^*(\mathbf{s}) + A^*(\mathbf{s})B(\mathbf{s}) \rangle$$
$$= \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \{\mathrm{Re}[A_{lm}(s)]\mathrm{Re}[B_{lm}(s)] + \mathrm{Im}[A_{lm}(s)]\mathrm{Im}[B_{lm}(s)]\}. \tag{31}$$

Here $B_{lm}(s)$ are the partial amplitudes of $B(\mathbf{s})$. We shall express them in terms of the known functions $B^{(0)}_{lm}(s)$.

Using the relation (Edmonds, 1957)

$$\exp(i\mathbf{s}.\mathbf{r}) = 4\pi \sum_{l=0}^{\infty} \sum_{m=-l}^{l} i^l j_l(sr)\, Y_{lm}(\Omega)\, Y^*_{lm}(\omega) \qquad (32)$$

one can write

$$B_{lm}(s) = \int B(\mathbf{s})\, Y^*_{lm}(\Omega)\,\mathrm{d}\Omega$$
$$= \int 4\pi \sum_{p=0}^{\infty} \sum_{q=-p}^{p} i^p j_p(sr)\, Y_{pq}(\Omega)\, Y^*_{pq}(\omega)$$
$$\times \Pi_{\alpha\beta\gamma}\left[\sum_{k=0}^{\infty} \sum_{j=-k}^{k} B^{(0)}_{kj}(s)\, Y_{kj}(\Omega)\right] Y^*_{lm}(\Omega)\,\mathrm{d}\Omega. \qquad (33)$$

Spherical harmonics are transformed under rotations as

$$\Pi_{\alpha\beta\gamma}[Y_{kj}(\Omega)] = \sum_{t=-k}^{k} D^{(k)}_{tj}(\alpha,\beta,\gamma)\, Y_{kt}(\Omega) \qquad (34)$$

where $D^{(k)}_{tj}$ are matrix elements of finite rotations (Edmonds, 1957). Substituting (34) into (33) and using (25), one gets

$$B_{lm}(s) = 4\pi(-1)^m \sum_{p=0}^{\infty} i^p j_p(sr) \sum_{q=-p}^{p} Y^*_{pq}(\omega)$$
$$\times \sum_{k=|l-p|}^{l+p} \begin{pmatrix} l & p & k \\ 0 & 0 & 0 \end{pmatrix} \sum_{j=-k}^{k} [(2l+1)(2p+1)$$
$$\times (2k+1)/4\pi]^{1/2} \begin{pmatrix} l & p & k \\ -m & q & m-q \end{pmatrix}$$
$$\times B^{(0)}_{kj}(s)\, D^{(p)}_{m-q,j}(\alpha,\beta,\gamma). \qquad (35)$$

This relationship expresses the partial amplitudes of the second subparticle as an analytical function of the parameters, describing its position and orientation. Therefore, combining (35) and (31), the cross term can be written as a function of the scattering vector and these parameters:

$$\langle |A(\mathbf{s})B(\mathbf{s})| \rangle = F(s,r,\theta,\varphi,\alpha,\beta,\gamma). \qquad (36)$$

If the cross term can be deduced from the experimental data (say in contrast-variation studies), these formulae can be used for the determination of the six parameters by minimizing the corresponding functional $\Phi = \Phi[r,\theta,\varphi,\alpha,\beta,\gamma]$. In this way positioning and mutual orientation of the two subparticles can be determined.

Equation (33) is simplified significantly in an important practical case, when the second body is spherically symmetric, $B^{(0)}(\mathbf{s}) = B^{(0)}(s)$. In this case its rotation produces no change and one gets

$$B_{lm}(s) = (4\pi)^{1/2} i^l j_l(sr)\, Y^*_{lm}(\omega)\, B^{(0)}(s). \qquad (37)$$

The approach described above may be applied, for example, as a complimentary tool to the triangulation method (Engelman & Moore, 1972). Theoretically, when locating ribosomal proteins with this approach, one needs only to measure the cross terms between the particle scattering and scattering from the marked protein. No measurements of distances between proteins are required [of course, this requires that the particle structure (shape) is known in advance].

