Journal of Quantitative Spectroscopy & Radiative Transfer 131 (2013) 202–214

Contents lists available at ScienceDirect

Journal of Quantitative Spectroscopy & Radiative Transfer

journal homepage: www.elsevier.com/locate/jqsrt

An optimization method for solving the inverse Mie problem based on adaptive algorithm for construction of interpolating database

Konstantin V. Gilev [a,b], Maxim A. Yurkin [a,b], Gleb V. Dyatlov [b,c], Andrei V. Chernyshev [a,b], Valeri P. Maltsev [a,b,*]

[a] Institute of Chemical Kinetics and Combustion, Institutskaya 3, Novosibirsk 630090, Russia
[b] Novosibirsk State University, Pirogova 2, Novosibirsk 630090, Russia
[c] Sobolev Institute of Mathematics, Koptyug Avenue 4, Novosibirsk 630090, Russia

ARTICLE INFO

Article history:
Received 20 June 2012
Received in revised form
30 July 2012
Accepted 2 August 2012
Available online 9 August 2012

Keywords:
Inverse light scattering problem
Global optimization
Lookup database

ABSTRACT

We introduce a numerical solution of the inverse light-scattering problem for a single non-absorbing spherical particle. The solution is implemented by global optimization at preliminary constructed database of light-scattering patterns. We propose an adaptive method for database construction, which aims both at providing satisfactory local accuracy and at avoiding large errors of the inverse map. Several databases were constructed varying the required accuracy of solution of the inverse problem and parameters used to characterize a sphere. We tested accuracy of the method on synthetic data for spheres with and without noise, on synthetic data for slightly prolate and oblate spheroids, and on experimental data of polystyrene microspheres measured with a scanning flow cytometer. The constructed databases have shown appropriate results in determination of the size and refractive index of a sphere from the angle-resolved light scattering with given accuracy.

1. Introduction

The problem of single particle characterization from light scattering arises in different fields, including astronomy, remote sensing, and analysis of aerosols and emulsions [1]. Optical methods play an important role in solution of this problem. These methods in analysis of individual particles, especially biological cells, can be divided into the following categories: (1) probe-field methods forming an electromagnetic field in smallest volume (confocal microscopy) and (2) full-field methods based on a detailed analysis of light scattered by a particle. Most powerful method that may be considered to belong to the second category is flow cytometry. A flow cytometer allows the measurements of light scattering from a single particle in fixed solid angles [2] or in a form of an entire angle-resolved light-scattering pattern (LSP) [3]. Recently Strokotov et al [4] modernized the flow cytometer for measurement of regular and polarized LSPs of individual particles simultaneously. New facilities of flow cytometry in the measurement of ample light-scattering data force us to develop new approaches in a solution of the inverse light-scattering (ILS) problem. Generally flow cytometers measure individual particles with a rate of 1000 particles per second that forms an extra requirement for the solution of the ILS problem—the fast determination of particle parameters (characteristics) from LSPs [5].

At present a practical analytical solution of the ILS problem for a single particle is unavailable. The most successful attempt is the result for spheres that allows determination of its parameters from the Mie scattering amplitudes, which include both the amplitude and phase components of a scattering field, measured over the

* Corresponding author at: Institute of Chemical Kinetics and Combustion, Institutskaya 3, Novosibirsk 630090, Russia. Tel.: +7 383 333 32 40; fax: +7 383 330 73 50.
E-mail address: maltsev@kinetics.nsc.ru (V.P. Maltsev).

0022-4073/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jqsrt.2012.08.001

whole angular interval [0°,180°] [6]. However, such complete measurements cannot be performed with flow cytometry.

The alternative approaches are to solve the inverse problem numerically. Spectral decomposition methods [7–9] allow one to determine the size of a sphere from the Gegenbauer of Fourier spectrum of the LSP. However, it provides no information about the refractive index. A parametric solution [10] is based on dependence of particle parameters on the phase-shift and fringe pitch of LSPs. It provides satisfactory accuracy only for spheres and is sensitive to the experimental noise [5]. Neural networks were also used to characterize single spheres [11–13]. This method requires only a learning sample of calculated LSPs and it was successfully used on experimental data of polystyrene spheres and spherized red blood cells [11]. However, application of neural networks is still an art, which requires fine-tuning of internal parameters for the particular scattering problem. That is why it has not been yet applied to non-spherical particles.

The most general approach for characterization of particles with relatively simple shape is optimization, i.e. direct fitting of the experimental signal to the computed LSPs. Oscillatory nature of the LSPs with strong dependence on particle parameters calls for robust global optimization techniques. In particular, stochastic global optimization techniques [14,15] and the DiRect method [15,16] were applied to spheres.

A similar problem, characterization of multi-layered concentric spheres, was approached by multi-start Levenberg-Marquardt [17] and the DiRect [18] methods. However, optimization methods come at a great computational cost which becomes unbearable for non-spherical particles, such as blood platelets and red blood cells, due to increasing number of parameters to explore and much slower computation of the LSP. New optimization methods have to allow determination of cell characteristics in reasonable time to be included in common routine of hematological analysis. Therefore, there is demand for acceleration of the optimization methods using some kind of preliminary exploration of a particular scattering problem. Since the problem of single particle characterization usually has to be solved multiple times in a row (in flow cytometry applications at least several thousand times), one-time investment of large computational time should be acceptable.

Conceptually the simplest approach for such acceleration is to calculate a large database of LSPs and to solve the ILS problem by the nearest-neighbor interpolation. In particular, this method was applied to spheres [19], spheroids [20] and biconcave disks [21,22]. However, in all cases of non-spherical particles the accuracy of the solution was not reliably assessed. The main problem of this approach, not addressed in the mentioned papers, is finding the optimum structure and size of the database. One should strike a compromise between computational time, both for the construction of the database and for each interpolation, and the accuracy of solution of the inverse problem. While a 'large enough' database may work for spheres, a careful compromise is required for practical feasibility of characterization of non-spherical particles. It is complicated by variability of sensitivity of the LSP to the particle parameters over their domain, calling for variable density of the database. Moreover, there may be a certain threshold in the dependence of interpolation error on grid density due to oscillatory nature of the LSP. Above this threshold the error continuously decreases with increasing density, as typical for interpolation. But below the threshold the errors abruptly increase to the values comparable with the size of the whole parameter domain (see Section 3). A related question is how large experimental noise is acceptable for a particular database.

Optimality of the interpolation database was previously addressed for a different problem, detecting defects in 2D systems [23,24]. However, the proposed adaptive algorithm of database construction focused only on interpolation error within the elementary cell, thus implicitly assuming that database density is large enough. There are also sampling methods based on nearest-neighbor interpolation [25,26], which share similar ideas. However, they construct a set of samples for a particular experimental signal instead of an universal database.

In this paper we propose a method to adaptively construct a database to characterize single particles from measured LSPs. This method aims at providing satisfactory uniqueness and local accuracy, avoiding large errors of the inverse map and limitation of the database constructed. After a general description of the problem and the method in Sections 2 and 3 respectively, we consider the simplest model, a homogeneous sphere, for a practical example in Section 4. In a recent paper [27] the inverse Mie problem is solved by constructing a set of starting points for the gradient-based optimization. This approach is based on detailed and rigorous analysis of the map, implementing LSP calculation, and its derivative, which guarantees perfect accuracy at least for the noise-free data. In contrast, this paper is based on the implicit assumption that a relatively small number of LSPs is the most complete information about the map one can obtain. From the current practical viewpoint this assumption seems to be valid for all non-spherical particles. Therefore, the method proposed in this paper should be applicable to many particle models described by a few parameters, but with no complete rigor due to the discreteness of the analysis.

2. Problem statement

The direct scattering problem under consideration consists in determination of the LSP from given parameters of a particle. It is implemented by a map f: $X \to Y$, where $X \subset \mathbf{R}^p$ is a domain of parameters of a particle, $Y \subset \mathbf{R}^d$ is a domain of LSP and $d \geq p$. We assume that f is a C^1-smooth and one-to-one map, whose value can be obtained (with sufficient accuracy) for any point in X. In particular, for a sphere the map is implemented using the Mie theory [28], which provides an analytical solution in terms of infinite series. Specifics of the Mie theory are also discussed in [27]. Typical LSP's for spherical particles, defined by Eq. (12) below, are presented in Fig. 1.

