

Комбинированная схема восстановления функции распределения частиц по размерам с использованием данных малоуглового рассеяния

В. В. Волков¹⁾, П. В. Конарев, А. Е. Крюкова

Институт кристаллографии им. А. В. Шубникова,
Федеральный научно-исследовательский центр “Кристаллография и фотоника” РАН, 119333 Москва, Россия

Поступила в редакцию 7 октября 2020 г.

После переработки 7 октября 2020 г.

Принята к публикации 7 октября 2020 г.

Для анализа полидисперсных систем в наноразмерном диапазоне существует ряд алгоритмов, позволяющих определять профиль распределения частиц по размерам из данных малоуглового рентгеновского и нейтронного рассеяния. Соответствующая задача наименьших квадратов плохо обусловлена, что может приводить к сильной зависимости решения от стартового приближения и параметров алгоритмов поиска. Процедура, основанная на решении системы линейных уравнений с регуляризацией, дает стабильное решение, но не свободное от артефактов. Решения, предлагаемые другими алгоритмами, которые ищут функцию распределения как в аналитическом виде, так и в виде непараметрической гистограммы, могут зависеть от параметров поиска и начального приближения. В работе предложена комбинированная схема использования этих алгоритмов, позволяющая повысить устойчивость решения и, как следствие, его надежность.

DOI: 10.31857/S1234567820210107

Введение. Свойства многих современных материалов определяются структурными особенностями в наноразмерном диапазоне, поэтому для их исследования часто применяют методы малоуглового рентгеновского и нейтронного рассеяния (МУРР и МУРН, соответственно), которые позволяют качественно и количественно исследовать строение вещества без специальной подготовки образцов в диапазоне разрешений от долей нанометра до нескольких микрон. Анализ таких систем представляет собой плохо обусловленную задачу, что может приводить к большим разбросам решений. Так, число обусловленности матрицы вторых производных минимизируемой целевой функции может достигать десятков и сотен миллионов, что приводит к сильной зависимости решения не только от ошибок в исходных данных, но и от реализации алгоритма поиска, и от величин стартовых значений параметров модели.

Для анализа полидисперсных систем по малоугловым данным существует ряд алгоритмов, среди которых можно выделить несколько основных:

(1) методы линейных наименьших квадратов, в том числе с регуляризацией решения по Тихонову (программа GNOM [1] из пакета ATSAS [2] и программа GIFT [3]);

(2) прямой поиск распределения частиц по размерам в виде гистограммы методами нелинейных наименьших квадратов (программа VOLDIS, неопубл.) и случайного поиска McSAS [4];

(3) постулирование вида распределения в аналитическом виде (например, нормального распределения или распределения Шульца [5]) и проведение приближения данных методами нелинейных наименьших квадратов (программа MIXTURE (и ее модифицированный вариант POLYMIX) [6] и программа SASFIT [7]).

Каждый из этих алгоритмов обладает своими преимуществами и недостатками, и использование только одного из них не всегда позволяет получать решения, свободные от артефактов. В данной работе нами предложена новая схема совместного использования алгоритмов, позволяющая повысить надежность восстановления функции распределения частиц по размерам. Работа схемы проиллюстрирована на примерах двухкомпонентных и трехкомпонентных систем полидисперсных сферических частиц с разнесенными и частично перекрывающимися распределениями компонент.

Алгоритмы восстановления функции распределения частиц по размерам по данным малоуглового рассеяния. Для вычисления функции распределения по размерам частиц в полидис-

¹⁾e-mail: volkicras@mail.ru

персных системах по данным интенсивности рассеяния $I(s)$, программы GNOM [1] и GIFT [3] решают интегральное уравнение