As a model example we shall consider SAS from a complex particle consisting of the ellipsoid with half-axes $a{:}b{:}c = 50{:}70{:}100$ Å and the sphere of radius $R = 30$ Å centred at the point $(x_0,y_0,z_0) = (50,50,50)$. The set of SAS curves from such a system is presented in Fig. 6. In order to describe the sensitivity of the cross term with respect to the position of the sphere, the cross section of the functional $\Phi(r,\theta,\varphi) = \Phi(x,y,z)$ corresponding to $z = z_0$ is shown in Fig. 7(*a*). The functional has distinct minima corresponding to the four equivalent positions of the sphere. In fact, the true solution is readily obtained starting from an arbitrary point.

Of course, the situation may be much more complicated. Fig. 7(*b*) presents similar results of computer simulations performed for the case when the ribosome model obtained by the shape determination (Fig. 5*b*) was taken as the first particle. The functional shows many local minima and its minimization becomes a non-trivial problem. However, since the global minimum does correspond to the true solution, it can be found with appropriate optimization techniques.

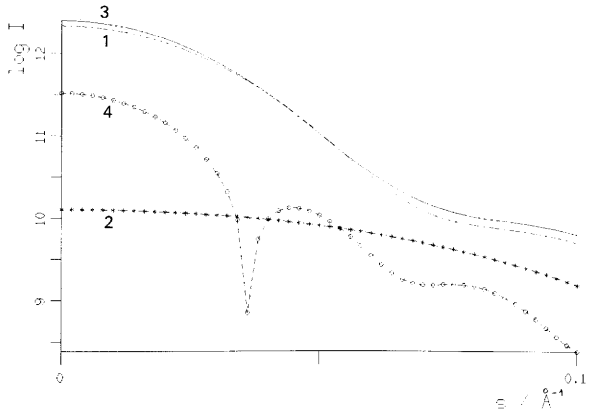

Fig. 6. Scattering from a complex particle (particle *A* ellipsoid, particle *B* sphere): (1) scattering from ellipsoid, (2) scattering from sphere, (3) total scattering, (4) cross term (negative for $s > 0{\cdot}037$ Å^{-1}). Intensities (1) and (2) are independent of the positions and orientations of the subparticles.

The model examples given above concern locating a small particle with respect to a large one. The cross term will be even more sensitive to the positioning of two particles of approximately the same dimensions. In particular, the method can be used to study conformational changes due to domain movements in proteins. Practical applications of the described approach are planned.

Concluding remarks

From the approaches presented in this survey one can see that mathematical methods play a significant role in modern SAS data analysis. They offer special possibilities in data treatment, improving significantly the quality of the data reduction. As to the use of spherical harmonics, this initially rather theoretical description now develops into a useful tool for structural studies of monodisperse systems.

The author would like to thank the GKSS Research Center for support he received. He is grateful to Dr V. Volkov for providing the source code of the minimization package *OPTIS*.

