Approaches for improving the quality of particle size distribution reconstructions from small-angle scattering data

V V Volkov¹, A E Kryukova¹, P V Konarev¹

¹Laboratory of Reflectometry and Small-angle Scattering, A.V. Shubnikov Institute of Crystallography of Federal Scientific Research Centre "Crystallography and photonics", Russian Academy of Sciences, Leninsky pr. 59, 119333, Moscow, Russia

E-mail: volkicras@mail.ru

Abstract. Using a series of two- and three-component model data of small-angle scattering from systems of polydisperse spherical particles, the possibility of determination of the particle size distribution function by regularization methods, a direct histogram search, and as a superposition of smooth analytical functions has been investigated. The advantages and disadvantages of these approaches are considered. It is shown that their combined use allows one to improve the quality particle size distribution and find acceptable solutions even in the case of significant data noise. The choice of the appropriate minimization algorithms is also an important factor for efficient data analysis. A number of iterative algorithms based on the gradient minimization method, the simulated annealing method, their combination, and the Nelder-Mead method is tested and their efficiency is compared for the small-angle scattering data of a silicasol solution.

1. Introduction

Many modern functional materials are multicomponent systems whose properties are determined by structural features and/or inclusions in the nanoscale range. The small-angle X-ray and neutron scattering (SAXS/SANS) methods are effective and often the only methods that allow qualitatively and quantitatively to study the structure of such materials without special preparation of samples and has practically no restrictions on their types. The measured scattering intensity contains contributions from all density fluctuations in the size range from fractions to several hundred nanometres. A typical task for most of the objects under study is not to reconstruct the structure or shape of inhomogeneities ("particles"), but to determine the particle size distribution functions while approximating their shape by simple geometrical bodies (spheres, ellipsoids, cylinders, etc.). Depending on the complexity of the model used, the starting values of its parameters and the amount of noise in the experimental data, the solutions may vary significantly. Therefore, the analysis of stability and the very possibility of reconstructing particle size distributions using SAXS/SANS data becomes extremely important.

The main problem in determination of the size distributions is a poor condition number of the inverse problem: the condition number of the matrix of second derivatives of the minimized objective function can reach tens and hundreds of millions, which leads to a strong dependence of the solution not only on errors in the experimental data, but also on the search algorithm implementation, and the starting values of the parameters of the model.

Small-angle scattering experiments performed on dilute dispersions contain a large amount of information. Using the Shannon’s sampling theorem [1], the number of independent parameters N, that
can be derived from a small-angle scattering curve is of the order of $D_{\text{max}}^2 s_{\text{max}}/\pi$ [2,3], where $D_{\text{max}}$ is the largest particle size in the system and $s_{\text{max}}$ is the largest scattering vector where reliable data can be collected. With modern instruments, $N_s$ can reach the values of a few hundred. The relatively large amount of information together with the linearity of the least-squares inverse problem of the distribution determination gives a potential for describing complex size distributions. Nevertheless, a reliable method for calculations is required. For the analysis of polydisperse systems by SAXS/SANS data there are a number of algorithms, among which several main realizations can be distinguished:

(i) Programs GNOM [4] from ATSAS package [5] and GIFT [6]. The distribution search consists of solving a system of linear equations with Tikhonov’s regularization, which provides an acceptable smoothness of the distribution contour (implemented in the program GNOM).

(ii) Direct search of a distribution histogram that is not described by an analytical function (implemented in the program VOLDIS (Volkov, unpublished) and McSAS [7]) using non-linear least squares or Monte Carlo optimization procedures.

(iii) A search of a distribution profile as a superposition of smooth analytical functions, e.g., normal or Schultz-Zimm distribution [8] (the programs MIXTURE [9] and its modified version POLYMIX from ATSAS package [5], and SASFIT [10]) using non-linear least-squares methods.

All these algorithms have the advantages and drawbacks, but the assessment of the uniqueness and reliability of the restored size distribution function for multi-component polydisperse systems remains very important and actual. In this study, we propose a novel scheme for the common use of the above algorithms, which allows one to increase the “reliability” of solutions. Its performance is illustrated on a number of two-component and three-component systems of polydisperse spherical particles with well-separated or partially overlapped size distributions of individual components. The efficiency of different minimization methods is also considered and compared.

