

## Weakly constrained Lucy-Richardson with applications to inversion of light scattering data

Alessandro Buccini · Marco Donatelli ·  
Fabio Ferri

Received: date / Accepted: date

**Abstract** Lucy-Richardson (LR) is a classical iterative regularization method largely used for the restoration of nonnegative solutions. LR finds applications in many physical problems, such as for the inversion of light scattering data. In these problems, there is often additional information on the true solution that is usually ignored by many restoration methods because the related measurable quantities are likely to be affected by non-negligible noise.

In this article we propose a novel Weakly Constrained Lucy-Richardson (WCLR) method which adds a weak constraint to the classical LR by introducing a penalization term, whose strength can be varied over a very large

---

A. Buccini  
Department of Mathematical Sciences,  
Kent State University,  
233 MSB, 1300 Lefton Esplanade,  
44242 Kent, Ohio,  
Tel.: +1 330 672 1481  
E-mail: abuccini@kent.edu

M. Donatelli  
Dipartimento di Scienza e Alta Tecnologia,  
Università degli Studi dell'Insubria,  
Via Valleggio 11,  
22100 Como, Italy,  
Tel.: +39-031-2386378  
Fax: +39-031-2386209  
E-mail: marco.donatelli@uninsubria.it

F. Ferri  
Dipartimento di Scienza e Alta Tecnologia,  
To.Sca.Lab,  
Università degli Studi dell'Insubria,  
Via Valleggio 11,  
22100 Como, Italy,  
Tel.: +39-031-2386251  
Tel.: +39-031-2386294  
Fax: +39-031-2386250  
E-mail: fabio.ferri@uninsubria.it

range. The WCLR method is simple and robust as the standard LR, but offers the great advantage of widely stretching the domain range over which the solution can be reliably recovered. Some selected numerical examples prove the performances of the proposed algorithm.

**Keywords** Lucy-Richardson Algorithm · Light Scattering · Particle Sizing

**Mathematics Subject Classification (2000)** 65R32 · 78A46 · 78A10

## 1 Introduction

In this work we consider the case of a Fredholm integral equation of the first kind

$$I(\theta) = \int K(\theta, R) N(R) dR. \quad (1)$$

where  $I(\theta)$  represents the data evaluated at the independent variable  $\theta$ ,  $N(R)$  is the unknown function that depends on the variable  $R$ , and  $K(\theta, R)$  is a known integral kernel with compact support. The latter condition implies that equation (1) is ill-posed, meaning that the solution is not unique and does not depend continuously on the data, *i.e.*, small perturbations in  $I$  greatly affect the solution  $N$ .

We are going to consider inverse problems of the form (1) coming from light scattering methods in optics [2, 6, 15–18], where the main goal of the optical technique is to recover the particle-size distribution  $N(R)$  which characterizes the sample under investigation. Thus the solution  $N(R)$  is defined to be positive ( $N > 0$ ) with a positive support ( $R > 0$ ). Among the various scattering techniques, the most popular ones are Dynamic Light Scattering (DLS), Static or Elastic Light Scattering (ELS) and Multi Wavelength Turbidimetry (MWT). All of them can be easily implemented by using simple experimental setups (see for example [7] and references therein), with the possibility of characterizing simultaneously a very large number of particles, the characterization being carried out *in situ* and almost *real time*. Recent applications of these methods to particle sizing can be found in Refs. [25, 28] for the DLS technique, in Refs. [21, 30] for ELS, and Refs. [14, 31] for MWT.

The goal of an efficient, well performing inversion algorithm is the accurate and fast recovery of the sample particle-size distribution over the largest possible range of particle radii. Indeed, inversion algorithms are expected to work pretty well only when the particle sizes to be recovered lie within a given range  $[R_{\min}, R_{\max}]$ , which depends on the range of the  $[\theta_{\min}, \theta_{\max}]$  independent variable being probed. Although what follows is of general validity, in this paper we will focus on the ELS technique, where  $I(\theta)$  represents the intensity of the light scattered by the sample at various scattering angles  $\theta$ . In the past we have shown [12] that the dynamical extension of the  $R$ -range (*i.e.*, the ratio  $[R_{\max}/R_{\min}]$ ) which can be probed, scales proportionally to  $[\theta_{\max}/\theta_{\min}]$  and, therefore can be rather limited if the latter ratio is not sufficiently large. Typically in the ELS technique,  $[R_{\max}/R_{\min}] \sim [\theta_{\max}/\theta_{\min}] \sim 10 - 100$ , see [12].

In order to obtain an inversion algorithm we have to discretize (1) and use a numerical method for solving the resulting linear system

$$AN = \mathbf{I}, \quad (2)$$

where  $\mathbf{N} \in \mathbb{R}^m$  and  $\mathbf{I} \in \mathbb{R}^n$  are the discretizations of  $N(R)$  and  $I(\theta)$ , respectively and  $A \in \mathbb{R}^{n \times m}$ . As mentioned above, since  $K$  is compact, the inverse problem in (1) is ill-posed. In the discrete version this property is inherited by the matrix  $A$  which is severely ill-conditioned, *i.e.*, its singular values decreases to zero rapidly with no significant gap [19]. In practice, since it is impossible to avoid the presence of noise in the data, the direct inversion of  $A$ , for instance be the Moore–Penrose pseudoinverse, leads to very poor approximations of the original desired  $\mathbf{N}$ . To recover a meaningful approximation of the true solution, a common approach is to resort to regularization, *i.e.*, use numerical methods that compute a solution of a nearby well-posed problem, see [3, 9] for classical regularization methods for inverse problems.

For large problems, a common approach is to adopt iterative methods that require only matrix vector-products. Among these iterative procedures, one of the most popular ones is undoubtedly the classical Lucy-Richardson (LR), known also as Richardson-Lucy method [26, 29], which has the remarkable feature of ensuring nonnegativity of the solutions. LR is also quite simple to implement, robust against noise and, provided that the iterative procedure is stopped after a properly chosen number of steps, does not require any parameter to be optimized. Furthermore, it is the specific nature of our problem that suggests the usage of the LR method, mainly because the noise present on the data is not additive, but multiplicative white Gaussian noise as described in Section 4.5, and the operator  $A$  is severely ill-conditioned and underdetermined. These features suggest that methods relying on the minimization of the Euclidean norm of the residual would not perform well on this specific problem. Moreover, in the considered setting, the knowledge of the norm of the noise is not available and only some statistical information on it can be retrieved, implying that the classical discrepancy principle cannot be employed. Thus, a method like LR seems to be more effective than other methods based on the minimization of the Euclidean norm.

On the other side, the LR method is known for being relatively slow and, as mentioned above and reported in [12]<sup>1</sup>, for being not so efficient when when the range of recoverable radii is too large. The latter limitation occurs because, for any given angular range  $[\theta_{\min}, \theta_{\max}]$ , as particles become smaller and smaller ( $R \ll \lambda/\theta_{\min}$ ), the angular distribution of the scattered light tends to be more and more uniform [ $I(\theta_{\min})/I(\theta_{\max}) \sim 1$ ], thus carrying less and less information on particle size. Conversely, for larger and larger particles ( $R \gg \lambda/\theta_{\max}$ ), the angular distribution varies widely with a high signal dynamic range [ $I(\theta_{\min})/I(\theta_{\max}) \gg 1$ ], but the shape of  $I(\theta)$  tends to be independent on particle size [23].

---

<sup>1</sup> After the publication of [12], it was realized that the method called “modification of the Chahine algorithm” proposed in that work, is identical to the LR algorithm.

In this paper we propose a new method called Weakly Constrained Lucy-Richardson (WCLR), which improves the classical LR method by adding a weak constraint to the true solution associated to a physical property (the particle volume fraction concentration) that is known or can be measured with rather high accuracy. The introduction of this additional knowledge accelerates remarkably the convergence rate of the method and, at the same time, largely improves the quality of the restoration. In particular it allows to enlarge the  $R$ -range by more than one order of magnitude, a property that is crucial in many real applications.

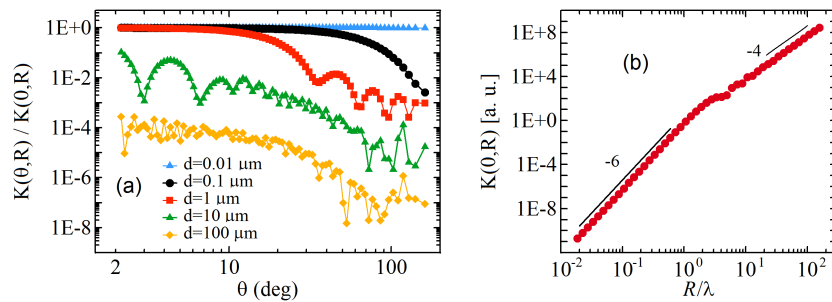
The price to pay for adding the constraint is the introduction of a damping parameter,  $\gamma$ , which has to be properly adjusted. However, this task is not critical at all because the strength of  $\gamma$  can be varied over a very broad range and its optimal value can be easily found by exploiting a second physical property (the particle number concentration), which needs to be known only very roughly. Note that using a *weak* rather than an *exact* constraint is a vital feature of the WCLR method. Indeed, while imposing an exact constraint has the advantage of requiring no damping parameter, its implementation (by using classical optimization algorithms for linear constrained problems) might still need the tuning of other parameters and, due to unavoidable errors always present in the estimate of the constraint, it does not provide necessarily an improvement in the quality of the restoration. Conversely, our WCLR is quite simple to implement, it is flexible and robust against small ( $\lesssim \pm 2\%$ ) uncertainties associated to the constraint and, except for  $\gamma$ , does not require any other parameter to be set in the same spirit of the classical LR.