$$I(s) = \int_{R_{\min}}^{R_{\max}} D_V(r) I_0(s, r) dr, \quad (1)$$

где $D_V(r)$ – искомая функция распределения, определенная в диапазоне размеров $[R_{\min}, R_{\max}]$, $I_0(s, r)$ – заданный форм-фактор частицы (в данной работе мы использовали сферы), $s = (4\pi/\lambda) \sin \theta$ – модуль вектор рассеяния, 2θ – угол рассеяния, λ – длина волны. Данное интегральное уравнение, представленное в конечных разностях в виде системы линейных уравнений с матрицей коэффициентов, составленной из интенсивностей рассеяния сферическими частицами, рассчитанными на экспериментальной сетке угловых отсчетов интенсивности и искусственной сетке радиусов от R_{\min} , (обычно от 1 Å) до предполагаемого максимального значения R_{\max} , представляет собой плохо обусловленную задачу. Для получения устойчивых решений, имеющих физический смысл, в программе GNOM применена регуляризация решения по Тихонову [8], т.е. минимизируется функционал, в который входит как квадрат невязки между экспериментальной и рассчитанной кривой рассеяния, так и квадрат нормы первой производной функции распределения, умноженный на параметр регуляризации и обеспечивающий гладкость контура [1]. Решение чувствительно к эффектам обрыва $I(s)$ и может содержать осцилляции, существенно искажающие форму распределения. Параметр регуляризации в данном подходе фактически представляет собой степень сглаживания решения, позволяющий повысить его устойчивость, но требует подбора области определения $[R_{\min}, R_{\max}]$ функции $D_V(r)$.

Два других алгоритма используют методы нелинейной минимизации, причем в одном случае функцию распределения непосредственно ищут в виде гистограммы (программы McSAS [4] и VOLDIS), в другом случае функция распределения задается в аналитическом виде (программа SASFIT [7]) или комбинации аналитических выражений (MIXTURE/POLYMIX [6]).

В последнем случае интенсивность рассеяния представляют в виде линейной комбинации вкладов рассеяния от K компонент, при этом каждая компонента есть система частиц заданной формы, распределение которых описывается одним аналитическим выражением с учетом, при необходимости, межчастичной интерференции:

$$I(s) = \sum_{k=1}^K v_k I_k(s), \quad (2)$$

где v_k – относительная объемная доля k -ого компонента с интенсивностью рассеяния $I_k(s)$, которая, в свою очередь, зависит от вида (объемной) функции распределения $D_{V_k}(r)$ и заданного квадрата форм-фактора частицы $I_{0_k}(s, r)$:

$$I_k(s) = \int_{R_{\min}}^{R_{\max}} D_{V_k}(r) I_{0_k}(s, r) dr, \quad (3)$$

где r – размер частицы. Тогда результирующее распределение есть линейная комбинация парциальных функций:

$$D_V(r) = \sum_{k=1}^K v_k D_{V_k}(r). \quad (4)$$

Решение задачи состоит в определении параметров аналитических распределений и их относительных вкладов в сумму (2) путем минимизации квадратичного отклонения теоретической интенсивности от экспериментальной.

В программе MIXTURE для поиска решения используется метод переменной метрики в варианте Бройдена–Флетчера–Голдфарба–Шанно [9] с простыми ограничениями на параметры, в программах POLYMIX и VOLDIS реализован модифицированный вариант алгоритма Левенберга–Марквардта [10], тогда как в программе McSAS применен относительно медленно сходящийся метод Монте-Карло. Использование стохастических методов, хотя и обладающих свойствами поиска глобального экстремума, в данном случае нам представляется необоснованным, так как нестабильность поиска непараметрической гистограммы распределения вызвана плохой обусловленностью задачи и ограниченностью точности машинных вычислений, а не ее неоднозначностью.

Сравнение эффективности восстановления функции распределения для двух и трехкомпонентных систем полидисперсных сферических частиц. В данной работе мы рассмотрели случаи двухкомпонентных и трехкомпонентных систем полидисперсных сферических частиц с разнесенными и частично перекрывающимися распределениями компонентов. Для этих систем были смоделированы наборы данных малоуглового рассеяния как без шума, так и с добавлением шума, подчиняющегося Пуассоновскому распределению. Структурные параметры систем приведены в табл. 1.

По незашумленным модельным данным все три алгоритма (программы GNOM, VOLDIS и

Таблица 1. Структурные параметры двух- и трехкомпонентных систем полидисперсных сферических частиц. R_i – средний радиус сфер i -компоненты, dR_i – ширина распределения i -компонента, v_i – объемная доля i -компоненты ($i = 1 \div 3$)