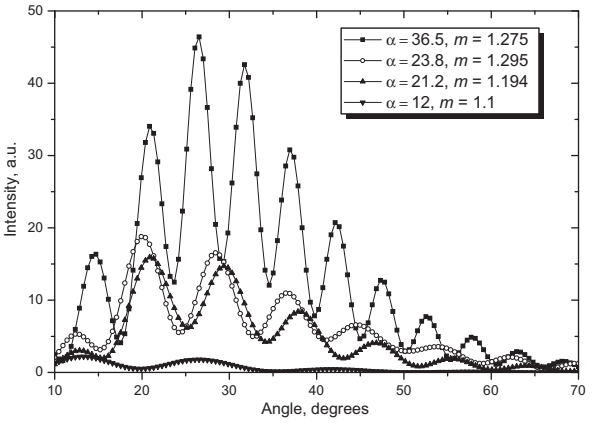

Fig. 1. The light-scattering profiles of spheres for several different size parameters α and refractive indices m.

Let us denote $x^\dagger = f^{-1}(y), \quad y \in f(X)$. The properties of f imply that f^{-1} is continuous. To incorporate experimental data with noise we extend f^{-1} to the whole $\mathbf{R}^d$

$$g : \mathbf{R}^d \to X,\ x^\delta = g(y) = \arg\min_{x\in X} \|y - f(x)\|, \tag{1}$$

which is an optimization problem with respect to the L_2-norm in the space of LSPs. However, we note that other norms can also be used. We also note, that our assumptions imply that the noise-free problem is (relatively) well-posed. This is discussed in more details in [27], including numerical analysis of the Jacobian of f.

The problem is to create a database (lookup table) to approximate map g (to determine particle parameters) with given accuracy. Although, a detailed database can be constructed for the inverse problem for sphere considered in this paper, this is expected to be computationally prohibitive for more complex particle shapes with greater number of parameters. For a sparse database even the linear interpolation may be unsuitable, therefore the nearest-neighbor interpolation seems to be generally the most reliable choice. The latter is further used in this paper, and can be defined as global optimization over database

$$g_1 : \mathbf{R}^d \to X_1, x_1 = g_1(y) = \arg\min_{x\in X_1} \|y - f(x)\|, \tag{2}$$

where $X_1 \subset X$ is a finite set of database points (their parameter values).

The error in the solution can be estimated as

$$\|x^\dagger - x_1\| \le \|x^\dagger - x^\delta\| + \|x^\delta - x_1\|,$$

where $\varepsilon^\delta = \|x^\dagger - x^\delta\|$ represents the noise error, $\varepsilon^d = \|x^\delta - x_1\|$ represents the discretization error. A careful consideration of experimental noise lies outside the scope of this paper, so we assume that noise is small enough, i.e., experimental LSPs lie in a 'small neighborhood' of $f(X)$, in which g is still continuous.

3. Algorithm for construction of interpolating database

Initially, the domain of parameters is normalized to be the unit hypercube, so influence of parameters is balanced to some extent. The proposed algorithm is adaptive and is based on subsequent division of the parameter domain into the rectangular cells (hyperrectangles) whose centers constitute the database according to the rules described further. New points are added to the database as a result of division of cells into smaller hyperrectangles. More specifically, we use the same structure of cells and division rules as in the global-optimization method DiRect [16]. According to these rules, a cell is divided along its largest size into three equal parts. Then the center of the central part coincides with the center of the original cell, and only two points together with newly calculated values of f (LSPs) are added to the database. If there are several equal largest sizes then division is performed along each of them serially.

Particular structure of the database is chosen from purposes of usability and simplicity of implementation. Moreover, it addresses particle parameters almost independently of each other, which is beneficial when parameters have very different physical significance, typical ranges, and influence on the LSP. However, the algorithm described below also applies to any other structure, e.g., division into simplices [23].

Let us further define two neighborhoods for any point $x \in X$:

1. $N(x)$ is discrete set of neighbors in the database, i.e., centers of those hyperrectangles that touch by facets the hyperrectangle around x.
2. $B(x,\varepsilon)$ defined as

$$B(x,\varepsilon) = \{u \in X \,\big|\, |u^j - x^j| < \varepsilon^j,\ 1 \quad j = 1,\ldots,p\}. \tag{3}$$

The superscript j stands for the j-th component of a vector in X, and ε is a p-component vector defined prior to constructing the database. Eq. (3) defines a ball with respect to the L_∞-norm with preliminary rescaling of the parameters (dividing them component-wise by ε). This particular norm is chosen in contrast to the L_2-norm due to possible incommensurability of different parameters discussed above. However, the algorithm may incorporate any other norm through corresponding change of Eq. (3).

The database, a set of pairs $(x_1, f(x_1))$, $x_1 \in X_1$, is constructed iteratively. At each step the database (each element x_1) is tested against condition:

$$\forall x_1 \in X_1 \quad \forall u \in N(x_1) \quad \frac{\forall v \in X_1}{B(x_1,\varepsilon)}$$
$$\|f(u) - f(x_1)\| < \|f(v) - f(x_1)\|. \tag{4}$$

If it is not satisfied the hyperrectangle centered at x_1 is divided further. First, this condition implies that $\forall x_1 \in X_1$ $N(x_1) \subset B(x_1,\varepsilon)$, i.e. the database needs to be sufficiently dense with respect to the distance between neighboring points in the parameter space, compared to the predefined accuracy ε. Second, condition (4) states that the LSPs of all neighboring to x_1 points in the database should be closer than the LSPs of all non-neighboring points outside the larger neighborhood $B(x_1,\varepsilon)$.

Construction of the database starts with an arbitrary division of the domain X into hyperrectangles.

The domain X itself does not need to be rectangular; it can be any connected region composed of hyperrectangles, satisfying the structure of the DiRect algorithm [16]. Our assumptions on f and g imply that the algorithm will converge after a finite number of steps, producing a database whose every element satisfies (4). Let us additionally define several distances in the space of LSPs:

$$\delta(x) = \inf_{u \in X \setminus B(x,\varepsilon)} \|f(u) - f(x)\|,$$

$$\delta_1(x) = \min_{u \in X_1 \setminus B(x,\varepsilon)} \|f(u) - f(x)\|,$$

$$\delta_2(x) = \max_{u \in N(x)} \|f(u) - f(x)\|. \quad (5)$$

where $\delta(x)$ is error radius in the data space, $\delta_1(x)$ is error radius over database, $\delta_2(x)$ is a distance to the most distant neighbor. A schematic interpretation of definition (5) is presented in Fig. 2.

Then condition (4) is equivalent to

$$\forall x_1 \in X_1 \quad \delta_2(x_1) < \delta_1(x_1), \quad (6)$$

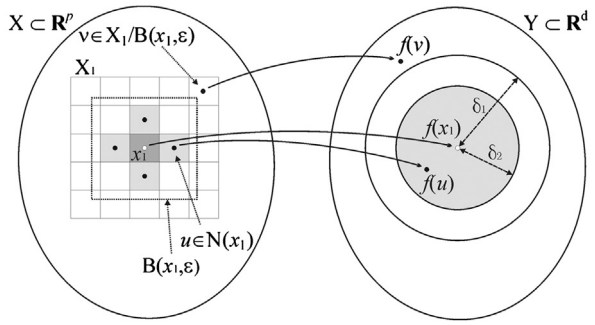

Fig. 2. Schematic interpretation of definition (5).

which must be satisfied for a database after the construction is complete. Last condition is also illustrated in Fig. 2.

Formally, the algorithm for database construction can be expressed as follows:

```
Normalize the domain to be enclosed in the unit hypercube
Generate initial points of database X_1
Do
                {
                    Determine the set S of points that do not satisfy
  condition (4)
                    For all x∈S
                      {
                          {Divide hyperrectangle corresponding to x
  into sub-rectangles
                      }
                    } while (S≠∅)
For all x_1∈X_1 calculate δ_2(x_1).
```

Several steps of the process are illustrated in Fig. 3 for a very simple model map L: $X \to \mathbf{R}^2$, $X \subset \mathbf{R}^1$. This map is a parameterization of a curve L with a single variable x. Here the value of ε is chosen so that for the database at step 1 geometrical sizes of $N(x)$ and $B(x,\varepsilon)$ are approximately equal. Two circles are drawn around $L(x_0)$: the filled one with radius $\delta_1(x_0)$ and the dashed one with radius $\delta_2(x_0)$. In this particular case the $\delta_1(x_0)$ is the same as the smallest distance to non-neighbors. At step 1 $\delta_1(x_0) < \delta_2(x_0)$, which leads to refinement of the database. At step 2, $(\delta_2(x_0) - \delta_1(x_0))$ became smaller than at previous step but still positive. At step 3, as a result of an additional division, $\delta_1(x_0) > \delta_2(x_0)$, i.e. condition (6) is satisfied. After the final step (Fig. 3) condition (6) is equivalent to the statement that a dashed circle around any point does not contain database points from the distant parts of the curve.