This work is structured as follows. In Section 2 we give some physical details about the problem we are going to analyze. In particular, Section 2.1 describes the discretization process, while Section 2.2 considers the constraints we are going to use. Section 3 is devoted to the formulation of the mathematical model and to the definition of our numerical method. The latter one is tested and compared with LR on some numerical examples in Section 4. Section 5 is devoted to concluding remarks and future work.

## 2 Physical details

In this section we give some insight into the physical problem we are going to consider, *i.e.*, the problem associated to the inversion of elastic light scattering (ELS) data, where the main goal is to recover the size distribution of the particles present in the sample.

According to ELS theory [23], when a sample made of polydisperse particles characterized by a refraction index different from that of the surrounding medium is illuminated by a laser light of wavelength  $\lambda$ , part of the radiation is going to be scattered at angles different from the incident direction. If the particles are homogeneously dispersed in the medium and their concentration is so low that they can be considered as non-interacting, the angular distribution of the overall scattered intensity,  $I(\theta)$ , is given by the sum of the intensities



**Fig. 1** (a): normalized behavior of the kernel  $K(\theta, R)$  appearing in (1) as a function of the scattering angle  $\theta$  for five particles with diameters  $d = 2R$  ranging from  $d = 0.01 \mu\text{m}$  to  $d = 100 \mu\text{m}$ ; the normalization is such that  $K(\theta = 0, R) = 1$ . (b): behavior of the kernel amplitude  $K(\theta = 0, R)$  as a function of the ratio  $R/\lambda$ . The straight lines with slopes 6 and 4 indicate that  $K(\theta = 0, R)$  grows as  $\sim R^6$  or  $\sim R^4$ , the crossover occurring at  $R \sim \lambda$ .

scattered by the single particles [23]. Thus the system is linear and  $I(\theta)$  can be written as (1) where  $\theta$  is the scattering angle (the angle between the incident laser beam and the direction at which the scattered light is detected),  $N(R)$  is the unknown number-concentration density [ $\text{cm}^{-3} \mu\text{m}^{-1}$ ] of particles of radius  $R$ , and  $K(\theta, R)$  is the (known) kernel of the system, representing the intensity scattered by a single particle of radius  $R$  at angle  $\theta$ . Typically,  $I(\theta)$  is detected at a finite number of angles  $\theta_i$  ( $i = 1, \dots, n$ ) within a bounded interval  $[\theta_{\min}, \theta_{\max}]$ .

If the particles are spheres, the kernel  $K(\theta, R)$  is provided by the Mie theory [22], according to which the angular distribution of  $I(\theta)$  scattered by a particle of radius  $R$  is mostly confined to the diffraction lobe  $\theta_{\text{diff}} \sim \lambda/2R$  and  $I(\theta)$  amplitude strongly increases with particle radius as  $I \sim R^6$  for small particles ( $R \ll \lambda$ ) and  $I \sim R^4$  for large particles ( $R \gg \lambda$ ). Figure 1(a) reports an example of the behaviors of  $I(\theta)/I(0) = K(\theta, R)/K(0, R)$  versus  $\theta$  over a range of  $[2 - 180 \text{ deg}]$  for particles of different diameters  $d = 2R$  from  $d = 0.01$  to  $100 \mu\text{m}$ . As one can notice, for small particles  $I(\theta)$  tends to be rather flat, whereas for large particles  $I(\theta)$  exhibits many oscillations and decays by many order of magnitude over the reported  $\theta$  range. At the same time the zero-angle amplitude  $I(0)$  varies widely, passing from  $I(0) \sim 10^{-11}$  at  $R/\lambda = 10^{-2}$  to  $I(0) \sim 10^8$  at  $R/\lambda = 10^2$ , see Figure 1(b). Thus, it is clear that the inversion of (1) might become an unbearable task when the particle size distribution  $N(R)$  to be recovered contains particles with very different radii.

## 2.1 Discretization of the Fredholm Integral Equation

We now describe the discretization of (1). Let us consider that only a finite number of  $\theta_i$  ( $i = 1, \dots, n$ ) can be accessed experimentally and within a limited range  $[\theta_1, \theta_n]$ . Thus, if the particle size distribution  $N(R)$  is approximated by a histogram constituted by  $m$  bins (or classes) delimited by the radii  $r_j$ ,

$j = 0, 1, \dots, m$ , the equation (1) becomes

$$I(\theta_i) = \sum_{j=1}^m A_{i,j} N_j, \quad i = 1, 2, \dots, n, \quad (3)$$

where  $N_j$  is the number concentration density [ $cm^{-3} \mu m^{-1}$ ] of the particles belonging to the  $j$ -th class of width  $\Delta r_j = r_j - r_{j-1}$  and

$$A_{i,j} = \int_{r_{j-1}}^{r_j} K(\theta_i, r) dr. \quad (4)$$

When the classes are narrow enough, we can pinpoint them in terms of their average radius  $R_j = (r_j + r_{j-1})/2$  and width  $\Delta R_j := \Delta r_j$ , where  $\Delta r_j = r_j - r_{j-1}$ . Thus  $N_j \Delta R_j$  represents the number concentration of particles belonging to the  $j$ -th class and the term  $A_{i,j}/\Delta R_j$  is the average intensity scattered at angle  $\theta_i$  by a single particle with average radius  $R_j$ . Note that (3) is a set of  $n$  linear equations in which the left-hand side  $I(\theta_i)$  are the data provided by the experiment, the matrix entries  $A_{i,j}$  are known and  $N_j$  are the unknowns to be recovered.

Although somewhat arbitrary, it is often convenient to choose the  $r_j$  grid points so that, within the range  $[r_0, r_m]$ , all the  $m$  classes are characterized by the same relative width  $\alpha = \Delta R_j/R_j$ . This can be accomplished by scaling  $r_j$  according to the geometrical progression

$$r_j = r_0 a^j,$$

where  $a = (R_m/R_1)^{1/(m-1)}$  and  $r_0 = 2R_1/(1+a)$ . In this way the average radius and the width of each class scale as

$$R_j = R_1 a^{j-1}, \quad \Delta R_j = \Delta R_1 a^{j-1}, \quad j = 1, \dots, m, \quad (5)$$

thus  $\alpha = 2[(a-1)/(a+1)]$  and for  $a \gtrsim 1$ ,  $\alpha \approx a - 1$ . A sketch layout and discretization scheme of the classes is reported in Figure 2. Typically if we want to cover three orders of magnitude in size, *i.e.*  $R_m/R_1 = 10^3$  with  $\alpha = 0.02$ , approximately  $m = 350$  classes are necessary.

## 2.2 Constraints

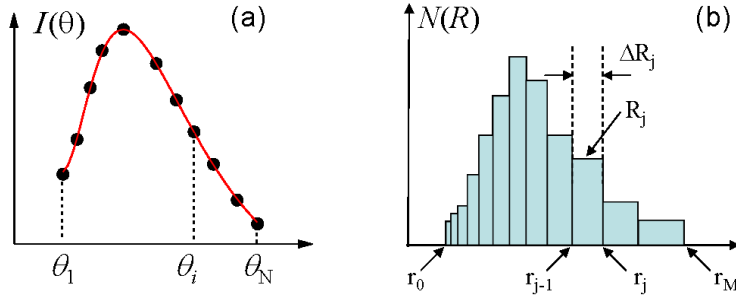
The very first constraint we can impose is that

$$N_j > 0, \quad j = 1, \dots, m, \quad (6)$$

This comes from the simple observation that the number of particle cannot be negative.

A more interesting constraint is related to the integral of  $N(R)$  obtaining

$$c_N = \sum_{j=1}^m N_j \Delta R_j, \quad (7)$$



**Fig. 2** Discretization scheme of equation (1)

where  $c_N$  represents the particle number concentration [ $cm^{-3}$ ], *i.e.*, the total number of particles contained in the sample divided by the sample volume. This constraint can be applied whenever the number of particles can be counted, a somewhat difficult task that can be carried out only under some experimental conditions.

The last constraint is related to the particle volume fraction concentration  $c_V$ , which is given by the total volume occupied by the particles divided by the sample volume

$$c_V = \sum_{j=1}^m N_j v_j \Delta R_j, \quad (8)$$

where

$$v_j = \frac{1}{\Delta R_j} \int_{R_j - \Delta R_j/2}^{R_j + \Delta R_j/2} (4/3)\pi R^3 dR \quad (9)$$

is the average volume of one particle belonging to the  $j^{th}$  class. Clearly, for very narrow classes ( $\Delta R_j/R_j \ll 1$ ),  $v_j \approx (4/3)\pi R_j^3$ .

*Remark 1* The constraint (8) on  $c_V$  is of particular significance because in most experiments the volume concentration is a quantity that can be measured quite easily and with high accuracy.