Номер системы	Компонент 1	Компонент 2	Компонент 3
1	$R_1 = 5.0 \text{ нм}$ $dR_1 = 1.5 \text{ нм}$ $v_1 = 0.70$	$R_2 = 12.0 \text{ нм}$ $dR_2 = 3.0 \text{ нм}$ $v_2 = 0.30$	—
2	$R_1 = 6.0 \text{ нм}$ $dR_1 = 4.0 \text{ нм}$ $v_1 = 0.10$	$R_2 = 11.0 \text{ нм}$ $dR_2 = 2.0 \text{ нм}$ $v_2 = 0.90$	—
3	$R_1 = 4.0 \text{ нм}$ $dR_1 = 0.7 \text{ нм}$ $v_1 = 0.20$	$R_2 = 15 \text{ нм}$ $dR_2 = 1.5 \text{ нм}$ $v_2 = 0.80$	—
4	$R_1 = 4.0 \text{ нм}$ $dR_1 = 0.5 \text{ нм}$ $v_1 = 0.70$	$R_2 = 8.0 \text{ нм}$ $dR_2 = 1.5 \text{ нм}$ $v_2 = 0.20$	$R_3 = 15.0 \text{ нм}$ $dR_3 = 3.0 \text{ нм}$ $v_3 = 0.10$
5	$R_1 = 3.0 \text{ нм}$ $dR_1 = 0.4 \text{ нм}$ $v_1 = 0.005$	$R_2 = 8.0 \text{ нм}$ $dR_2 = 2.0 \text{ нм}$ $v_2 = 0.078$	$R_3 = 22.0 \text{ нм}$ $dR_3 = 3.0 \text{ нм}$ $v_3 = 0.917$
6	$R_1 = 37.0 \text{ нм}$ $dR_1 = 7.0 \text{ нм}$ $v_1 = 0.01$	$R_2 = 65.0 \text{ нм}$ $dR_2 = 4.0 \text{ нм}$ $v_2 = 0.89$	$R_3 = 130.0 \text{ нм}$ $dR_3 = 5.5 \text{ нм}$ $v_3 = 0.10$

MIXTURE/POLYMIX) восстанавливают решение с удовлетворительной для практики точностью (относительная ошибка не превышает 0.1–0.2%, на графиках отклонение не видно, и поэтому рисунки мы не приводим). Естественно, по зашумленным кривым рассеяния, показанным на рис. 1–2, удастся найти только приближенное распределение. При этом форма функции распределения, восстанавливаемая программой GNOM (синие пунктирные кривые на рисунках), оказывается наиболее отличающейся от точного решения (сплошные красные линии на рисунках). Если положения пиков распределений определяются достаточно точно (отклонения не превышают 5%), то ширины распределений оказываются намного больше заданных при построении моделей. Кроме того, найденные распределения имеют артефакты в виде осцилляций из-за эффекта обрыва функции интенсивности рассеяния. Регуляризация, используемая в программе GNOM, хотя и позволяет находить достаточно гладкие решения (требование гладкости и приводит к уширению пиков) с малой амплитудой артефактов, тем не менее может неверно передавать форму распределения в области начальных (малых) размеров из-за чрезмерного ограничения первой производной по контуру.

Результаты программы непараметрического поиска гистограммы распределения VOLDIS для зашумленных данных (кривые на рисунках показаны

точками) также содержат отличия от точного решения (сплошные линии), причем особенно неустойчивыми также оказываются области малых размеров (особенно хорошо это видно на примере рис. 2(5)). Этот эффект связан с тем, что парциальные кривые рассеяния от частиц близких размеров очень похожи, и программа, увеличивая вклад частиц какого-либо радиуса, компенсирует чрезмерный рост интенсивности рассеяния уменьшением вклада от соседних частиц с близкими размерами. При малых же размерах сходство форм кривых рассеяния (например, по критерию корреляции) намного больше, чем в области больших размеров. Это связано с тем, что рассеяние от малоразмерных частиц в экспериментальной области углов представляет собой монотонные медленно спадающие функции, тогда как при больших размерах парциальные кривые содержат особенности в виде максимумов (по крайней мере, на производных), имеющих разные угловые положения.

Программа POLYMIX (аналогично MIXTURE) позволяет находить решения, практически совпадающие с точными во всех представленных случаях (сплошные красные кривые на рисунках, визуально не отличимые от точных контуров) при старте с приближений, близких к решениям, полученным программами GNOM и VOLDIS. Однако, решения могут содержать и значительные отличия, если стартовое приближение далеко от истинного. Как было показано ранее в работах [11, 12], диапазон области