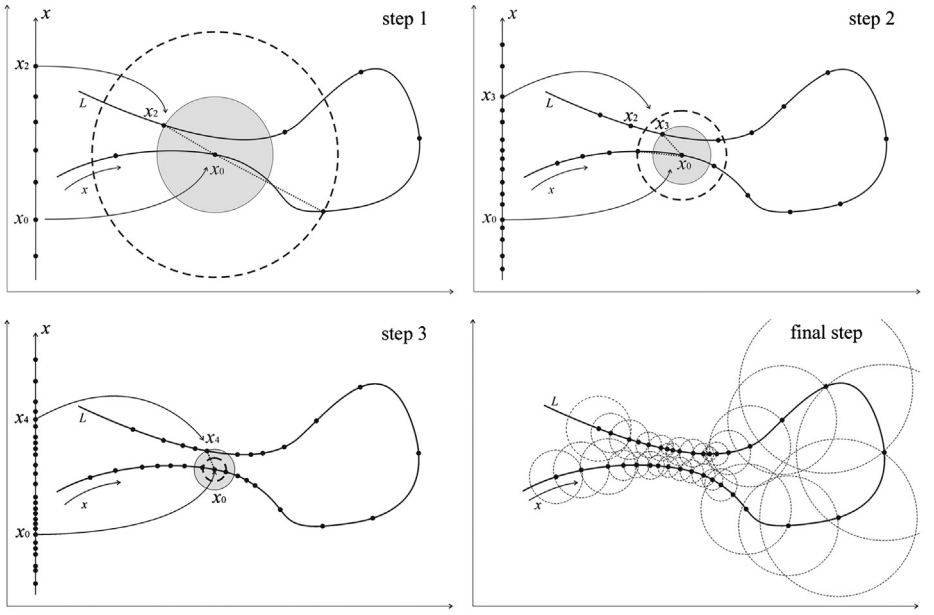

Fig. 3. Schematic description of the proposed algorithm. A filled circle represents the smallest distance to a non-neighbor $\delta_1(x)$, a dashed circle represents the largest distance to a neighbor $\delta_2(x)$.

From a practical viewpoint a desirable property of g_1 is its satisfactory accuracy on the whole $f(X)$:

$$\forall x \in X \quad g_1(f(x)) \in B(x,\varepsilon). \tag{7}$$

Diagram interpretation of the latter condition is presented at Fig. 4.

However, we know neither the values of f outside X_1 nor any bounds on its derivative. This implies that any condition on the database can be neither necessary nor sufficient for (7). In other words, condition (6) or any similar one can only imply anything about g_1 over $f(X_1)$, while condition (7) or any similar practical requirement is always satisfied over $f(X_1)$, since $\forall x_1 \in X_1 \; g_1(f(x_1))=x_1$. Nevertheless, define

$$Z(x) = \{x_1 \in X_1 \,|\, \|f(x_1)-f(x)\| < \delta_1(x)\}, \tag{8}$$

which is a minimal set of points that has to be removed from X_1 to make correspondingly reduced g_1 violate condition (7) at x. By construction, $\forall x_1 \in X_1 \; x_1 \in Z(x_1) \subset B(x_1,\varepsilon)$. Then condition (4) is equivalent to $N(x_1) \subset Z(x_1)$. Thus after the final construction step the database is redundant in the sense, that it is required to remove at least x and $N(x)$ from X_1 to break condition (7) at x.

Once the division is complete, we can estimate the sensitivity of the LSP to particle characteristics using the quantities defined in Eq.(5). All of them are positive due to continuity of g. Moreover, $\delta_1(x)$ is a non-smaller discrete approximation to $\delta(x)$, which in turn is an important measure that guarantees a certain accuracy of the inverse map:

$$\forall u, x \in X \quad \|f(u)-f(x)\| < \delta(x) \Rightarrow u \in B(x,\varepsilon). \tag{9}$$

A desirable property of $\delta_1(x)$ is a similar statement

$$\forall u, x \in X \quad \|f(u)-f(x)\| < \delta_1(x) \Rightarrow u \in B(x,\varepsilon) \tag{10}$$

which, however, cannot be proven due to discrete nature of available information (see the discussion above). In other words, there is no guarantee that $\delta_2(x) < \delta(x)$, although it may be so for many points in the database. This is illustrated in Fig. 6. No divisions around x_0 are required by condition (6), because $\delta_2(x_0) < \delta_1(x_0)$. However, $\delta(x_0)$ is less than both $\delta_2(x_0)$ and $\delta_1(x_0)$, so for some points x on the arc between x_3 and x_4 $||f(x)-f(x_0)|| < \delta_2(x_0)$ does not imply $x \in B(x_0,\varepsilon)$.

A schematic interpretation of conditions (9) and (10) is presented in Fig. 5.

There are several ways to sharpen condition (6) to better detect situations like shown in Fig. 6, e.g.:

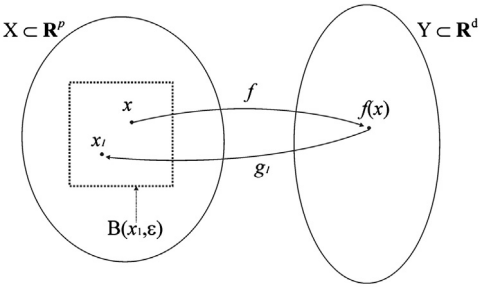

Fig. 4. Graph interpretation of condition (7).

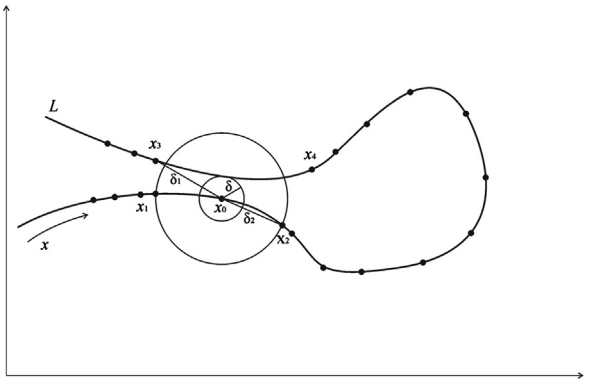

Fig. 5. Schematic interpretation of conditions (9) and (10).

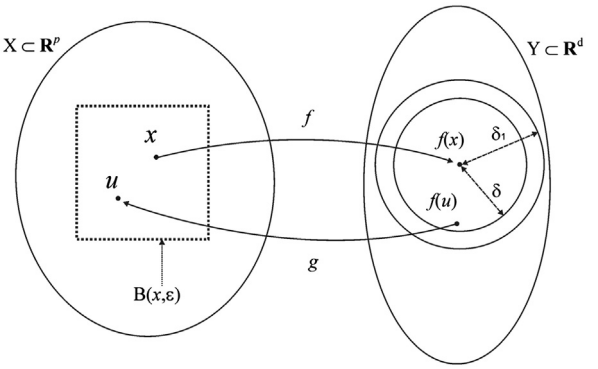

Fig. 6. An example of database that satisfies (4) but not (10).

- considering not only points from the database for the definition of $\delta_2(x)$, but also simplices constructed on any set of neighboring points;
- replacing the inequality in (6) by $2\delta_2(x) < \delta_1(x)$.

However, as we discussed above, strong assumptions, e.g. about derivatives of f, are required to make any such method rigorous, thus we leave them for a future research.

Values of δ_2 can be used to test the reliability of the inverse solution. If the solution is such that

$$\|y-f(g_1(y))\| \le \delta_2(g_1(y)), \tag{11}$$

then it is *likely* that the true solution $g(y)$ belongs to $B(g_1(y),\varepsilon)$, taking into account the limitations discussed above. A schematic interpretation of the latter condition is presented at Fig. 7.

However, violation of condition (11) does not necessarily imply that solution is largely wrong. So test (11) is expected to be useful only in a statistical sense when applied to large set of data. Its applicability to experimental data with noise is even more arguable, because then $y \neq f(g(y))$ and hence $||y-f(g_1(y))||$ is no more a direct measure of closeness of $g(y)$ and $g_1(y)$. A correction of condition (11) to alleviate this problem is possible, but it requires detailed knowledge of the structure of

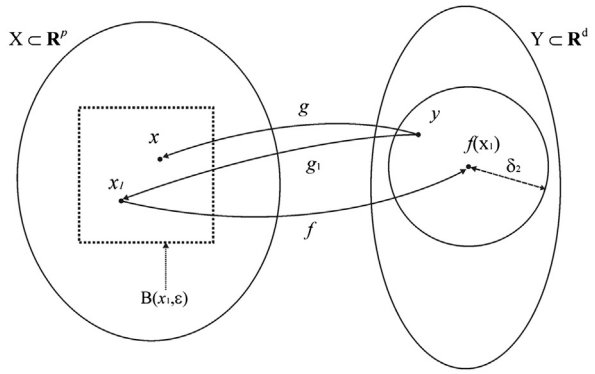

Fig. 7. Schematic interpretation of condition (11).

experimental noise, which lies out of scope of this paper. So when processing experimental data below, we present results both with and without using condition (11) as a threshold.