In our method we would like to exploit both concentration constraints (7) and (8) (the positiveness constraint, (6), being fulfilled automatically, see below), but they are not equivalent from a physical point of view. As mentioned out above, whereas an accurate value of  $c_V$  can be easily obtained experimentally, the estimate of  $c_N$  might be somewhat troublesome and affected by large errors. Thus, we propose a weakly constrained version of the LR algorithm based on the  $c_V$  constraint alone, and we will use the estimate of  $c_N$  only for cross-checking the self-consistency of the inversion procedure, *i.e.*, for estimating the damping parameter that weight the constraint on  $c_V$ .

### 3 The Weakly Constrained Lucy-Richardson method

We now describe how the physical model translates into the linear algebra language. The linear system (3) is compactly rewritten as

$$AN = \mathbf{I} \quad (10)$$

where  $A \in \mathbb{R}^{n \times m}$ ,  $\mathbf{N} \in \mathbb{R}^m$  and  $\mathbf{I} \in \mathbb{R}^n$ .

Similarly, the three constraints of Section 2.2 can be rewritten as:

- (i) From (6)  $\mathbf{N} > 0$ , meaning that  $N_j > 0$ , for  $j = 1, \dots, m$ ;
- (ii) From (7)

$$c_N = \mathbf{N}^t \Delta \mathbf{R}, \quad (11)$$

where  $0 < \Delta \mathbf{R} \in \mathbb{R}^m$ , with  $(\Delta \mathbf{R})_j = \Delta R_j$  defined in (5) for  $j = 1, \dots, m$ ;

- (iii) From (8)

$$c_V = \mathbf{N}^t \mathbf{V} = \sum_{j=1}^m V_j N_j. \quad (12)$$

where  $0 < \mathbf{V} \in \mathbb{R}^m$  and  $V_j = v_j \Delta R_j$ , being  $v_j$  defined in (9).

This information will be used in the following to define a simple and effective iterative procedure to compute a solution of (10), where, except from the positiveness (i), the constraints (ii)-(iii) are not all strictly satisfied, but are used to improve the computed approximation or to estimate possible parameters.

Note that equations (11) and (12) can be seen as weighted  $\ell_1$ -norm because both  $\mathbf{V}$  and  $\Delta \mathbf{R}$  are nonnegative, so we can define

$$\|\mathbf{N}\|_{1, \mathbf{V}} = \mathbf{N}^t \mathbf{V}, \quad \|\mathbf{N}\|_{1, \Delta \mathbf{R}} = \mathbf{N}^t \Delta \mathbf{R}.$$

As stated at the end of Section 2.2, we are going to use directly only a weighted version of the constraint (iii).

In order to insert the constraint (iii) we opportunely pad the matrix  $A$  and the right-hand side  $\mathbf{I}$ . Let  $\gamma > 0$  be a fixed real number and define

$$\varphi = \frac{\langle A_{i,j} \rangle}{\langle \mathbf{V}_j \rangle}, \quad (13)$$

where  $\langle A_{i,j} \rangle = \frac{1}{NM} \sum_{i,j} A_{i,j}$  and  $\langle \mathbf{V}_j \rangle = \frac{1}{N} \sum_j \mathbf{V}_j$  are the arithmetic averages of the entries of the matrix  $A$  and the vector  $\mathbf{V}$ , respectively. We define

$$\tilde{A} = \begin{pmatrix} A \\ \gamma \varphi \mathbf{V}^t \end{pmatrix} \quad \text{and} \quad \tilde{\mathbf{I}} = \begin{pmatrix} \mathbf{I} \\ \gamma \varphi c_V \end{pmatrix}. \quad (14)$$

Note that, since the factor  $\gamma \varphi V_j = \gamma \langle A_{i,j} \rangle (V_j / \langle \mathbf{V}_j \rangle)$  appearing in (14), must have the same dimensional units as  $\langle A_{i,j} \rangle$ ,  $\gamma$  is correctly a dimensionless parameter. As a consequence, its effect of (14) is independent on the units of both  $A_{i,j}$  and  $V_j$ .



In this way we have inserted the constraint (iii) in our system weighted by  $\gamma$ . The value of  $\gamma$  determines the strength of the constraint and its effectiveness. In particular the larger  $\gamma$  the stronger the effect of the constraint will be.

The new extended system becomes

$$\tilde{A}\mathbf{N} = \begin{pmatrix} A \\ \gamma\varphi\mathbf{V}^t \end{pmatrix} \mathbf{N} = \tilde{\mathbf{I}} = \begin{pmatrix} \mathbf{I} \\ \gamma\varphi c_V \end{pmatrix} \quad (15)$$

Since, the entries of the matrix  $\tilde{A}$  and  $\tilde{\mathbf{I}}$  are nonnegative and, according to (i), we are looking for a nonnegative solution of (15), the LR methods appears to be an excellent candidate for undertaking this task. Indeed, provided that the initial guess  $\mathbf{N}^0 > 0$ , the  $k+1$  approximated solution of equation (15) can be recursively written in term of the  $k$ -iterate as

$$\mathbf{N}^{k+1} = \frac{\mathbf{N}^k}{\mathbf{a}} \circ \left( \tilde{A}^t \cdot \frac{\tilde{\mathbf{I}}}{\tilde{\mathbf{I}}^k} \right), \quad k = 0, 1, \dots$$

where  $\frac{\cdot}{\cdot}$  and  $\bullet \circ \bullet$  are respectively the entry-wise division and multiplication and  $\bullet \cdot \bullet$  is the usual matrix-vector multiplication. The vector  $\mathbf{a} \in \mathbb{R}^m$  is defined as

$$a_j = \sum_{i=1}^{n+1} \left( \tilde{A}^t \right)_{j,i} = \sum_{i=1}^n A_{i,j} + \gamma\varphi V_j, \quad j = 1, \dots, m, \quad (16)$$

and  $\tilde{\mathbf{I}}^k$  is

$$\tilde{\mathbf{I}}^k = \tilde{A}\mathbf{N}^k = \begin{pmatrix} A\mathbf{N}^k \\ \gamma\varphi \|\mathbf{N}^k\|_{1,\mathbf{V}} \end{pmatrix}.$$

Let us consider the  $j$ -th component of  $\mathbf{N}^{k+1}$ :

$$N_j^{k+1} = \frac{N_j^k}{a_j} \left[ \tilde{A}^t \cdot \frac{\tilde{\mathbf{I}}}{\tilde{\mathbf{I}}^k} \right]_j = \frac{N_j^k}{a_j} \xi_j, \quad (17)$$

the factor  $\xi_j$  is

$$\xi_j = \sum_{i=1}^{n+1} \left( \tilde{A}^t \right)_{j,i} \frac{\left( \tilde{\mathbf{I}} \right)_i}{\left( \tilde{\mathbf{I}}^k \right)_i} = \sum_{i=1}^{n+1} \left( \tilde{A} \right)_{i,j} \frac{\left( \tilde{\mathbf{I}} \right)_i}{\left( \tilde{\mathbf{I}} \right)_i^k} = \sum_{i=1}^n A_{i,j} \frac{I_i}{I_i^k} + \gamma\varphi V_j \frac{c_V}{c_V^k}, \quad (18)$$

with  $c_V^k = \|\mathbf{N}^k\|_{1,\mathbf{V}}$ . Combining (16) and (18) with (17), we obtain

$$N_j^{k+1} = N_j^k \frac{\sum_{i=1}^n A_{i,j} \frac{I_i}{I_i^k} + \gamma \langle A_{i,j} \rangle \frac{V_j}{\langle \mathbf{V}_j \rangle} \frac{c_V}{c_V^k}}{\sum_{i=1}^n A_{i,j} + \gamma \langle A_{i,j} \rangle \frac{V_j}{\langle \mathbf{V}_j \rangle} V_j}, \quad j = 1, \dots, m. \quad (19)$$

where we have explicitly reported the expression for  $\varphi$  given in (13). We can see that constraint (iii) is not blended with the other term, it is decoupled from the data fitting part and is weighted by  $\gamma$ . Moreover, the nonnegativity of  $\mathbf{N}^k$  is simply preserved starting with  $\mathbf{N}^0 > 0$ , e.g.,  $\mathbf{N}^0 = \mathbf{1}$ , where  $\mathbf{1}$  represents the vector with entries all equals to 1.

### 3.1 Heuristic interpretation

We now want to give an heuristic interpretation of the formulation of (15). A standard approach for passing from a constrained least square problem to an unconstrained problem is the well-known quadratic penalization technique [5]. In our case, considering the constraint (iii), we obtain the minimization problem

$$\min_{\mathbf{N}} \|\mathbf{A}\mathbf{N} - \mathbf{I}\|_2^2 + (\gamma\varphi)^2 \left( \|\mathbf{N}\|_{1,\mathbf{V}} - c_V \right)^2, \quad (20)$$

where  $\|\cdot\|_2$  denotes the Euclidean norm and  $(\gamma\varphi)^2$  is the penalization parameter. This can be seen as a regularized version of the problem (2), where the parameter  $(\gamma\varphi)^2$  balances the trade off between the data fitting and the penalization term.