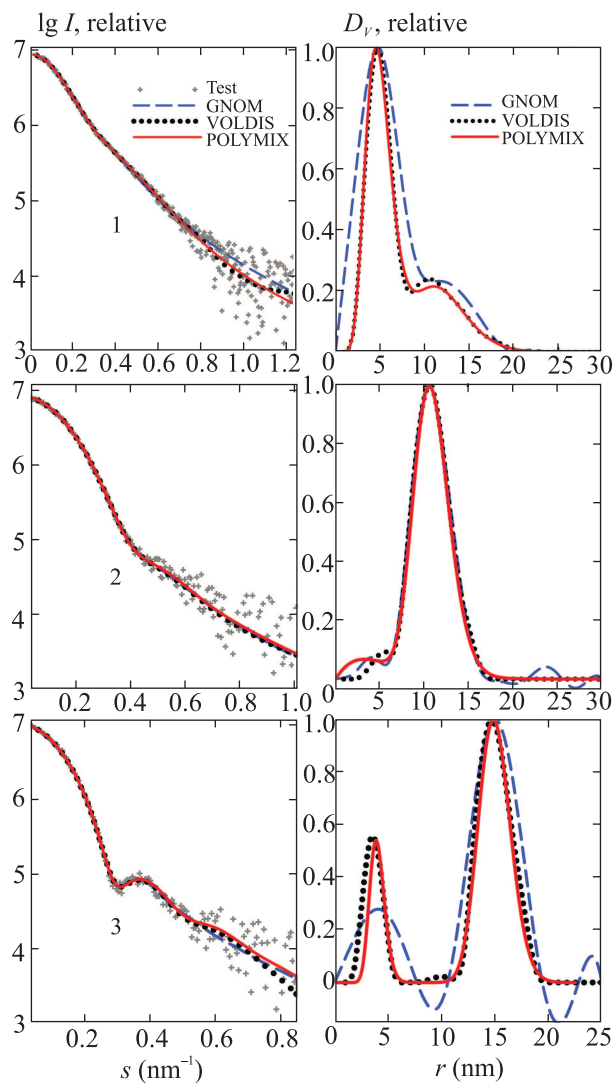


Рис. 1. (Цветной онлайн) Столбец слева: малоугловые данные от двухкомпонентной системы полидисперсных сферических частиц, номера строк соответствуют моделям 1–3 из табл. 1. Строки: крестики – зашумленные модельные данные, синяя пунктирная кривая – решение GNOM, черные точки – решение VOLDIS; сплошная красная кривая – решение POLYMIX, полученное с использованием стартовых параметров, оцененных по решениям GNOM и VOLDIS. Рассеяние от точной модели визуально совпадает с решением POLYMIX. Столбец справа: контуры полученных распределений по радиусам частиц, обозначения кривых те же

стартовых значений параметров распределений, с которых возможно восстановление правильного решения, зависит от многих факторов, в частности, от относительного вклада параметров компонентов смеси в общую кривую рассеяния и от уровня и вида шума в данных. Поэтому естественной кажется мысль ис-

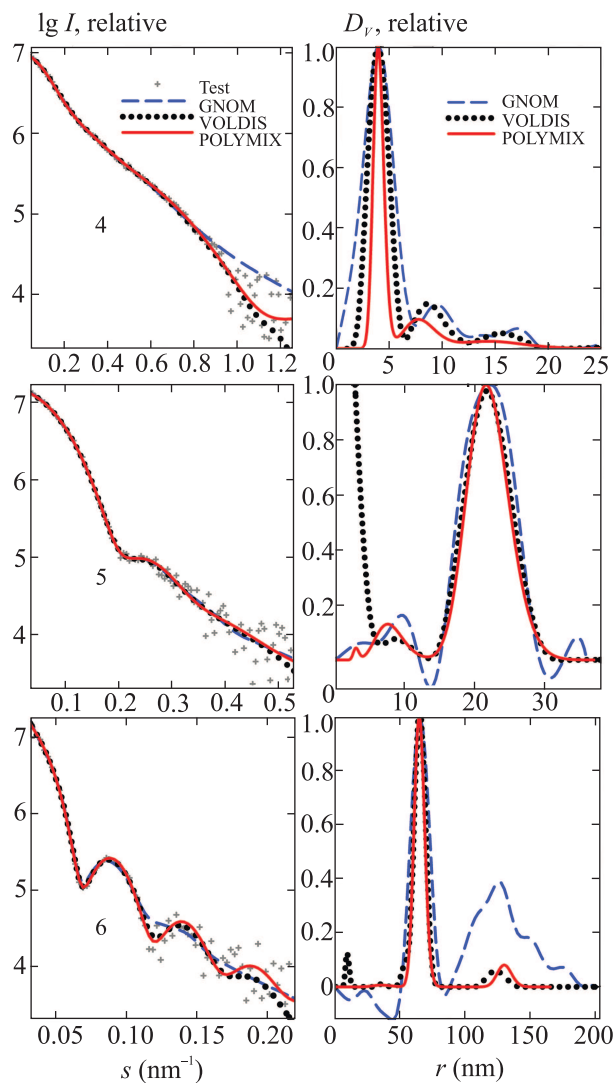


Рис. 2. (Цветной онлайн) Данные от трехкомпонентной системы полидисперсных сферических частиц, номера строк соответствуют моделям 4–6 из табл. 1. Обозначения и цвета аналогичны рис. 1

пользовать в качестве стартовых параметры, оцененные по решениям GNOM и MIXTURE/POLYMIX.