4. Implementation

4.1. Construction of the databases

In this paper we construct a database for homogeneous spherical particles. The LSP is defined as a uniform discretization of a function

$$I(\theta)=\frac{w(\theta)}{2\pi}\int_0^{2\pi} \mathrm{d}\varphi S_{11}(\theta,\varphi). \tag{12}$$

In the range of $\theta\in[10^\circ,70^\circ]$ using $k=128$ points. S_{11} is an element of the Mueller scattering matrix [28], θ and φ are the polar and azimuth angle respectively, and $w(\theta)$ is the weighting function [18]

$$w(\theta)=\frac{1^\circ}{\theta}\exp\left[-2\ln^2\left(\frac{\theta}{54^\circ}\right)\right]. \tag{13}$$

Calculation of a single value of map f for spheres using the Mie theory [28] takes several milliseconds on personal computer. We note that a LSP is measured for a single fixed wavelength.

A sphere is characterized by $d=2$ parameters: the size parameter $\alpha\in[5,40]$ and the refractive index (relative to the surrounding medium) $m\in[1.05,1.3]$. The size parameter is connected with radius r as $\alpha=2\pi r n_0/\lambda$, where the wavelength λ of the incident light is 660 nm; and the refractive index of surrounding media (buffer solution) m_0 is 1.337. The following required accuracy is chosen from practical reasons: $\delta\alpha=5$ and $\delta m=0.1$, so after normalization into the unit hypercube $\varepsilon=(\varepsilon^1,\varepsilon^2)=(0.14,0.4)$. The process is initialized by 9×9 uniform grid and finally leads to database C with number of points N_{db} equal to 9366 points. The corresponding set P_1 is presented in Fig. 8(C) together with a rectangle $\varepsilon^1\times\varepsilon^2$.

We also study the dependence of the database structure on chosen particle parameters and required accuracy ε. For this purpose we constructed five more databases. Parameters of all databases are summarized in Table 1 and the databases themselves are shown in Fig. 8. Databases A and B differ from C by lack of accuracy tests for m and α respectively. That is equivalent to setting respective component of ε equal to 1. Database D is a refined version of C with improved accuracy of both parameters. Alternatively a sphere can be characterized with α and the phase-shift parameter $\rho=2\alpha(m-1)\in[0.5,24]$. These parameters have been shown to better correlate with the LSP than α and m [10]. Databases E and F are constructed in coordinates (α,ρ), but only half of the square in this parameter space is covered, which is approximately an image of the original square in (α,m) coordinates. Since $\delta\rho=2[\alpha\delta m+(m-1)\delta\alpha]$, no single value can exactly correspond to used $\delta\alpha$ and δm. If the values of the latter are taken the same as for database C, then $\delta\rho$ varies from 1.5 to 11. We chose these two limiting cases for construction of databases F and E respectively.

Note that the choice of parameters has a great impact on the structure and size of the database. As expected, total number of points in the database inversely correlates with ε. Moreover, for all databases the density significantly varies over P, which proves the need for an adaptive procedure.

4.2. Testing of the databases

4.2.1. Exact synthetic data for spheres

Constructed databases were tested on a set of 1000 theoretical LSPs of homogenous spheres with parameters randomly chosen in P. For all databases we have used threshold (11), i.e. the data, for which result has not passed the test, were discarded. However, we have also tried the database C without the threshold; corresponding results are denoted by C_1. We denote the number of LSPs (of the initial 1000) that pass the threshold as N_{th}. We also process the same data with the DiRect method. This is a global optimization method which is supposed to give the reference (almost exact) solution albeit at a relatively large computational cost. The summary of the results, including mean absolute errors (MAE) of α and m and values of N_{th}, are presented in Table 2 together with results for all other test cases. An immediate conclusion from this table is that the proposed method is about 100 times faster than the rigorous global optimization technique (DiRect) for a single LSP. It is reasonable to assume that computational speed of the latter is comparable to that of other global optimization methods, e.g. [14,15]. Therefore, the method proposed in this paper is also much faster than other published methods to characterize spheres from LSPs.

Distribution of errors for this particular case is shown in Fig. 9. The errors of the DiRect algorithm are negligibly small, which is expected for noise-free data. Note that $N_{\mathrm{th}}\geq 966$ for these data; hence, there is almost no difference between C and C_1. Going from A to D the errors decrease in correlation with increasing N_{db}. Both databases E and F constructed in coordinates (α,ρ) are generally less accurate than other databases of comparable sizes (although there is no direct analog for database E). These conclusions are also supported by values of MAE presented in Table 2. Another general conclusion is strong negative correlation between errors in α and m, which is

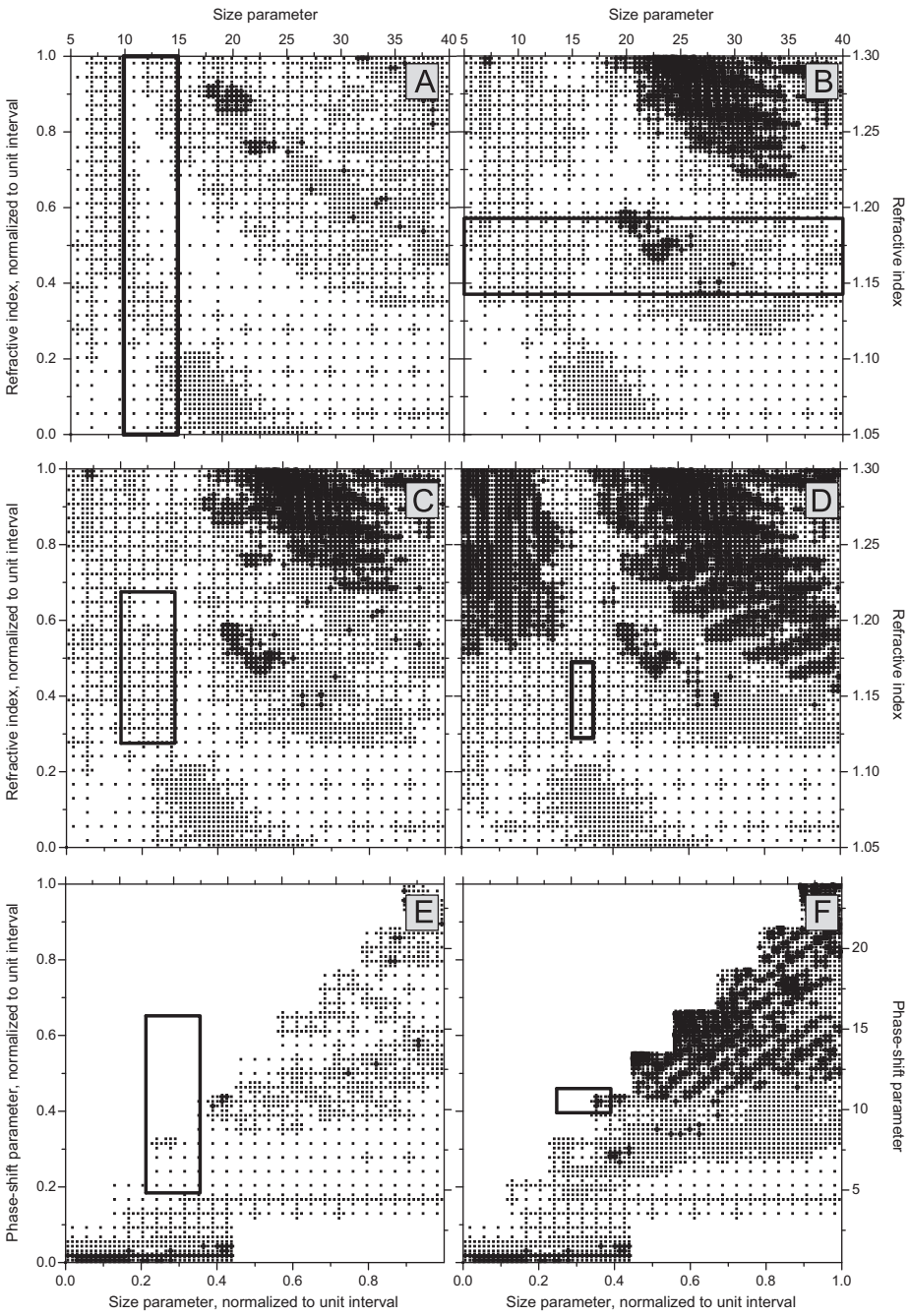

Fig. SEQ Fig. * ARABIC 8. Diagrams of the constructed databases in the parameter space. Databases A–D are constructed in terms of size parameter α and refractive index m, while E and F, in terms of α and the phase-shift parameter ρ. Rectangles represent required accuracy ε. Two scales are shown for each parameter corresponding to original and normalized values.

discussed in detail in [27]. For databases E and F some points with the greatest errors (less than 2% of the total amount of points) are not shown in Fig. 9 because they fall out of plot bounds.