Define

$$\begin{aligned} \Psi(\mathbf{N}) &= \|\mathbf{A}\mathbf{N} - \mathbf{I}\|_2^2 + (\gamma\varphi)^2 \left( \|\mathbf{N}\|_{1,\mathbf{V}} - c_V \right)^2 \\ &= \mathbf{N}^t A^t \mathbf{A} \mathbf{N} - 2\mathbf{N}^t A^t \mathbf{I} + \mathbf{I}^t \mathbf{I} + (\gamma\varphi)^2 \left( \mathbf{N}^t \mathbf{V} \mathbf{V}^t \mathbf{N} - 2c_V \mathbf{N}^t \mathbf{V} + c_V^2 \right). \end{aligned}$$

The gradient of  $\Psi(\mathbf{N})$  is

$$\nabla \Psi(\mathbf{N}) = 2[A^t \mathbf{A} \mathbf{N} - A^t \mathbf{I} + (\gamma\varphi)^2 (\mathbf{V} \mathbf{V}^t \mathbf{N} - c_V \mathbf{V})].$$

Assume that  $V_j \neq 0$  for all  $j$ . Then  $\Psi(\mathbf{N})$  is coercive. In fact

$$\Psi(\mathbf{N}) \geq \gamma\varphi \left( \min_j \{V_j\} \mathbf{N}^t \mathbf{1} - c_V \right)^2 \rightarrow \infty \text{ as } \|\mathbf{N}\|_2 \rightarrow \infty.$$

Thus the minimum of  $\Psi$  exists, by Weierstrass theorem, and satisfies

$$\nabla \Psi(\mathbf{N}) = 0. \quad (21)$$

The functional  $\Psi$  is convex, but not strictly convex, since, in general,  $\mathcal{N}(A) \cap \mathcal{N}(\mathbf{V}^t) \neq \{\mathbf{0}\}$ , where by  $\mathcal{N}(A)$  we denote the null space of  $A$ . Thus, condition (21) is only necessary and the minimum is not unique. However, since the functional is coercive, we have that the minimum  $\mathbf{N}^*$  of  $\Psi$  satisfies  $\|\mathbf{N}^*\|_2 < \infty$ . In other words we have proven the following

**Proposition 1** *The functional  $\Psi$  defined above admits a global minimizer  $\mathbf{N}^*$  such that*

- (a)  $\|\mathbf{N}^*\|_2 < \infty$ ;
- (b)  $\nabla \Psi(\mathbf{N}^*) = \mathbf{0}$ .

Condition (21) can be rewritten equivalently as

$$(A^t \gamma\varphi \mathbf{V}) \begin{pmatrix} A \\ \gamma\varphi \mathbf{V}^t \end{pmatrix} \mathbf{N} = (A^t \gamma\varphi \mathbf{V}) \begin{pmatrix} \mathbf{I} \\ \gamma\varphi c_V \end{pmatrix}, \quad (22)$$

Recalling the definitions of  $\tilde{A}$  and  $\tilde{\mathbf{I}}$ , we can see that (22) is simply the normal equations of (15). This, coupled with (i), leads to the idea of looking for a nonnegative solution of (15).

### 3.2 Estimation of $\gamma$

The parameter  $\gamma$  weights the constraint (iii) on  $c_V$  defined in (12), but larger values of  $\gamma$  not necessarily lead to better restorations in practice (see Figures 4 and 6). This shows that the use of a constrained optimization algorithm does not necessarily provide better reconstructions.

The estimation of the optimal value of  $\gamma$ , *i.e.*, the one that minimizes the reconstruction error, can be somewhat tricky and the choice of  $\gamma$  is not straightforward, thus we propose to use an a posteriori strategy using the constraint (ii) on  $c_N$ .

Let us call  $\mathbf{N}_\gamma$  the nonnegative solution of (22) obtained with a certain choice of  $\gamma$  and suppose to know exactly  $c_N$ . Thus, we expect that the best choice for  $\gamma$  is the one that, beside providing the best reconstruction for  $\mathbf{N}_\gamma$ , minimizes also the error on  $c_N$ . Therefore, we choose  $\gamma = \gamma_{\text{opt}}$  such that

$$\gamma_{\text{opt}} = \arg \min_{\gamma} \{ |c_N - \|\mathbf{N}_\gamma\|_{1, \Delta \mathbf{R}}| \}. \quad (23)$$

In practice, as we will see in Section 4, the choice rule (23) is not so strict because there is a large range of  $\gamma$ -values around  $\gamma_{\text{opt}}$  where the reconstruction is equally good and even a value of  $\gamma$  very far from  $\gamma_{\text{opt}}$  would provide accurate results. This feature is of fundamental importance because, whenever the constraint  $c_N$  is not known and can be only roughly estimated (as it might happen experimentally),  $\gamma_{\text{opt}}$  cannot be determined with high accuracy and the condition (23) would be inapplicable.

To obtain an estimation of  $\gamma_{\text{opt}}$  we compute  $\mathbf{N}_\gamma$  for a small set of possible  $\gamma$  and select the one that minimizes  $|c_N - \|\mathbf{N}_\gamma\|_{1, \Delta \mathbf{R}}|$ .

Summarizing, our *weakly constrained LR (WCLR)* algorithm is the following:

1. fix  $\mathbf{N}^0 = \mathbf{1}$  and a small set of possible values for  $\gamma$ ;
2. compute  $\mathbf{N}_\gamma$  for every  $\gamma$ ;
3. choose the solution  $\mathbf{N}_{\gamma_{\text{opt}}}$  corresponding to  $\gamma_{\text{opt}}$  defined in (23).

## 4 Numerical Examples

In this section we report some numerical examples aimed at ascertaining how our algorithm performs against the classical LR method, *i.e.*, when  $\gamma = 0$ . We will also show how to find the optimal value  $\gamma_{\text{opt}}$ , consistently with what described above.

### 4.1 Generation of matrix A

The  $n \times m$  matrix  $A$  was computed by numerically integrating equation (4) with the kernel  $K$  provided by the Mie theory [22] and illustrated in Figure 1. The number of angles was  $n = 100$ , scaled according to a geometrical progression

with  $\theta_{\min} = 2 \text{ deg}$  and  $\theta_{\max} = 180 \text{ deg}$ . The bins for the recovered distribution were also chosen accordingly to a geometrical progression (see Section 2.1) with  $R_{\min} = 10^{-3} \mu\text{m}$  and  $R_{\max} = 10^3 \mu\text{m}$  and their number was  $m = 600$ . In this way all the bins were characterized by the same relative width  $\Delta R_j/R_j \approx 0.023$ .

#### 4.2 Generation of artificial test data

The generation of artificial test data was carried out by supposing to know a true solution for the system  $\mathbf{N}_{\text{true}}$  and computing the noise-free data as

$$\mathbf{I}_{\text{true}} = \mathbf{A}\mathbf{N}_{\text{true}}.$$

Then the real data were obtained by adding to  $\mathbf{I}_{\text{true}}$  a (fractional) random white Gaussian noise, so that the noise level is proportional to  $\mathbf{I}_{\text{true}}$  and independent from point to point. If we indicate with  $\epsilon$  the fractional noise level, the noisy  $\mathbf{I}$  becomes

$$\mathbf{I} = \mathbf{I}_{\text{true}} \circ (\mathbf{1} + \epsilon \mathbf{e}), \quad (24)$$

where  $\mathbf{e}$  is a vector whose entries are realizations of a random variable such that  $(\mathbf{e})_j \sim \mathcal{N}(0, 1)$ . Typical values for  $\epsilon$  are in the interval  $[10^{-3}, 10^{-2}]$ .

#### 4.3 Inversion procedure and stopping criteria

The artificial test data in (24) were inverted by fixing a value of the parameter  $\gamma$  and using the iterative algorithm given by (19). The iterative procedure was stopped according to the following criteria:

- (1) First of all we impose a minimum number of iterations  $k_{\min} = 10^5$ , which is necessary for the WCLR (or the LR) algorithm to work properly. The minimum number of iterations  $k_{\min}$  was estimated by inverting noiseless data and checking that  $10^5$  iterations are enough for retrieving the expected distribution with high accuracy ( $RRE \lesssim 10^{-2}$ , see below, Section 4.4). We also checked that for noisy data with noise levels typical of our problem ( $\sim 1\%$ ) when the procedure would prefer to stop at a number of iterations  $k < k_{\min}$  (see below points 3. and 4.), forcing it to continue up to  $k_{\min} = 10^5$  does not spoil significantly the quality of the distribution.
- (2) We also impose a maximum number of iterations  $k_{\max} = 10^6$ , which ensures that the inversion stops even when the criteria below reported are not met.
- (3) For any  $k_{\min} < k < k_{\max}$  the procedure is stopped accordingly to a *relative discrepancy principle*, which is a modified version of the classical discrepancy principle, largely used with iterative regularization methods [9]. As relative discrepancy parameter we define the quantity

$$\epsilon_k = \frac{1}{\sqrt{n}} \left\| \frac{\mathbf{I}^k - \mathbf{I}}{\mathbf{I}} \right\|_2, \quad (25)$$

where  $\mathbf{I}^k = A\mathbf{N}^k$  and the division is component-wise. The value of  $\epsilon_k$  in (25) corresponds to the mean relative standard deviations between the input noisy data (24) and the ones recovered after  $k$  iterations. Note that  $\epsilon_k$  differs from the classical parameter used in the discrepancy principle ( $\|\mathbf{I}^k - \mathbf{I}\|_2$ ) because, aside for the constant term  $1/\sqrt{n}$ , it weights the relative and not the absolute square deviations between input and reconstructed data. This choice was enforced by the fact that the noise added to the data was chosen to be proportional to  $\mathbf{I}_{\text{true}}$  and, therefore only an  $\epsilon_k$  defined as in (25) ensures that all the data points are equally weighted. Moreover, the exact noise level  $\epsilon\|\mathbf{I}_{\text{true}} \circ \mathbf{e}\|_2$  of the single experiment is unknown and hence the classical discrepancy principle cannot be used. According to (25), the algorithm was therefore stopped after  $k = k^*$  iterations so that

$$\epsilon_{k^*} \leq \epsilon, \quad (26)$$

where  $\epsilon$  is the fractional noise level introduced in (24).