Схема комбинированного использования алгоритмов восстановления функции распределения. Основываясь на сравнении полученных результатов восстановления функции распределения частиц по размерам разными алгоритмами, мы предлагаем следующую схему их комбинированного использования:

1) использовать программу VOLDIS для оценки гистограммы распределения, числа компонентов (т.е. вкладов, представленных унимодальными распределениями) и среднего радиуса частиц в каждом компоненте. При этом особое внимание следует обратить

на область углов рассеяния от 0 до начального экспериментального. Теоретическая кривая интенсивности должна монотонно экстраполировать экспериментальный контур, в противном случае найденное решение будет содержать артефакты в виде значительных вкладов больших размеров частиц. Для того чтобы этого избежать, необходимо провести повторный поиск, искусственно уменьшая R_{\max} ;

2) использовать программу GNOM для нахождения предварительной оценки функции распределения, указывая в ней максимальный размер частиц R_{\max} , подобранный на предыдущем шаге. Цель данного численного эксперимента является получение гладкого контура, уточнения числа и параметров компонентов;

3) уточнить параметры распределения с помощью программ POLYMIX/MIXTURE, используя в качестве стартового приближения параметры распределений, оцененные на предыдущих шагах.

Предложенная схема позволяет улучшить устойчивость восстановления функции распределения частиц по размерам с использованием малоугловых данных. На третьем шаге схемы можно дополнительно проводить комбинирование локальных и глобальных минимизационных методов, что позволит еще больше расширить диапазон сходимости к правильному решению и, таким образом, повысить достоверность определения распределений по размерам, как было показано в работе [13].

Работа выполнена при поддержке Министерства науки и высшего образования в рамках выполнения работ по Государственному заданию ФНИЦ “Кристаллография и фотоника” РАН, а также

при частичной финансовой поддержке Российского фонда фундаментальных исследований (грант # 19-32-90190).

1. D. I. Svergun, *J. Appl. Crystallogr.* **25**, 495 (1992).
2. D. Franke, M. V. Petoukhov, P. V. Konarev, A. Pankovich, A. Tuukkanen, H. D. T. Mertens, A. G. Kikhney, N. R. Hajizadeh, J. M. Franklin, C. M. Jeffries, and D. I. Svergun, *J. Appl. Crystallogr.* **50**, 1212 (2017).
3. O. Glatter, *J. Appl. Crystallogr.* **12**, 166 (1979).
4. I. Bressler, B. R. Pauw, and A. F. Thuenemann, *J. Appl. Crystallogr.* **48**, 962 (2015).
5. G. V. Schulz, *Z. Phys. Chem. Abt. B* **30**, 379 (1935).
6. P. V. Konarev, V. V. Volkov, A. V. Sokolova, M. H. J. Koch, and D. I. Svergun, *J. Appl. Crystallogr.* **36**, 1277 (2003).
7. I. Bressler, J. Kohlbrecher, and A. F. Thuenemann, *J. Appl. Crystallogr.* **48**, 1587 (2015).
8. A. N. Tikhonov and V. Ya. Arsenin, *Solution of ill Posed Problems*, Wiley, N.Y. (1977), 258 p.
9. P. E. Gill, W. Murray, and M. H. Wright, *Practical Optimization*, Academic Press, London (1981), 401 p.
10. J. J. More, *The Levenberg-Marquardt Algorithm, Implementation and Theory. Lecture Notes in Mathematics*, ed. by G. A. Watson, Springer-Verlag, Berlin (1978), v. 630, p. 105.
11. A. E. Kryukova, P. V. Konarev, and V. V. Volkov, *Crystallogr. Rep.* **63**, 26 (2018).
12. A. E. Kryukova, A. S. Kozlova, P. V. Konarev, V. V. Volkov, and V. E. Asadchikov, *Crystallogr. Rep.* **63**, 531 (2018).
13. A. E. Kryukova, P. V. Konarev, V. V. Volkov, and V. E. Asadchikov, *J. Mol. Liq.* **283**, 221 (2019).