2. Algorithms for restoring the size distribution functions using small-angle scattering data

Consider a polydisperse system in which the particle size distribution can be described as the sum of single-mode distributions described by analytical expressions. We call such partial distributions "components." Let "$k$" denote the index of the component, "$K$" their total number. Then, the small-angle scattering pattern is a linear combination:

$$I(s) = \sum v_k I_k(s)$$

where $v_k$ and $I_k(s)$ are the volume fraction and the scattering intensity from the $k$-th component, respectively, $s$ is the momentum transfer $s = (4\pi/\lambda) \sin \theta$, where $2\theta$ is the scattering angle, $\lambda$ is the X-ray or neutron wavelength. The scattering intensity from the $k$-th component can be represented as

$$I_k(s) = \int D_{\text{vol}}(r) I_{0k}(s,r) dr$$

where $D_{\text{vol}}(r)$ and $I_{0k}(s,r)$ denote the single-mode volume size distribution of the particles and the normalized squared form-factor for the $k$-th component, respectively.

The use of equations (1) and (2) permits one to analyze arbitrary polydisperse systems using SAS data. A program MIXTURE [9], based on equations (1, 2), was developed to characterize mixtures of particles with different shapes (core-shell spheres, core-shell cylinders, ellipsoids, dumb-bells) containing up to $K = 10$ components. Each component is characterized by its relevant distribution parameters (e.g. the average size $R$, the half-width $dR$, the volume fraction, etc.). To solve the non-linear least-squares problem, we used here the program POLYMIX, which is based on the MIXTURE algorithm, in which the Broyden-Fletcher-Goldfarb-Shanno [11] minimization kernel is replaced by one of the most effective implementations of the Levenberg-Marquardt algorithm - the DN2FB [http://www.netlib.org/port/] procedure with simple bounds on variables, based, in turn, on NI2SOL [12] algorithm. The same minimization method is used in the distribution histogram direct search of the program VOLDIS. A key feature of VOLDIS is the search for a series of distribution profiles by minimizing the quadratic residual calculated from the difference between the experimental and model scattering intensities, the latter being calculated from a distribution profile smoothed to various degrees. From the set of solutions corresponding to various degrees of smoothness, the maximum smooth distribution is selected with an acceptable residual value. Results obtained using VOLDIS (and
to an even greater extent, McSAS [7]) are often prone to false maxima; therefore, they cannot be considered as final answers. However, the obtained information can be effectively used to assign starting values in the programs MIXTURES / POLYMIX.

3. Examples of two and three-component polydisperse systems: comparison of the algorithm efficiency

The series of two-component and three-component model small-angle scattering data of polydisperse spherical particles have been generated. The cases of separated and partially overlapped component distributions are considered. Each component has a Schulz-Zimm type of size distribution. The data have been simulated without noise and with the addition of Poisson type noise. The structural parameters of the systems are presented in Table 1.

**Table 1. Structural parameters for two- and three-component systems of polydisperse spherical particles.** Here, $R_i$ – the average radius of the $i$-th component, $dR_i$ – the width of the $i$-th component, $v_i$ – volume fraction of the $i$-th component (i=1÷3).