- (4) Finally whenever, for  $k_{\min} < k < k_{\max}$ , condition (26) is not met but  $\epsilon_k$  attains a minimum and begins to increase (the sequence  $\{\epsilon_k\}$  decreases for small values of  $k$ , but is not always monotonic decreasing for large  $k$ ), the procedure is stopped in correspondence of the minimum.

Since the value of  $\epsilon_k$  can oscillate, we say that we have reached a minimum for  $\epsilon_k$  if  $\epsilon_k < \epsilon_{k+1} < \epsilon_{k+2} < \dots < \epsilon_{k+10}$ , *i.e.*, if  $\epsilon_k$  has increased for 10 consecutive iterations.

When such a minimum is reached before  $k_{\min}$ , the procedure is stopped at  $k = k_{\min}$ .

In the following we will denote with  $K_{\text{it}}$  the number of iterations at which the procedure is actually stopped and the corresponding  $\epsilon_{K_{\text{it}}}$  will be called Mean Relative Standard Deviation (*MRSD*).

#### 4.4 Evaluation parameters

The accuracy of the inversion algorithm was evaluated by comparing the retrieved distribution with the true one. However, since from a physical point of view, volume (or mass) distributions are much more significant than number distributions, we compared retrieved and true distributions on the basis of volume-fraction density distributions, defined as

$$\phi(R) = N(R)v(R),$$

here  $\phi(R)$  has the dimensions of  $[\mu m^{-1}]$ .

For assessing the accuracy of the inversion procedure, we define a  $\gamma$ -dependent Relative Restoration Error (*RRE*) as

$$RRE(\gamma) = \frac{\|\phi_\gamma - \phi_{\text{true}}\|_2}{\|\phi_{\text{true}}\|_2},$$

which corresponds to the relative average root mean square (*r.m.s.*) deviations between the retrieved and true mass distributions.

Similarly, for assessing the accuracy on the reconstruction of the two parameters  $c_N$  and  $c_V$  that characterize the true distribution, we define the quantities

$$D_N(\gamma) = \frac{\left| \|\mathbf{N}_\gamma\|_{1,\Delta\mathbf{R}} - c_N \right|}{c_N} \quad \text{and} \quad D_V(\gamma) = \frac{\left| \|\mathbf{N}_\gamma\|_{1,\mathbf{V}} - c_V \right|}{c_V},$$

which represent the relative errors between  $c_N$  and  $c_V$  and the corresponding recovered parameters.

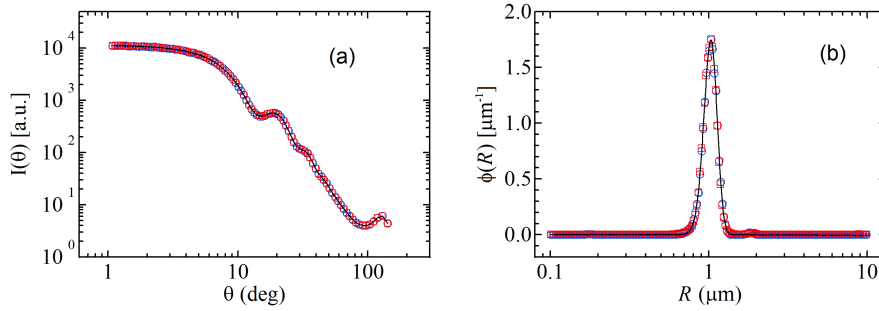
#### 4.5 Numerical results

In the following numerical tests, the WCLR and LR algorithms were compared for several values of  $\gamma$  in a given range of recoverable radii. The latter one was chosen to be a subset of the original range defined in Subsection 4.1, so that  $R_{\max}/R_{\min} = 100$ . For each  $\gamma$ , the tests were repeated 30 times, with different noise realizations (of the same  $\epsilon$  level).

*Test 1* In the first test we show that, when the distribution to be recovered is characterized by particles that produce signals whose features are: (a) asymptotically constant at low angles and (b) exhibit a dynamic range between first and last angle of several order of magnitudes (see  $d = 0.1 \mu m$  or  $d = 1.0 \mu m$  curves in Figure 1(a)), the original LR and our WCLR algorithms are quite equivalent. To this aim, we selected as true number distribution  $\mathbf{N}_{\text{true}}$  a Gaussian centered in the middle of the recoverable range  $[R_{\min}, R_{\max}]$ , *i.e.*, with an average value  $\langle R \rangle = 1 \mu m$  and standard deviation  $\sigma_R = 0.1 \mu m$ . The particle number concentration was (arbitrarily) chosen to be  $c_N = 10^{16} \text{ cm}^{-3}$  and the *r.m.s.* noise level added to the data was  $\epsilon = 0.01$ . The inversion was carried out by trimming the recoverable particle radii in the range  $[0.1 \mu m - 10 \mu m]$ , so that number of bins was  $m = 200$ . Correspondingly, the matrix used in this test was obtained by selecting a proper subset ( $100 \times 200$ ) of the original ( $100 \times 600$ ) matrix computed at section 4.1. The value of  $\gamma$  ranges from zero (original LR) to  $\gamma = 10^6$  and  $K_{\text{it}} = 10^6$ .

The findings of this test show that our algorithm performs equally well independently of the  $\gamma$ -value ( $0 - 10^4$ ) and its performances were quite similar to the ones provided by the original LR algorithm. This is shown in Figure 3 where data reconstructions and average recovered distributions are highly accurate and, as matter of fact, indistinguishable between our algorithm (run with  $\gamma = 1$ ) and the classical LR algorithm ( $\gamma = 0$ ).

*Test 2* The effective difference between the two algorithms becomes evident only when the particles are close to the boundaries of the  $[R_{\min}, R_{\max}]$  range. For this second test we selected a Gaussian distribution characterized by large

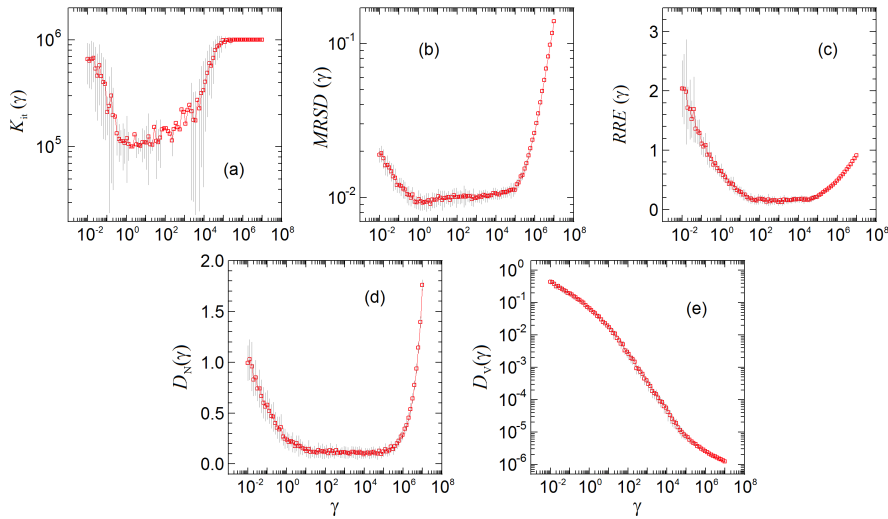


**Fig. 3** Test 1: Comparison between the original LR (blue symbols) and our WCLR algorithm (red symbols) run with  $\gamma = 1$  for a Gaussian distribution with an average value  $\langle R \rangle = 1 \mu\text{m}$  and standard deviation  $\sigma_R = 0.1 \mu\text{m}$ . (a) Reconstructed (symbols) and true (line) signals. (b) Reconstructed (symbols) and true (line) distributions. The two algorithms performs equally well and, as matter of fact, both results (a) and (b) are almost indistinguishable.

particles, *i.e.*, with  $\langle R \rangle = 100 \mu\text{m}$  and  $\sigma_R = 10 \mu\text{m}$ . The noise level added to the data was  $\epsilon = 0.01$ , the inversion was carried out in the range  $[10 \mu\text{m} - 1000 \mu\text{m}]$  (so that  $m = 200$ ), and  $\gamma$  was varied in the range  $[10^{-2} - 10^7]$ .