Fig. 10 shows 100 largest (in terms of $|\delta\alpha|+|\delta m|$) errors for each database (method) in even more detail, indicating both actual and calculated values of the parameters. In particular, it shows that the greatest errors for databases E and F occur for small α and large m and are mainly along the m-axis.

4.2.2. *Noisy synthetic data for spheres*

The same LSPs as used in previous test case were contaminated with the additive white Gaussian noise. The amplitude σ of the noise is chosen independently for every LSP as a random value in the range [0,0.2]

multiplied by the root-mean-squared value of the corresponding LSP.

These modified LSPs were processed by all databases and the DiRect method. Distribution of obtained errors and detailed review of 100 largest of them are shown in Figs. 11 and 12 respectively. For databases C, E, and F and for the DiRect method some points fall out of plot bounds in Fig. 11. The errors for all databases except C_1 is comparable to that for the noise-free data but at a cost of discarding about 80% of input data. Moreover, N_{th} correlates with errors and inversely correlates with N_{db}, which can be explained through relation of all these values to δ_2. If the threshold is not used, then errors are about twice greater (compare C_1 and C). This means that discretization errors (map g_1, implemented by database C_1, without noise) and errors due to noise (map g, implemented by DiRect method, with noise) do add up for results of database C_1. The fact that database D is more accurate than the DiRect method is delusive, since the errors of the former are calculated from a smaller set of more accurate LSPs.

Pros and cons of using the threshold largely depend on the particular application and the structure of the noise. In particular, an important open question is whether N_{th} LSPs are representative for the whole set of measurements (see also Section 4.2.3). However, some kind of threshold is clearly desirable when the noise is large enough to render map g discontinuous. In this case, even exact optimization may lead to a different part of the manifold $f(X)$ compared to the location of the true value. Hence, obtained errors of parameters may be comparable to the size of the whole domain X. That is exactly what is shown in Fig. 12 (DiRect). This problem can be partly alleviated by determining the probability distribution function over X for each LSP instead of a single nearest point, which can be done through the Bayesian inference [18]. However, the information obtained from such noisy data is still very small.

In case of Fig. 12 an optimal threshold should filter out only about 20 LSPs that result in large errors both for DiRect method and for database C_1. So the threshold (11), which filters out about 800 LSPs, is way too restrictive. We leave a problem of estimating the optimal threshold from the constructed database for a future research, although Bayesian approach [18] may provide an answer.

Table 1
Parameters of databases constructed with the adaptive algorithm.

Database	$\delta\alpha$	δm or $\delta\rho$	ε^1	ε^2	N_{db}
A	5	$\delta m=0.25$	0.14	1	3254
B	35	$\delta m=0.1$	1	0.4	8664
C	5	$\delta m=0.1$	0.14	0.4	9366
D	2	$\delta m=0.05$	0.057	0.2	16074
E	5	$\delta\rho=11$	0.14	0.47	1725
F	5	$\delta\rho=1.5$	0.14	0.064	7769

4.2.3. Exact synthetic data for spheroids

We also studied stability of the inverse solution with respect to particle deformations. These so-called model errors are always present in real applications, when a simple shape model is used. Each sphere in the original test set was replaced by an equi-volume spheroid with axes ratio randomly and independently chosen in [0.9,1.1]. Orientation of each spheroid (its symmetry axis) with respect to incident direction is chosen randomly in [0°,90°]. Calculation of LSPs of spheroids was performed with the T-matrix method [29].

These LSPs were processed by all databases and the DiRect method. Distribution of obtained errors and detailed review of 100 largest of them are shown in Figs. 13 and 14 respectively. Again, for databases E and F a minor number of points fall out of plot bounds in Fig. 13. Overall, the effect of shape deformation on LSPs is less than that of the Gaussian noise in the previous test case. Results of DiRect are reliable for all input LSPs, and corresponding MAEs are about 1.5 times smaller than for the spheres with noise (Table 2). Also, majority of the LSPs pass the threshold for all databases ($N_{th} \geq 872$). That is why MAEs for the databases approximately equal to

Table 2
Summary of the results of processing of synthetic and experimental data for spheres.

	N_{db}	A 3254	B 8664	C 9366	C_1 9366	D 16074	E 1725	F 7769	DiRect –
	Time[a] (ms)	7	18	19	19	36	5	15	1550
Ideal sphere	MAE(α)	0.17	0.16	0.14	0.14	0.11	0.26	0.18	0.002
	MAE(m), 10^{-2}	0.44	0.39	0.34	0.35	0.24	0.94	0.64	0.01
	N_{th}	988	973	984	1000	982	972	966	1000
Noisy sphere	MAE(α)	0.19	0.22	0.17	0.30	0.15	0.28	0.20	0.16
	MAE(m), 10^{-2}	0.44	0.50	0.39	0.71	0.28	1.11	0.77	0.41
	N_{th}	234	213	193	1000	148	357	219	1000
Ideal spheroid	MAE(α)	0.23	0.22	0.20	0.20	0.17	0.31	0.24	0.12
	MAE(m), 10^{-2}	0.56	0.51	0.46	0.49	0.39	1.03	0.77	0.32
	N_{th}	976	904	913	1000	872	969	914	1000
Experiment	mean(α)	23.65	23.42	23.35	23.36	23.66	23.36	23.40	23.86
	mean(m)	1.193	1.196	1.194	1.195	1.194	1.205	1.197	1.207
	N_{th}	238	115	90	751	149	541	325	751

[a] Time for computing a single inverse solution (value of g_1).

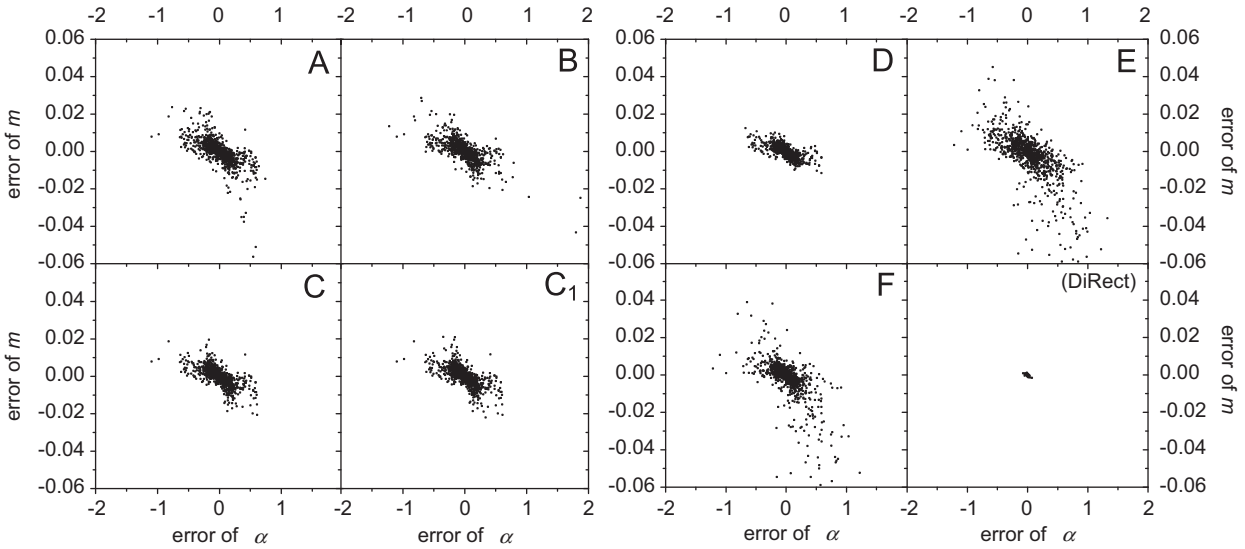

Fig. 9. Errors in determination of parameters using constructed databases A–F and DiRect algorithm to process 1000 theoretical LSPs of spheres with random parameters from domain P (see text for details).