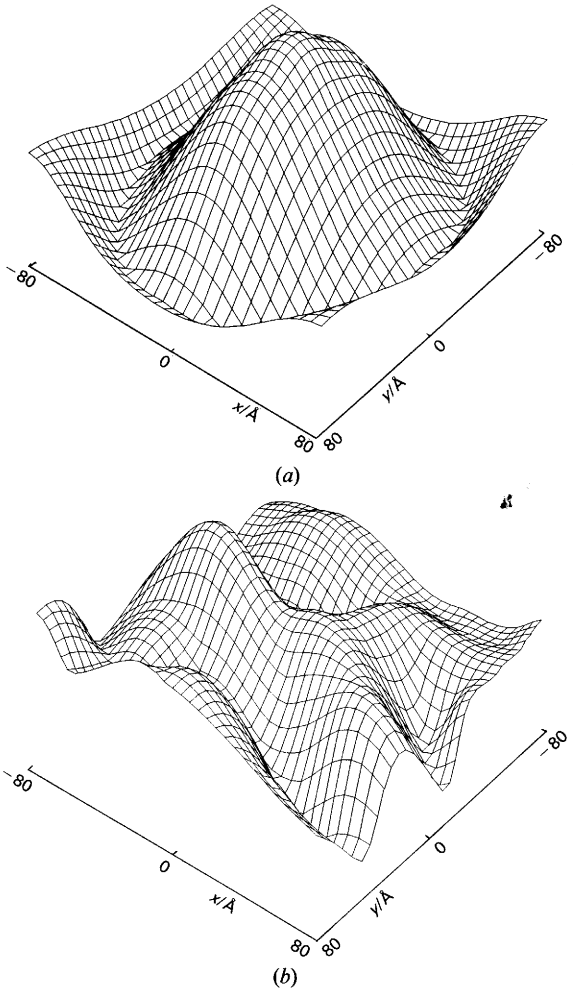

Fig. 7. Cross sections of the functional to be minimized for the two complex particles: (1) ellipsoid + sphere; (2) ribosome model (Fig. 5*b*) + sphere.

References

EDMONDS, A. R. (1957). *Angular Momentum in Quantum Mechanics*. Princeton Univ. Press.

ENGELMAN, D. & MOORE, P. B. (1972). *Proc. Natl Acad. Sci. USA*, **69**, 1997–1999.

FEIGIN, L. A. & SVERGUN, D. I. (1987). *Structure Analysis by Small-Angle X-ray and Neutron Scattering*. New York: Plenum Press.

GILL, P. E., MURRAY, W. & WRIGHT, M. H. (1981). *Practical Optimization*. London: Academic Press.

GLATTER, O. (1977). *J. Appl. Cryst.* **10**, 415–421.

HÜTSCH, M. (1991). Paper presented at 8th Int. Conf. on Small-Angle Scattering. Unpublished.

KÜHNHOLZ, O. (1991). *J. Appl. Cryst.* **24**, 811–814.

MANGANI, M., PULITI, P. & STEFANON, M. (1988). *Nucl. Instrum. Methods Phys. Res.* A**271**, 611–616.

MOORE, P. B. (1980). *J. Appl. Cryst.* **13**, 168–175.

PROVENCHER, S. W. (1982). *Comput. Phys. Commun.* **27**, 213–227, 229–242.

RONTO, GY., TOTH, K., TIMMINS, P., FEIGIN, L. A. & SVERGUN, D. I. (1991). In preparation.

STUHRMANN, H. B. (1970*a*). *Acta Cryst.* A**26**, 297–306.

STUHRMANN, H. B. (1970*b*). *Z. Phys. Chem.* (*Frankfurt am Main*), **72**, 177–184, 185–198.

STUHRMANN, H. B. (1985). *Adv. Polym. Sci.* **67**, 124–163.

STUHRMANN, H. B. & KIRSTE, R. G. (1965). *Z. Phys. Chem.* **46**, 247–250.

STUHRMANN, H. B., KOCH, M. H. J., PARFAIT, R., HAAS, J., IBEL, K. & CRICHTON, R. R. (1977). *Proc. Natl Acad. Sci. USA*, **74**, 2316–2320.

STUHRMANN, H. B. & SVERGUN, D. I. (1991). *Acta Cryst.* A**47**. In the press.

SVERGUN, D. I., FEIGIN, L. A. & SCHEDRIN, B. M. (1982). *Acta Cryst.* A**38**, 827–835.

SVERGUN, D. I., SEMENYUK, A. V. & FEIGIN, L. A. (1988). *Acta Cryst.* A**44**, 244–250.

TIKHONOV, A. N. & ARSENIN, V. YA. (1979). *Solution of Ill-Posed Problems*. New York: Wiley.

YONATH, A. (1990). Personal communication.

YONATH, A., LEONARD, K. R. & WITTMANN, H. G. (1987). *Science*, **236**, 813–816.