Differently from the first test, the effect of changing  $\gamma$  is quite relevant, as shown in Figure 4 where the behaviors of the parameters  $K_{\text{it}}$ , which we recall that denotes the number of iterations, (a),  $MRSD$  (b),  $RRE$  (c),  $D_N$  (d), and  $D_V$  (e) are reported as a function of  $\gamma$ . First of all we observe that there is a broad range of  $\gamma \sim [10^{-1} - 10^5]$  where the inversion procedure was stopped at a number of iterations  $K_{\text{it}} \leq 10^6$  (see Figure 4(a)). Correspondingly, within this range, the reconstruction parameter  $MRSD$  (see Figure 4(b)) was comparable with the noise level added to the data, meaning that the iterative procedure was much faster and stopped accordingly to the relative discrepancy principle (26).

Then, whereas  $D_V$  (see Figure 4(d)) decreases monotonically with increasing  $\gamma$  (which is consistent with the fact that the stronger the constraint, the higher the accuracy of its recovered value), the parameters describing the accuracy of the retrieved distribution,  $RRE$  (see Figure 4(c)) and the accuracy on the recovered number concentration,  $D_N$  (see Figure 4(e)), exhibit very broad valleys whose flat regions cover almost the same range of  $\gamma \sim [10^1 - 10^5]$ . Thus, the choice of an optimal value for  $\gamma$  is not critical at all and any value chosen in the central part of this interval (for example  $\gamma_{\text{opt}} \sim [10^2 - 10^4]$ ) leads both to small errors in the recovery of  $c_N$  and to very accurate distribution reconstructions, as shown in Figure 5(e-f-g). Conversely, for values of  $\gamma$  outside this range, the recovery of  $c_N$  becomes increasingly less and less accurate and, at the same time, the distributions are more and more poorly recovered, as shown in all the other panels of the figure. In particular, we would like to point out the remarkable mismatching between the true distribution and the one recovered in Figure 5(a), showing that the classical LR algorithm ( $\gamma = 0$ ) is totally unable to perform such a task.



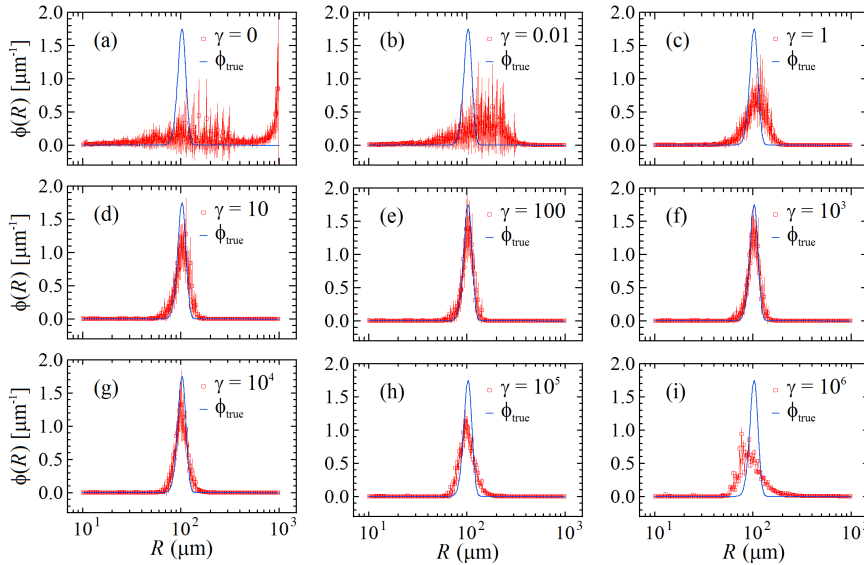
**Fig. 4** Test 2: Behavior as a function of  $\gamma$  of the average parameters  $K_{it}$  (number of iterations) (a),  $MRSD$  (b),  $RRE$  (c),  $D_N$  (d), and  $D_V$  (e) for a Gaussian distribution with  $\langle R \rangle = 100 \mu m$  and  $\sigma_R = 10 \mu m$ . The error bars are the standard deviations associated to the various parameters deriving from the noise ( $\epsilon = 0.01$ ) present on the data (error bars larger than the data points are not displayed).

Finally, we would like to point out that the rather similar behaviors between  $RRE$  and  $D_N$  guarantees that, in a real experiment where the parameter  $RRE$  cannot be measured because  $\phi_{true}(R)$  is not known, the optimal range for  $\gamma_{opt}$  can be inferred by looking at the behavior of  $D_N$  (Figure 4(d)). The fact that such a range is remarkably broad ( $\sim 2 - 3$  orders of magnitude) ensures that even a huge uncertainty on the value of  $c_N$  would not affect significantly the accuracy with which  $\phi_{true}(R)$  will be reconstructed.

*Test 3* In this test we selected a distribution close to the left side of the range, namely a Gaussian with  $\langle R \rangle = 0.05 \mu m$  and  $\sigma_R = 0.005 \mu m$ . In this case, as shown in Figure 1, the  $I(\theta)$  data are rather flat; therefore, for not corrupting completely their behavior, the noise level added to the data was  $\epsilon = 0.001$ . The inversion was carried out in the range  $[0.005 \mu m - 0.5 \mu m]$  (so that  $m = 200$ ), and  $\gamma$  was varied in the range  $[10^{-7} - 10^2]$ .

Figure 6 shows the behavior of the parameters  $K_{it}$  (Number of iterations) (a),  $MRSD$  (b),  $RRE$  (c),  $D_N$  (d), and  $D_V$  (e), whereas the comparison between the recovered distribution and  $\phi_{true}(R)$  is shown in Figure 7 varying  $\gamma$  in the range  $[10^{-6} - 10^3]$ . As for the large particles, there is an even larger optimal range  $\gamma_{opt} \sim [10^{-3} - 10^1]$ , where convergence rate is faster and both distribution reconstruction and  $c_N$  recovery are very accurate. We recall that  $\gamma_{opt}$  can be safely identified by the criterion (23) thanks to the very broad and shallow minimum region of the curve of  $D_N$  in Figure 6(d). Finally, note





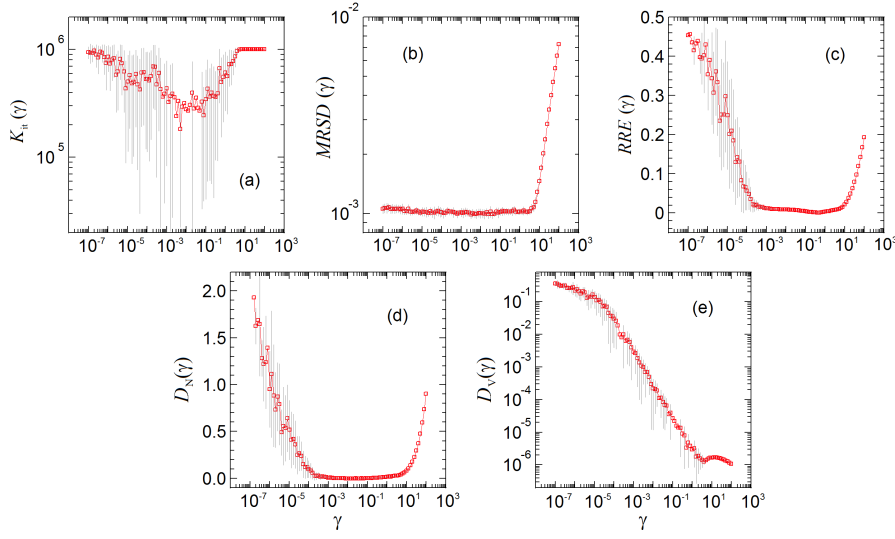
**Fig. 5** Test 2: Comparison between the average recovered distributions (red symbols) obtained with our algorithm at various  $\gamma$ -values ( $\gamma = 0$  is LR) and a true Gaussian distribution  $\phi_{\text{true}}(R)$  with  $\langle R \rangle = 100 \mu\text{m}$  and  $\sigma_R = 10 \mu\text{m}$ . The error bars are the standard deviations associated to the bins of the recovered histogram due to the noise ( $\epsilon = 0.01$ ) presents in the data.

that, also in this case, the original LR algorithm is totally unable to recover the true distribution (see Figure 7(a)).

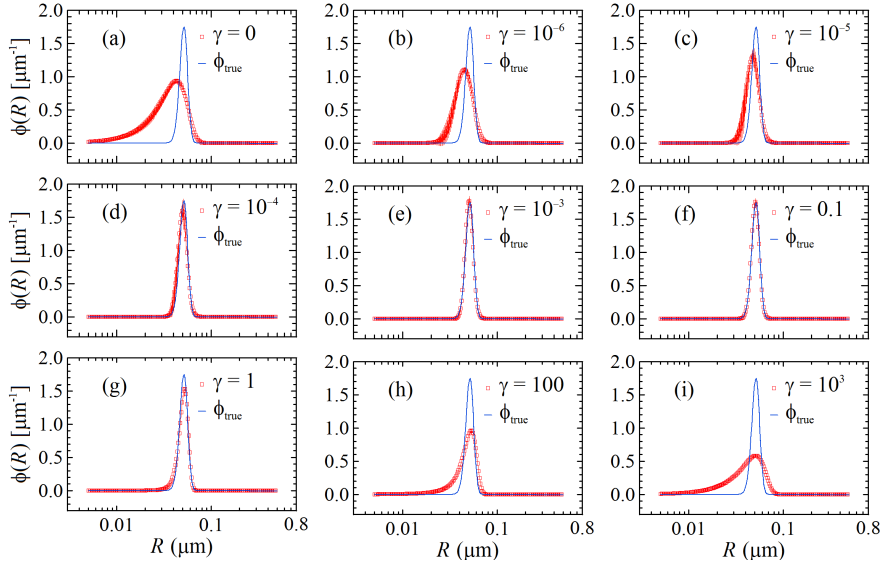
#### 4.6 Errors on the $c_V$ constraint

In this section we report an example of the effects of small systematic errors on the constraint  $c_V$  with respect to the quality of the recovered distributions. We investigated the specific case of the true distribution used in Figure 7, *i.e.*, a Gaussian distribution with  $\langle R \rangle = 0.05 \mu\text{m}$  and  $\sigma_R = 0.005 \mu\text{m}$ , and we supposed to have errors  $\delta c_V$  varying within a range of  $\pm 10\%$ . Then for any fixed value of  $\gamma$  and we investigated how the recovered distributions deteriorates as  $\delta c_V$  is increased. This analysis was carried out following the same procedure described for the previous tests, and was repeated for 8 different values of  $\gamma$  spanning the range  $\gamma \sim [10^{-5} - 10^2]$  (1 point per order of magnitude).