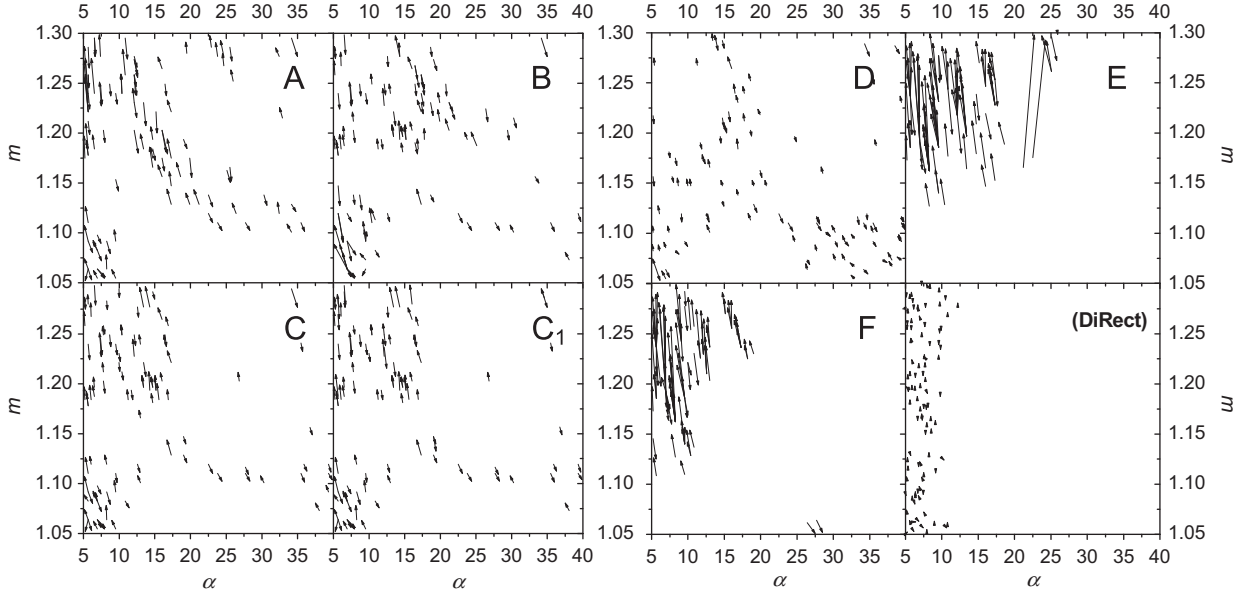

Fig. 10. Same as Fig. 9 but showing 100 largest errors as arrows with head and tail corresponding to the true and calculated values respectively.

those for the noise-free spheres plus a constant amount (0.06 for α and 0.0012 for m). Finally, we can conclude that both databases and DiRect are tolerant to shape deformation of the considered magnitude.

4.2.4. Experimental data for polystyrene spheres

Finally we tested the databases on experimental data. The LSPs of polystyrene microspheres (Invitrogen C37253) in the buffer solution were measured by a SFC [4]. The angular range, the wavelength, and the refractive index of the medium were the same as described in Section 4.1. According to the producer's data, mean α of microspheres is 25 with coefficient of variation 4%. Relative refractive index of bulk polystyrene at this wavelength is 1.185 [30], while 1.183 produced the best fit to diffuse reflectance and transmittance of a suspension of 1 μm polystyrene microspheres [31].

Calculated distributions of the sample over α and m are shown in Fig. 15. Corresponding mean values and N_{th} are given in Table 2. Results of DiRect generally agree with reference values cited above. However, there are small systematic differences due to distortion of experimental LSPs caused by variation of flow speed and particle position in the flow of the SFC. Results of all databases are rather accurate within the experimental errors themselves. Moreover, mean of α converge to its value obtained by DiRect with increasing N_{db}. However, mean of m seem to converge to value 1.194, which is

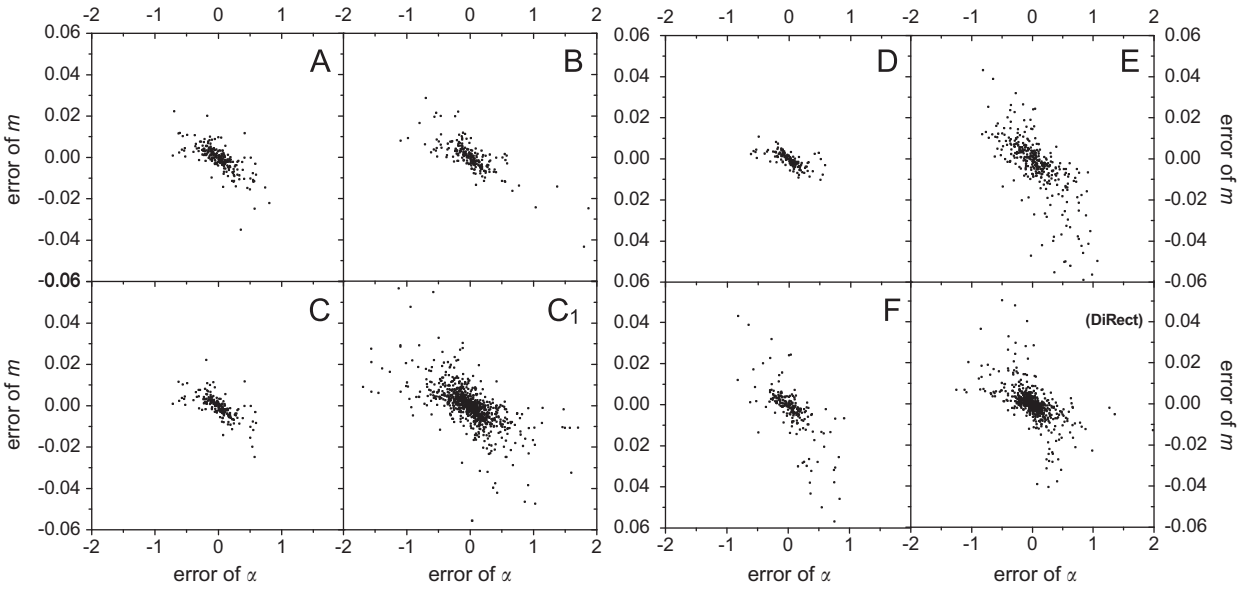

Fig. 11. Same as Fig. 9 but for LSPs of spheres with added noise.

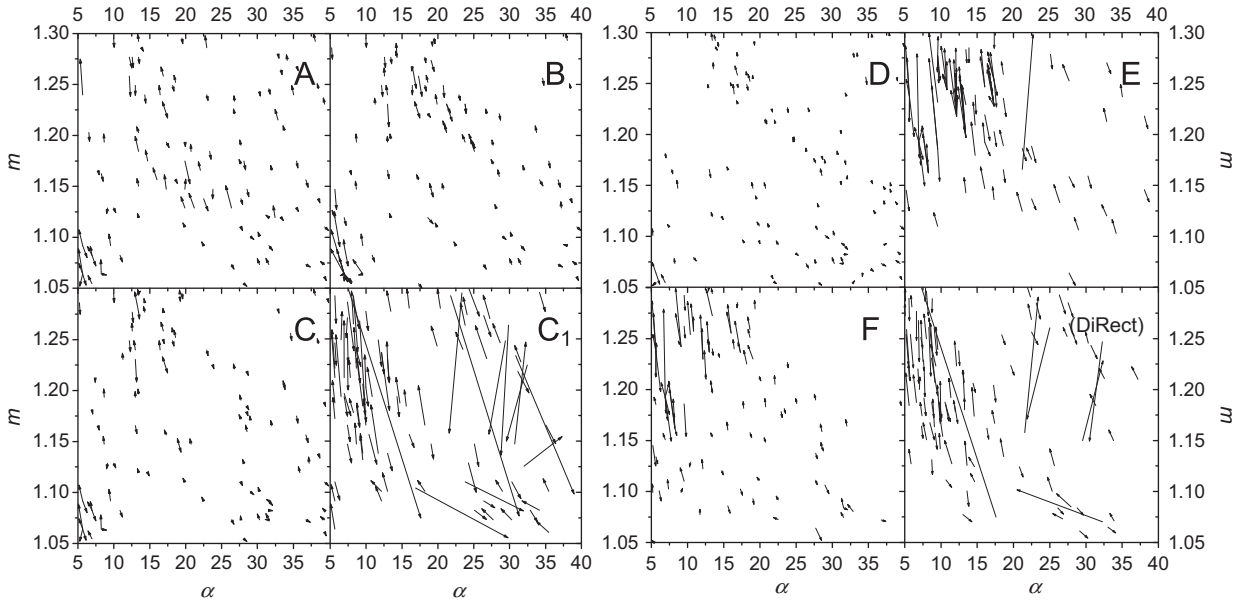

Fig. 12. Same as Fig. 10 but for LSPs of spheres with added noise.

different from the result of DiRect (1.207) and is actually closer to the reference values. However all these differences are much smaller than the required accuracy ε, and hence may constitute a discretization artifact. Unlike all previous test cases, N_{th} shows little correlation with N_{db}, emphasizing the need for a more sophisticated threshold.