| Test number | Component 1 | Component 2 | Component 3 |
|-------------|-------------|-------------|-------------|
| 1           | $R_1 = 5.0$ nm | $R_2 = 12.0$ nm | --- |
|             | $dR_1 = 1.5$ nm | $dR_2 = 3.0$ nm | --- |
|             | $v_1 = 0.70$ | $v_2 = 0.30$ | --- |
| 2           | $R_1 = 6.0$ nm | $R_2 = 11.0$ nm | --- |
|             | $dR_1 = 4.0$ nm | $dR_2 = 2.0$ nm | --- |
|             | $v_1 = 0.10$ | $v_2 = 0.90$ | --- |
| 3           | $R_1 = 8.0$ nm | $R_2 = 12$ nm | --- |
|             | $dR_1 = 4.0$ nm | $dR_2 = 3.0$ nm | --- |
|             | $v_1 = 0.60$ | $v_2 = 0.40$ | --- |
| 4           | $R_1 = 4.0$ nm | $R_2 = 15$ nm | --- |
|             | $dR_1 = 0.7$ nm | $dR_2 = 1.5$ nm | --- |
|             | $v_1 = 0.20$ | $v_2 = 0.80$ | --- |
| 5           | $R_1 = 4.0$ nm | $R_2 = 8.0$ nm | $R_3 = 15.0$ nm |
|             | $dR_1 = 0.5$ nm | $dR_2 = 1.5$ nm | $dR_3 = 3.0$ nm |
|             | $v_1 = 0.70$ | $v_2 = 0.20$ | $v_3 = 0.10$ |
| 6           | $R_1 = 3.0$ nm | $R_2 = 8.0$ nm | $R_3 = 22.0$ nm |
|             | $dR_1 = 0.4$ nm | $dR_2 = 2.0$ nm | $dR_3 = 3.0$ nm |
|             | $v_1 = 0.005$ | $v_2 = 0.078$ | $v_3 = 0.917$ |
| 7           | $R_1 = 5.0$ nm | $R_2 = 8.0$ nm | $R_3 = 16.0$ nm |
|             | $dR_1 = 2.0$ nm | $dR_2 = 1.0$ nm | $dR_3 = 5.0$ nm |
|             | $v_1 = 0.10$ | $v_2 = 0.70$ | $v_3 = 0.20$ |
| 8           | $R_1 = 37.0$ nm | $R_2 = 65.0$ nm | $R_3 = 130.0$ nm |
|             | $dR_1 = 7.0$ nm | $dR_2 = 4.0$ nm | $dR_3 = 5.5$ nm |
|             | $v_1 = 0.01$ | $v_2 = 0.89$ | $v_3 = 0.10$ |

In the beginning, the size distribution functions from noiseless test data were reconstructed using the programs GNOM, VOLDIS, and POLYMIX. It was possible to find the solutions close to exact ones, using the above-mentioned algorithms for all simulated data sets. It demonstrated the high potential and efficiency of the algorithms permitting to distinguish even minor component fractions and partially overlapped contributions from different components in the system.

However, the situation becomes more complicated when the noise is added to the data. In this case, the useful angular range of the data was estimated by the program SHANUM [13] according to Shannon’s sampling theorem [1]. The corresponding data are presented in Figures 1 and 2 (left panels,
dots with error bars). The fits from GNOM, VOLDIS, and POLYMIX are shown by dashed, dashed-dotted and solid lines, respectively. The restored size distribution functions are displayed in the right panel of Figures 1 and 2. The exact solutions coincide with the POLYMIX fits (solid lines).

Figure 1. Two component polydisperse data sets. (A) Test 1 (B) Test 2 (C) Test 3 (D) Test 4 (as defined in Table 1). Left panel: Simulated data with added noise (dots with error bars) and the fits from GNOM, VOLDIS, and POLYMIX (dashed, dashed-dotted and solid lines, respectively). Right panel: the restored size distribution functions.

Figure 2. Three-component polydisperse data sets. (A) Test 5 (B) Test 6 (C) Test 7 (D) Test 8 (as defined in Table 1). Left panel: Simulated data with added noise (dots with error bars) and the fits from GNOM, VOLDIS, and POLYMIX (dashed, dashed-dotted and solid lines, respectively). Right panel: the restored size distribution functions.

As can be seen from Figures 1 and 2, the shapes of the size distribution functions restored by the program GNOM appear to be significantly different from the exact solutions and contain artifacts in the form of oscillations due to the termination effects. The reconstructed size distributions by the program VOLDIS also contain differences from the exact solutions, especially in the small size
regions. Only the program POLYMIX permits to restore ‘true’ solutions in all cases presented, but its efficiency strongly depends on the starting approximation. As was shown earlier in [14, 15], the range of the starting values of the parameters from which the ‘true’ solution can be restored depends both on the relative contribution from mixture components and on the level of noise contained in the data.

4. Scheme for the common use of the algorithms

We propose the following scheme for the combined use of the algorithms:

(Step 1): run the program VOLDIS for the calculation of the size distribution histogram and determine mean radii of corresponding particles;

(Step 2): use the maximum size estimated by VOLDIS, run the program GNOM to find the preliminary shape of the distribution function;

(Step 3): find the size distribution profile using the programs MIXTURE / POLYMIX setting starting values of distribution parameters obtained at the steps (1) and (2).

The proposed scheme is designed to improve the stability and reliability of the reconstructed size distribution functions. At the final step of the scheme, it is possible to additionally combine local and global minimization methods, which will further expand the range of convergence to the “true” solution, as was shown in [16].