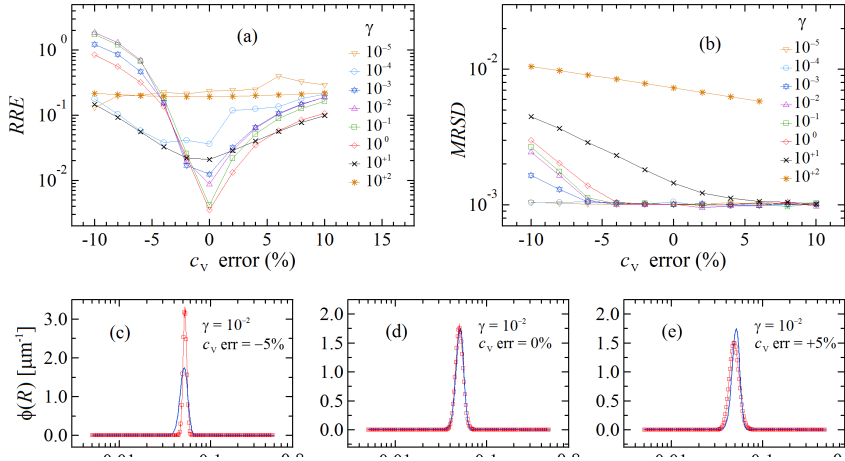
The main result is summarized in the first panel of Figure 8, where the behavior of the parameter  $RRE$  is reported as a function of the percentage error  $\delta c_V$ . As expected, within the optimal range  $\gamma_{\text{opt}} \sim [10^{-3} - 10^1]$ , the minimum of  $RRE$  occurs at zero error and increases as  $\delta c_V$  becomes larger and larger, but as long as  $\delta c_V \lesssim \pm 2\%$ , its value remains  $\lesssim 2 - 3 \times 10^{-2}$ . This trend is clearly not followed by values of  $\gamma$  that are outside this optimal range because, even at zero error as shown in Figure 6(a), the inversion algorithm



**Fig. 6** Test 3: Behavior as a function of  $\gamma$  of the average parameters  $K_{it}$  (a),  $MRS D$  (b),  $RRE$  (c),  $D_N$  (d), and  $D_V$  (e) for a Gaussian distribution with  $\langle R \rangle = 0.05 \mu m$  and  $\sigma_R = 0.005 \mu m$ . with  $\langle R \rangle = 100 \mu m$  and  $\sigma_R = 10 \mu m$ . The error bars are the standard deviations associated to the various parameters deriving from the noise ( $\epsilon = 0.001$ ) present on the data (error bars larger than the data points are not displayed).



**Fig. 7** Test 3: Comparison between the average recovered distributions (red symbols) obtained with our algorithm at various  $\gamma$ -values ( $\gamma = 0$  is LR) and a true Gaussian distribution  $\phi_{true}(R)$  with  $\langle R \rangle = 0.05 \mu m$  and  $\sigma_R = 0.005 \mu m$ . The error bars are the standard deviations associated to the bins of the recovered histogram due to the noise (a  $\epsilon = 0.001$ ) present on the data.



**Fig. 8** Effects of systematic errors on the  $c_V$  constraint with respect to the quality of the recovered distributions in the case of a true Gaussian distribution  $\phi_{\text{true}}(R)$  with  $\langle R \rangle = 0.05 \mu\text{m}$  and  $\sigma_R = 0.005 \mu\text{m}$ : (a) Behavior of  $RRE$  versus the percentage error  $\delta c_V$ ; (b) behavior of  $MRSD$  versus the percentage error  $\delta c_V$ ; (c)-(d)-(e): comparison between the true distribution (solid line) and the average recovered distributions (red symbols) obtained with  $\gamma = 10^{-2}$  and  $-5\%$ ,  $0\%$  and  $+5\%$  percentage errors. The standard deviations associated to the bins of the recovered histograms are due to the noise ( $\epsilon = 0.001$ ) present on the data.

is not able to work properly. The other important parameter describing the performances of the inversion algorithm in presence of systematic errors on  $c_V$  is the behavior of  $MRSD$  (see Figure 8(b)). As shown, within the optimal range  $\gamma_{\text{opt}} \sim [10^{-3} - 10^1]$  and for errors  $\lesssim \pm 5\%$ , all the  $MRSD$  values are compatible with the fractional statistical errors on the data ( $\epsilon = 10^{-3}$ ). Outside these ranges,  $MRSD$  tends to increase because convergence is slowed down and  $10^6$  iteration are not sufficient or because signal reconstructions are very poor whatsoever.

In the last three panels of Figure 8 we report an example of the distributions recovered for  $\gamma = 10^{-2}$  at errors on  $c_V$  equal to  $-5\%$ ,  $0\%$  and  $+5\%$ . As expected, the central distribution (panel d,  $0\%$  error) matches very accurately the input ones ( $RRE \approx 8.7 \times 10^{-3}$ ), whereas the other two are clearly less accurate, being  $RRE \approx 3.2 \times 10^{-1}$  for the  $-5\%$  error and  $RRE \approx 8.8 \times 10^{-2}$  for the  $+5\%$  error, but still very meaningful from a physical point of view.

## 5 Conclusions

In this paper we have proposed a simple and effective variant of the LR method by adding a weak constraint to be imposed as a penalty term on the recovered solution. The new method, called Weakly Constrained Lucy-Richardson (WCLR), was applied and tested on the inversion of simulated elastic light scattering data, whose aim is to recover the number-size,  $N_{\text{true}}(R)$ , or volume-

size,  $\phi_{\text{true}}(R)$ , distributions of the particles composing the sample over the largest possible interval of particle sizes.

For this problem, there are two physical constraints related to the knowledge of the particle number,  $c_N$ , and volume,  $c_V$ , concentrations, which correspond, mathematically, to the integral of  $N_{\text{true}}(R)$  and  $\phi_{\text{true}}(R)$ , respectively. Both constraints can be determined experimentally but, whereas  $c_V$  can be measured with high accuracy, the estimation of  $c_N$  can be affected by large errors. Thus in this paper we used as a constraint only  $c_V$  and exploited  $c_N$  for determining a *a posteriori* criterion for adjusting the strength,  $\gamma$ , of  $c_V$ .

By the use of extensive numerical simulations, we have shown that the optimal value  $\gamma_{\text{opt}}$ , *i.e.*, the one that provides the highest accuracies in the recovery of both  $c_N$  and  $\phi_{\text{true}}(R)$  (or  $N_{\text{true}}(R)$ ), is not critical at all because there is a broad range ( $\sim 2 - 4$  orders of magnitude) of  $\gamma$ -values around  $\gamma_{\text{opt}}$  where both restorations are equally good. Thus, in a real experiment where  $\phi_{\text{true}}(R)$  is not known, the optimal  $\gamma_{\text{opt}}$  can be inferred only by looking at  $c_N$  and, thanks to the fact that the  $\gamma$ -range is so wide, even a huge uncertainty on the value of  $c_N$  would not affect significantly the accuracy on the recovered distributions.

We have also provided a new protocol for stopping the iterative procedure, which is based on the four criteria described in Sect. 4.3. Although such criteria rely on the choice of the two parameters  $k_{\text{min}}$  and  $k_{\text{max}}$  which force the procedure to stop at  $k_{\text{it}}$  so that  $k_{\text{min}} < K_{\text{it}} < k_{\text{max}}$ , the actual values of  $K_{\text{it}}$  very rarely are cast on the extremes of such a range. Thus, the particular choice for  $k_{\text{min}}$  and  $k_{\text{max}}$  does not affect the results of the method. It must be said, however, that the tuning of  $k_{\text{min}}$  is a little bit tricky because convergence depends on the features of the distribution to be recovered and, at the same time, on the noise level present on the data. In our case where such a level is  $\sim 1\%$  or less, we found that for not too narrow bell shaped distributions (relative width  $\sigma / \langle R \rangle \sim 10\%$  or larger) the use of a single value  $k_{\text{min}} = 10^5$  was a reasonable compromise. In the extremely rare case that the best restoration is obtained at an iteration lower than  $k_{\text{min}}$ , we do not observe the insurgence of artifacts, or they are at least fairly weak, because WCLR algorithm has a very stable convergence, especially against reasonable levels ( $\sim 1\%$ ). In any case, for higher noise levels, a preprocessing denoising procedure could always be applied. In conclusion therefore, although  $k_{\text{min}}$  could have been optimized from case to case, the single value  $k_{\text{min}} = 10^5$  allowed us to recover accurately distributions spanning the entire  $[R_{\text{min}}, R_{\text{max}}]$  range, without introducing a further adjusting parameter in the procedure.