Typical experimental LSP is shown in Fig. 16 together with corresponding fits by the DiRect method and database C. One can see that the discretization error of the database (difference between two fits) is smaller than the experimental noise (distortion).

5. Conclusion

The adaptive method proposed in this paper allows one to create a non-uniform database to approximate highly-nonlinear map with a given accuracy without access to derivatives. Nearest-neighbor interpolation on the database is further used as an approximate inverse map. The main novelty of this method is that possible large errors of the inverse map caused by two points distant in the argument space but close in the image space, are detected and addressed. Such behavior of direct

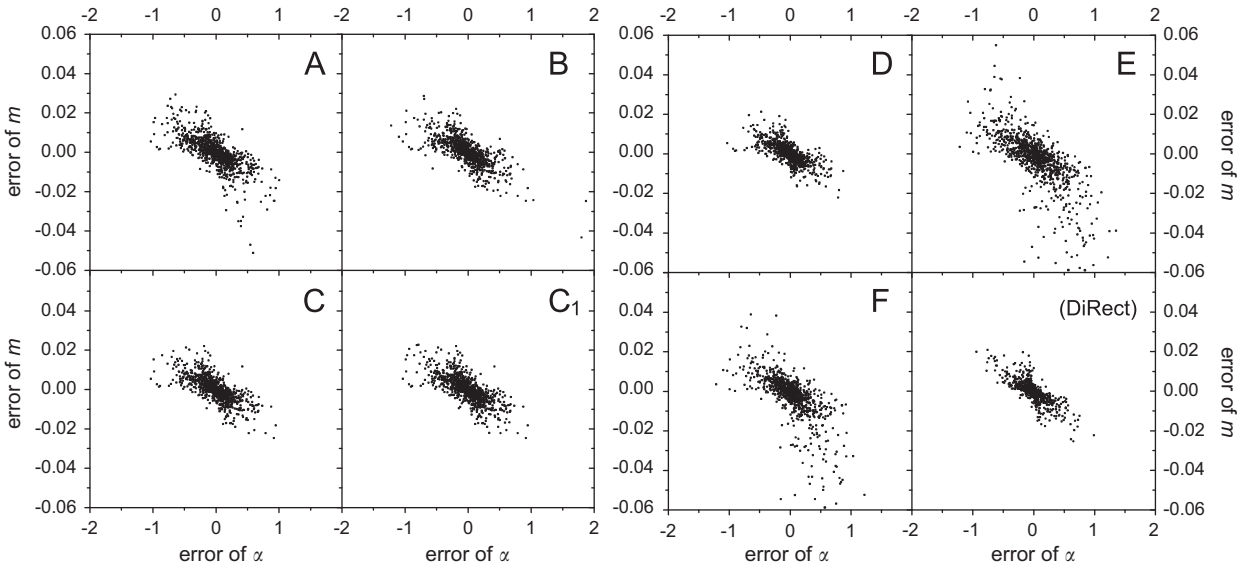

Fig. 13. Same as Fig. 9 but for LSPs of spheroids.

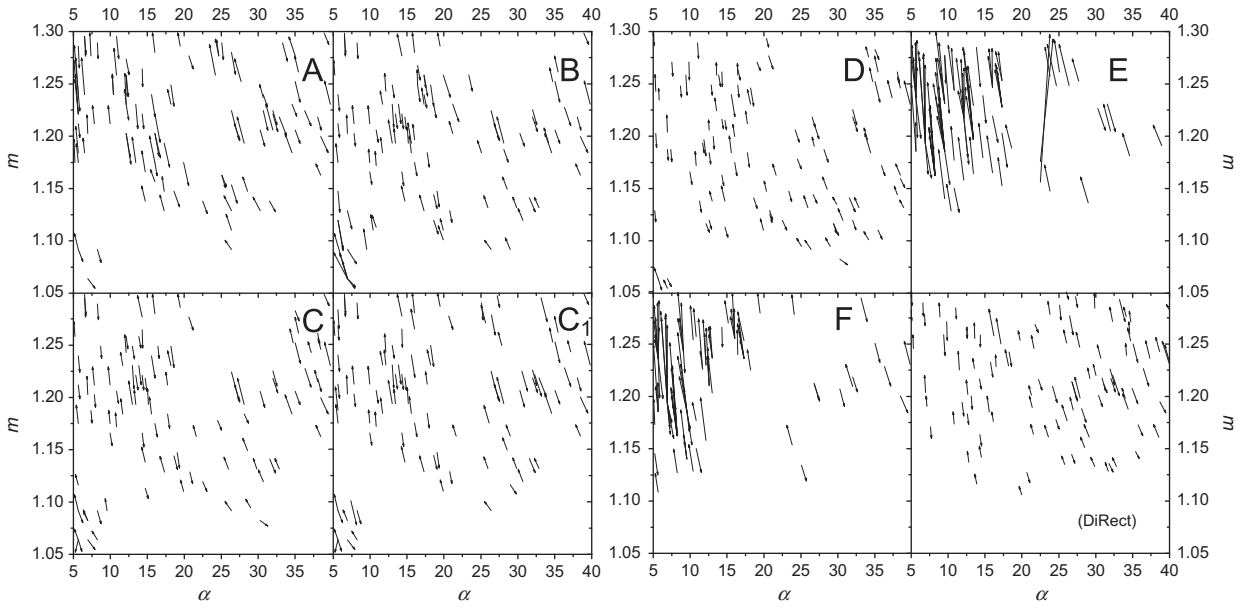

Fig. 14. Same as Fig. 10 but for LSPs of spheroids.

map often occur, e.g., in inverse scattering problems and significantly deteriorates performance of many other methods to solve these problems.

We tested the proposed method for solution of ILS problem of a non-absorbing homogeneous sphere using the light-scattering patterns measured with the scanning flow cytometer. The results agree with a reference global-optimization method DiRect both for synthetic (with or without noise) and experimental data. Moreover, we tested the database against the model errors caused by shape deformation of a sphere. The method contains a number of options controlling the construction and usage of a database. These options include a particular set of parameters to characterize the particle, norms in argument and image spaces, shape of elementary cells to divide the argument space into (hyperrectangles, simplices, etc.) and subdivision algorithm, and threshold to indicate accurate solution. Tuning of these parameters to maximize the speed and accuracy of the inverse map remains an open problem.

In future we plan to apply this method to ILS problems for more complicated particle shapes, such as spheroids, cylinders, and biconcave disks. These problems are of great practical interest for characterizations of blood platelets, rod-shaped bacteria, and red blood cells.