The efficiency of different minimization methods using MIXTURE / POLYMIX target function is compared to the experimental data set from a silicasol solution representing a mixture of spherical particles with the average sizes $R_1=5.9$ nm and $R_2=13.9$ nm. The following schemes are tested: quasi-Newton method due to Broyden-Fletcher-Goldfarb-Shanno (BFGS) [11], simulated annealing algorithm (SA) [17], simplex method of Nelder-Mead (NM) [18] and the combination of BFGS and SA methods. As can be seen from Figure 3 the area of convergence to the “true” solution is increasing in the following sequence: (NM) – (SA) – (BFGS) – (BFGS+SA).

Figure 3. The contour plots of solution stability range for the two-component mixture of silicasol particles in solution depending on the pair of parameters ($R_1$, $R_2$). The interception of the dotted lines corresponds to the theoretical (“true”) solution. If the restored parameters are located within 5% in the vicinity of ‘true’ solution, then the solution is treated as successful (white areas), otherwise, the solution is treated as false (black areas). The employed minimization schemes: (A) NM, (B) SA, (C) BFGS, (D) BFGS+SA.

5. Conclusions

A number of algorithms for retrieving the particle size distributions from small-angle scattering data sets were tested on a series of two- and three-component artificial systems of polydisperse spherical particles. It was demonstrated that all these algorithms are able to find the ‘true’ solutions for noiseless data sets, however, they have significant problems on the real-type data with the added noise. Based on the simulation results a novel scheme for the common use of the algorithms is proposed that improves the quality of the particle size distribution reconstructions from small-angle scattering data.
The joint use of the minimization methods (such as BFGS and SA algorithms) can additionally expand the ranges of convergence to the ‘true’ solution and, thus, increase the reliability of the analysis results.

**Acknowledgements**
This work was supported by the Ministry of Science and Higher Education of the Russian Federation within the State assignment FSRC ‘Crystallography and Photonics’ RAS and by the Russian Foundation for Basic Research (project № 19-32-90190). SAXS measurements were performed using the equipment of the Shared Research Center FSRC “Crystallography and Photonics” RAS and was supported by the Russian Ministry of Education and Science (project RFMEFI62119X0035).

**References**

[1] Shannon C E and Weaver W *The Mathematical Theory of Communication* 1949 (Urbana: University of Illinois Press)

[2] Damaschun G, Mueller J J and Puerschel H V 1968 *Monatsh. Chem.* **99** 2343–48

[3] Taupin D and Luzzati V 1982 *J. Appl. Cryst.* **15** 289–300

[4] Svergun D I 1992 *J. Appl. Cryst.* **25** 495-503

[5] Franke D, Petoukhov M V, Konarev P V, Pankovich A, Tuukkanen A, Mertens H D T, Kikhney A G, Hajizadeh N R, Franklin J M, Jeffries C M and Svergun D I 2017 *J. Appl. Cryst.* **50** 1212-25

[6] Glatter O 1980 *J. Appl. Cryst.* **13** 7–11

[7] Bressler I, Pauw B R and Thünemann A F 2015 *J. Appl. Cryst.* **48** 962-69

[8] Schulz G. V. 1935 *Z. Phys. Chem. Abt. B30*, 379-98

[9] Konarev P V, Volkov V V, Sokolova A V, Koch M H J and Svergun D I 2003 *J. Appl. Cryst.* **36** 1277-82

[10] Breßler I, Kohlbrecher J and Thünemann A F 2015 *J. Appl. Cryst.* **48** 1587-98

[11] Gill P.E., Murray W., Wright M.H. *Practical Optimization* 1981 (London: Academic Press)

[12] More J J *The Levenberg-Marquardt Algorithm, Implementation and Theory. Lecture Notes in Mathematics* 1978 (Berlin: Springer-Verlag)

[13] Konarev P V and Svergun D I 2015 *IUCr J.* **2** 352-360.

[14] Kryukova A E, Konarev P V and Volkov V V 2017 *J. Phys. Conf. Ser.* **941** 012069

[15] Kryukova A E, Konarev P V and Volkov V V 2018 *Cryst. Reports** **63** 26-31

[16] Kryukova A E, Konarev P V, Volkov V V and Asadchikov V E 2019 *J. Mol. Liq.* **283** 221-24

[17] Kirkpatrick S, Gelatt C D and Vecchi M P 1983 *Science* **220** 671–80

[18] Nelder J A and Mead R 1965 *Computer Journal* **7** 308-13