Our numerical simulations show that, when the particle sizes to be recovered lie within a range extension comparable with that of the measurements, WCLR and LR are fairly equivalent: they are both simple, robust and, provided that the iterative procedure is stopped after a large enough number of steps, equally accurate. The only slight difference is that, whereas LR truly requires no adjusting parameter, in the case of WCLR, the only adjustable parameter would be  $\gamma$ , but as mentioned above, it can be varied over such a large range of values that, as a matter of fact, does not need to be optimized.

When the two algorithms are tested over a wider range of particle sizes, WCLR permits to extend widely the range over which the distributions can be reliably recovered with respect to LR. In particular, by using the LR algorithm we would not be able to recover particle size distributions with particles as large as  $100\mu m$  or as small as  $50nm$ . This is clearly shown in panels (a) of Figs. 5 and 6. In the specific case of elastic light scattered data inversion investigated in this work where the probed scattering angles were between  $[2^\circ - 180^\circ]$ , such a range can be stretched more than one order of magnitude, passing from  $[R_{\min} - R_{\max}] \sim [0.1 - 10] \mu m$  to  $\sim [0.05 - 100] \mu m$ .

In Subsection 4.6 we also investigated the effect of having small systematic errors on the knowledge of the constraint  $c_V$  with respect to the quality of the reconstructed distributions. Our findings show that, as long as the error on  $c_V$  is within  $\pm 2\%$ , the quality of the reconstructed distributions is well preserved and the use of a weak constraint is advantageous with respect to a strong constraint. This feature is fundamental in practical applications where, typically, the relative accuracy on  $c_V$  is of the order of a few percents.

As we stated in the Introduction the nature of the problem itself suggested that the LR algorithm would have been a good choice, while other methods, in particular the ones which rely on the minimization of the Euclidean norm, would have been ineffective.

The main drawback of the proposed approach is still the requirement of a fairly high number of iterations, even though we have shown that the number of iterations required by WCLR is much smaller than the one required by the standard LR method. However, the dimension of the problem at hand, while not small, is not too large (as seen in the examples in Section 4) and thus the overall computation effort is limited even when a high number of iterations is required.

It will be subject of future studies the combination of the techniques proposed in this paper with more efficient methods from the literature, starting from the work in [8, 27]. In particular, it would be interesting to insert the methods from [13] in our framework, as well as to combine our approach with Modulus-Method for ill-posed problems, see [1] and references therein.

Finally, we would like to point out that variants of the classical Lucy-Richardson, like those in [4, 24], could be useful for further improving our algorithm. We would also like to stress that, although in this article our WCLR method has been tested and its performances ascertained for the specific optical problem related to the inversion of the elastic light scattering data, the proposed method can be in principle applied to any other problem where some constraints associated to integrals of the distribution to be recovered are known. Examples of such problems are again in the fields of optics with the technique known as multi-spectral extinction turbidity [10, 11] or in the field of stereology, an interdisciplinary methodology that is concerned with the recovering of the 3D properties of a sample from its 2D sections (see for example [20] and references therein). Work is in progress for extending WCLR to these problems.

## 6 Acknowledgments

The author would like to thank the reviewers for their insightful comments that greatly improved the readability and the overall quality of this work. The work of the first two authors is supported in part by MIUR - PRIN 2012 N. 2012MTE38N and by a grant of the group GNCS of INdAM.

## References

1. Bai, Z.Z., Buccini, A., Hayami, K., Reichel, L., Ying, J.F., Zheng, N.: Modulus-based iterative methods for constrained Tikhonov regularization. *Journal of Computational and Applied Mathematics* **319**, 1–13 (2017)
2. Baltes, H.P.: Inverse scattering problems in optics. Springer-Verlag, Berlin (1980)
3. Bertero, M., Boccacci, P.: Introduction to inverse problems in imaging. CRC press (1998)
4. Biggs, D.S.C., Andrews, M.: Acceleration of iterative image restoration algorithms. *Appl. Opt.* **36**(8), 1766–1775 (1997)
5. Björck, Å.: Numerical methods for least squares problems. Siam (1996)
6. Bohren, C.F., Hirdleman, E.D.: Feature on optical particle sizing. *Appl. Opt.* **30**, 4685–4987 (1991)
7. Brown, W.: Dynamic Light Scattering. Clarendon, Oxford (1993)
8. Calvetti, D., Landi, G., Reichel, L., Sgallari, F.: Non-negativity and iterative methods for ill-posed problems. *Inverse Problems* **20**(6), 1747 (2004)
9. Engl, H.W., Hanke, M., Neubauer, A.: Regularization of inverse problems, vol. 375. Springer Science & Business Media (1996)
10. Ferri, F., Bassini, A., Paganini, E.: Modified version of the Chahine algorithm to invert spectral extinction data for particle sizing. *Applied Optics* **34**(25), 5829–5839 (1995)
11. Ferri, F., Bassini, A., Paganini, E.: Commercial spectrophotometer for particle sizing. *Applied optics* **36**(4), 885–891 (1997)
12. Ferri, F., Righini, G., Paganini, E.: Inversion of low-angle elastic light-scattering data with a new method devised by modification of the Chahine algorithm. *Applied optics* **36**(30), 7539–7550 (1997)
13. Gazzola, S., Nagy, J.G.: Generalized arnoldi–tikhonov method for sparse reconstruction. *SIAM Journal on Scientific Computing* **36**(2), B225–B247 (2014)
14. Glasse, B., Riefler, N., Fritsching, U.: Intercomparison of numerical inversion algorithms for particle size determination of polystyrene suspensions using spectral turbidimetry. *Journal of Spectroscopy* **2015**, 645,879 (2015)
15. Glatter, O.: Fourier transformation and deconvolution. In: P. Linder, T. Zemb (eds.) Neutrons, X-rays and light: Scattering methods applied to soft condensed matter. North Holland, Elsevier (2002)
16. Glatter, O.: The inverse scattering problem in small angle scattering. In: P. Linder, T. Zemb (eds.) Neutrons, X-rays and light: Scattering methods applied to soft condensed matter. North Holland, Elsevier (2002)
17. Glatter, O.: Static light scattering of large systems. In: P. Linder, T. Zemb (eds.) Neutrons, X-rays and light: Scattering methods applied to soft condensed matter. North Holland, Elsevier (2002)
18. Gouesbet, G., Gréhan, G.: Optical particle sizing: theory and practice. Academic press, New York (1969)
19. Hanke, M., Hansen, P.C.: Regularization methods for large-scale problems. *Surv. Math. Ind* **3**(4), 253–315 (1993)
20. Hilliard, J.E., Lawson, L.: Stereology and stochastic geometry, vol. 28. Kluwer Academic Publisher, The Netherlands (2003)
21. Hu, B., Shen, J., Duan, T.: Vector similarity measure for particle size analysis based on forward light scattering. *Applied Optics* **54**, 3855–3862 (2015)
22. Hulst, H.C.: Light scattering by small particles. New York: Dover Publications (1981)

23. Kerker, M.: The scattering of light and other electromagnetic radiation (1969)
24. Khan, M., Morigi, S., Reichel, L., Sgallari, F.: Iterative methods of Richardson-Lucy-type for image deblurring. *Numerical Mathematics: Theory, Methods and Applications* **6**(01), 262–275 (2013)
25. Liu, X., Shen, J., Thomas, J.C., Clementi, L.A., Sun, X.: Multiangle dynamic light scattering analysis using a modified chahine method. *Journal of Quantitative Spectroscopy and Radiative Transfer* **113**, 489–497 (2012)
26. Lucy, L.B.: An iterative technique for the rectification of observed distributions. *The astronomical journal* **79**, 745 (1974)
27. Nagy, J.G., Strakoš, Z.: Enforcing nonnegativity in image reconstruction algorithms. In: *International Symposium on Optical Science and Technology*, pp. 182–190. International Society for Optics and Photonics (2000)
28. Naim, M., Boualem, A., Ferre, C., Jabloun, M., Jaloča, A., Ravier, P.: Multiangle multiangle dynamic light scattering for the improvement of multimodal particle size distribution measurements. *Soft Matter* **11**, 28 (2015)
29. Richardson, W.H.: Bayesian-based iterative method of image restoration\*. *JOSA* **62**(1), 55–59 (1972)
30. Roig, A.R., Alessandrini, J.L.: Particle size distributions from static light scattering with regularized non-negative least squares constraints. *Part. Part. Syst. Charact.* **23**, 431–437 (2006)
31. Wang, L., Sun, X.G., Xing, J.: Determination of particle size distribution by light extinction method using improved pattern search algorithm with tikhonov smoothing functional. *Journal of modern Optics* **59**, 1829–1840 (2012)