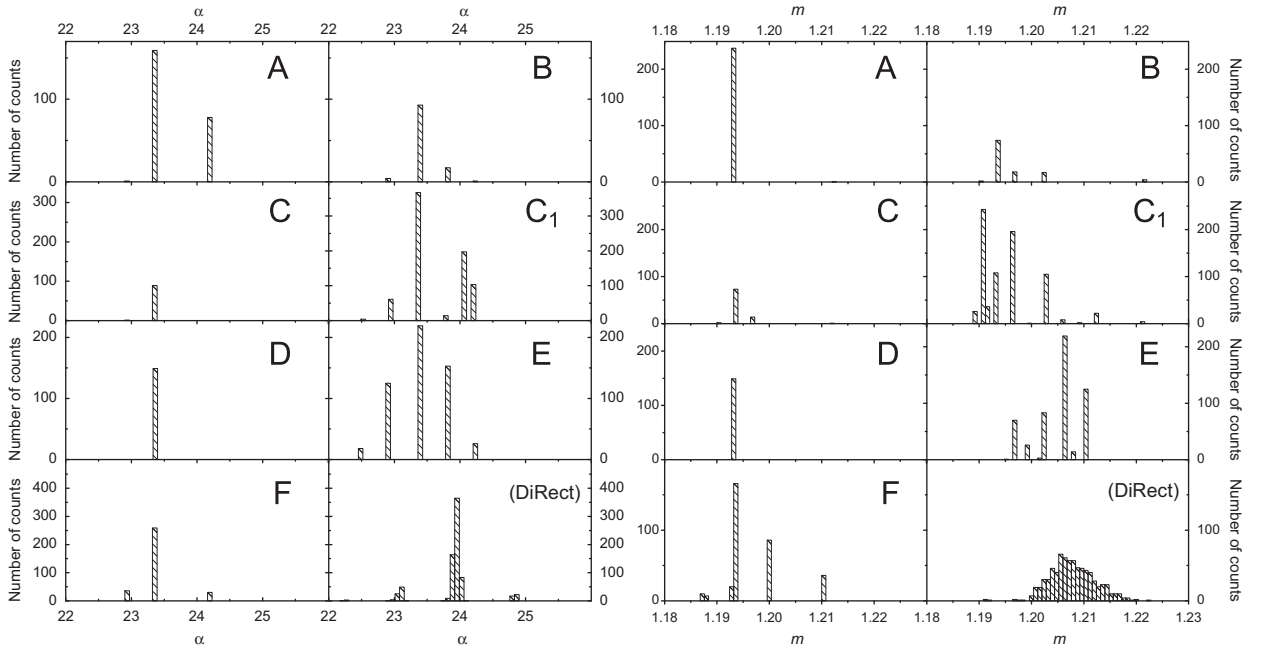

Fig. 15. Distributions of a sample of polystyrene microspheres over α and m, obtained by processing LSPs measured with a SFC by databases A–F and the DiRect method.

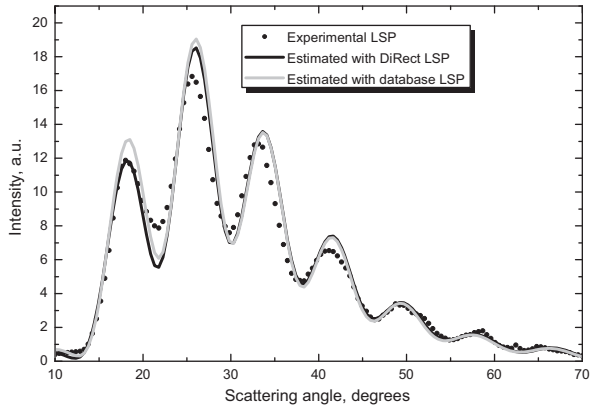

Fig. 16. Sample experimental LSP of a polystyrene microsphere and corresponding theoretical LSPs fitted with DiRect (α=23.27 and m=1.2022) and database C (α=23.36, m=1.1966).

Adoption of this method into hematological analysis should increase a power of characterization of blood cells from physical characteristics.

Acknowledgment

This work was supported by grant from the program of Presidium of the Russian Academy of Science, No 2009-27-15, program of the Russian Government "Research and educational personnel of innovative Russia" (contracts P1039, P422, 14.740.11.0729, and 14.740.11.0921), by grant from Government of Russian Federation 11.G34.31.0034, and by the President of the Russian Federation Program for State Support of the Leading Scientific Schools (grant NSh-65387.2010.4).

References

[1] Mishchenko MI, Hovenier JW, Travis LD. Light scattering by nonspherical particles: theory measurements, and applications. New York: Academic Press; 2000.
[2] Givan AL. Flow cytometry: first principles. 2nd ed.. New York: Wiley–Liss; 2001.
[3] Maltsev VP. Scanning flow cytometry for individual particle analysis. Rev Sci Instrum 2000;71:243–55.
[4] Strokotov DI, Moskalensky AE, Nekrasov VM, Maltsev VP. Polarized light-scattering profile—advanced characterization of nonspherical particles with the scanning flow cytometry. Cytometry 2011;79A: 570–9.
[5] Maltsev VP, Semyanov KA. Characterisation of bio-particles from light scattering.Utrecht: VSP; 2004.
[6] Ludlow IK, Everitt J. Inverse Mie problem. J Opt Soc Am A 2000;17(12):2229–35.
[7] Ludlow IK, Everitt J. Application of Gegenbauer analysis to light-scattering from spheres. Theory. Phys Rev E 1995;51:2516–26.
[8] Min SL, Gomez A. High-resolution size measurement of single spherical particles with a fast Fourier transform of the angular scattering intensity. Appl Opt 1996;35:4919–26.
[9] Semyanov KA, Tarasov PA, Zharinov AE, Chernyshev AV, Hoekstra AG, Maltsev VP. Single-particle sizing from light scattering by spectral decomposition. Appl Opt 2004;43:5110–5.
[10] Maltsev VP, Lopatin VN. Parametric solution of the inverse light-scattering problem for individual spherical particles. Appl Opt 1997;36:6102–8.
[11] Berdnik VV, Gilev KV, Shvalov A, Maltsev V, Loiko VA. Characterization of spherical particles using high-order neural networks and scanning flow cytometry. J Quant Spectrosc Radiat Transfer 2006;102:62–72.
[12] Berdnik VV, Loiko VA. Retrieval of size and refractive index of spherical particles by multiangle light scattering: neural network method application. Appl Opt 2009;48:6178–87.
[13] Wang Z, Ulanowski Z, Kaye PH. On solving the inverse scattering problem with RBF neural networks: noise-free case. Neural Comput Appl J 1999;8:177–86.

[14] Zakovic S, Ulanowsk ZJ, Bartholomew-Biggs MC. Application of global optimization to particle identification using light scattering. Inverse Probl 1998;14:1053.
[15] Bartholomew-Biggs MC, Ulanowski ZJ, Zakovic S. Using global optimization for a microparticle identification problem with noisy data. J Global Optim 2005;32:325–47.
[16] Jones DR, Pertunnen CD, Stuckman BE. Lipschizian optimization without the Lipschitz constant. J Optim Theory Appl 1993;79: 157–81.
[17] Zharinov AE, Tarasov PA, Shvalov AN, Semyanov KA, van Bockstaele DR, Maltsev VP. A study of light scattering of mononuclear blood cells with scanning flow cytometry. J Quant Spectrosc Radiat Transfer 2006;102:121–8.
[18] Strokotov DI, Yurkin MA, Gilev KV, van Bockstaele DR, Hoekstra AG, Rubtsov NB, et al. Is there a difference between T- and B-lymphocyte morphology? J Biomed Opt 2009;14:064036–312.
[19] Li W, Jaffe JS. Sizing homogeneous spherical particles from intensity-only angular scatter. J Opt Soc Am A 2010;27:151–8.
[20] Kolesnikova IV, Potapov SV, Yurkin MA, Hoekstra AG, Maltsev VP, Semyanov KA. Determination of volume, shape and refractive index of individual blood platelets. J Quant Spectrosc Radiat Transfer 2006;102:62–72.
[21] Yurkin MA, Semyanov KA, Tarasov PA, Chernyshev AV, Hoekstra AG, Maltsev VP. Experimental and theoretical study of light scattering by individual mature red blood cells with scanning flow cytometry and discrete dipole approximation. Appl Opt 2005;44: 5249–56.
[22] Yurkin MA. Discrete dipole simulations of light scattering by blood cells. PhD thesis. University of Amsterdam; 2007.
[23] Pavo J, Gyimothy S. Adaptive inversion database for electromagnetic nondestructive evaluation. NDT&E Int 2007;40:192–202.
[24] Gyimothy S, Pavo J. Qualification of the inverse problem of defect reconstruction using optimized mesh database. COMPEL 2005;24: 436–45.
[25] Mosegaard K, Sambridge M. Monte Carlo analysis of inverse problems. Inverse Probl 2002;18:R29.
[26] Yang K, Chapman NR, Ma Y. Estimating parameter uncertainties in matched field inversion by a neighborhood approximation algorithm. J Acoust Soc Am 2007;121:833–43.
[27] Dyatlov GV, Gilev KV, Yurkin MA, Maltsev VP. An optimization method with precomputed starting points for solving the inverse Mie problem. Inverse Probl 2012;28:045012.
[28] Bohren CF, Huffman DR. Absorption and scattering of light by small particles.New York: Wiley; 1983.
[29] Mishchenko MI, Travis LD. Capabilities and limitations of a current FORTRAN implementation of the T-matrix method for randomly oriented, rotationally symmetric scatterers. J Quant Spectrosc Radiat Transfer 1998;60:309–24.
[30] Kasarova SN, Sultanova NG, Ivanov CD, Nikolov ID. Analysis of the dispersion of optical plastic materials. Opt Mater 2007;29: 1481–90.
[31] Ma X, Lu JQ, Brock RS, Jacobs KM, Yang P, Hu X. Determination of complex refractive index of polystyrene microspheres from 370 to 1610 nm. Phys Med Biol 2003;48:4165–72.