

An efficient gradient method for maximum entropy regularizing retrieval of atmospheric aerosol particle size distribution function

Yanfei Wang*

Division of Oil and Gas Resource, Institute of Geology and Geophysics, Chinese Academy of Sciences, P.O. Box 9825, Beijing 100029, PR China

Received 27 July 2007; received in revised form 15 November 2007; accepted 16 November 2007

Abstract

The aerosol particle size distribution function can be retrieved by solving the moment problem. This is an ill-posed problem, since it is known that, we are often faced with limited/insufficient observations in remote sensing and the observations are contaminated. To overcome the ill-posed nature, regularization techniques such as Phillips–Twomey's regularization, Tikhonov's smooth regularization as well as some iterative methods were developed. Since the particle size distribution function is always nonnegative, and we are often faced with incomplete data, the concept of maximum entropy from information theory and statistic mechanics can be used for this purpose. Therefore, in this paper, we study the maximum entropy-based regularization method and develop a nonmonotone gradient method for solving the corresponding optimization problem. Our numerical tests for both synthetic problem and practical problem are given to show the efficiency and feasibility of the proposed algorithm.

© 2007 Elsevier Ltd. All rights reserved.

PACS: 02.30.Zz; 42.68.Jg; 42.68.Wt; 92.60.Mt; 84.40.Xb; 02.60.Pn

Keywords: Aerosol particle size distribution; Moment problem; Inversion; Maximum entropy regularization; Gradient method

1. Introduction

It is well-known that the characteristics of the aerosol particle size, which can be represented as a size distribution function in the mathematical formalism, say $n(r)$, plays an important role in climate modeling due to its uncertainty (Houghton et al., 1996). So, the determination of particle size distribution function becomes a basic task in aerosol research (Bohren & Huffman, 1983; Davies, 1974; McCartney, 1976; Twomey, 1977). Since the relationship between the size of atmospheric aerosol particles and the wavelength dependence of the extinction coefficient was first suggested by Ångström (1929), the size distribution began to be retrieved by extinction measurements.

For sun-photometer, the attenuation of the aerosols can be written as the integral equation of the first kind

$$\tau_{\text{aero}}(\lambda) = \int_0^{\infty} \pi r^2 Q_{\text{ext}}(r, \lambda, \eta) n(r) dr + \varrho(\lambda), \quad (1)$$

where r is the particle radius; $n(r)$ is the columnar aerosol size distribution (i.e., the number of particles per unit area per unit radius interval in a vertical column through the atmosphere); η is the complex refractive index of the aerosol

* Tel.: +86 10 8299 8132; fax: +86 10 6201 0846.

E-mail address: yfwang_ucf@yahoo.com

particles; λ is the wavelength; $q(\lambda)$ is the error/noise; $Q_{\text{ext}}(r, \lambda, \eta)$ is the extinction efficiency factor from Mie theory. Since aerosol optical thickness (AOT) can be obtained from the measurements of the solar flux density with sun-photometers, one can retrieve the size distribution by the inversion of AOT measurements through the above equation. This type of method is called extinction spectrometry, which is not only the earliest method applying remote sensing to determine atmospheric aerosol size characteristics, but also the most mature method thus far.

To overcome the oscillations in recovering of the particle size distribution function $n(r)$, various techniques have been developed, e.g., direct regularization methods (Bockmann, 2001; Phillips, 1962; Shaw, 1979; Shifrin & Zolotov, 1996; Twomey, 1963; Wang, Fan, Feng, Yan, & Guan, 2006) and iterative methods (Bockmann & Kirsche, 2006; Chahine, 1970; Ferri, Bassini, & Paganini, 1995; Grassl, 1971; Lumme & Rahola, 1994; Twomey, 1975; Voutilainen & Kaipio, 2000; Wang, Fan, & Feng, 2007; Yamamoto & Tanaka, 1969). We consider an entropy constrained regularization model in this paper, which is usually used for moment problems in different applications (Wang, 2007). Since the commercial device, sun-photometer CE 318, which we used to obtain observation data (Wang et al., 2006, 2007) can be modeled numerically by moment problem, therefore introducing entropy constraint is reasonable. Note that the CE 318 can only supply four aerosol channels, i.e., only four observations are obtained, which are insufficient for the retrieval of the particle size distribution function $n(r)$ by solving Eq. (1). Therefore, numerical difficulty occurs. To overcome the numerical difficulty, we first propose a maximum entropy regularization model and then propose a fast but efficient gradient method to solve the regularization model.

The structure of the paper is organized as follows: Section 2 uses the moment problem to formulate the mathematical model and presents the standard regularization methods. Section 3 provides the proposed methods. In Section 3.1, we formulate our model by introducing entropy and nonnegative constraints. In Section 3.2, we propose an efficient solution method: nonmonotone gradient iterative method of BB type. Section 3.3 presents the numerical implementation details for aerosol retrieval. In Section 4, we first do theoretical simulations; then we use the ground-based remotely sensed measurements to verify the numerical results. In Section 5, some concluding remarks are given. Finally, for clarity, we provide appendices for introducing the Barzilai and Borwein's gradient method (BB type of method) and for explanation of making code by our method.

Throughout the paper, we use the following notations: “:=” denotes “defined as”, “ \vec{n} ” denotes the discretization of a function n , “argmax” and “argmin” denote “maximization for an argument” and “minimization for an argument”, respectively, “max” and “min” denote “maximizing” and “minimizing” some functional, respectively, $\text{diag}(\cdot)$ denotes a diagonal matrix, “ A^T ” denotes the transpose of matrix A and “s.t.” denotes “subject to”.

2. Forward model and standard regularizing methods

2.1. Mathematical model formulation by moment problem

Using moment methods for retrieval of aerosol particle size distributions has been studied by Wright (2000) and Wright et al. (2002) in their excellent papers. We consider a different way by considering observation by sun-photometer and solution by regularization. Assume that $n(y)$ is a real-valued density function in the interval $[a, b]$ and $n \geq 0$, the moment of a function $g(y)$ about $n(y)$ is defined as

$$\langle g(y) \rangle = \int_a^b g(y)n(y) dy.$$

If we set $g(y) = k_i(y)$, $i = 1, 2, \dots, m$, the m known functionals k_i , the particle size distribution measurement system can be considered as a finite moment problem of the following type (Wang, 2007):

$$\int_a^b k_i(y)n(y) dy = o_i, \quad i = 1, 2, \dots, m. \quad (2)$$

Certainly, we have in mind that $o_i = o(x_i)$ and $k_i(\cdot) = k(x_i, \cdot)$. In (2), $[a, b]$ is the integral interval which characterizes the lower and upper limits of the size range of interests; o_i for each i is an error-free observation; x_i is a parameter related to particle wavelength; k_i is a weighted function (or more generally, the kernel function) that characterizes the classification, losses and detection properties of the measurement system.

Practically, the observations are usually contaminated by noise. Hence, the relation (2) between the observation o_i and the size distribution $n(y)$ is

$$d_i(x) = o_i(x) + \varrho_i(x) = \int_a^b k_i(y)n(y) dy + \varrho_i(x), \quad i = 1, 2, \dots, m, \tag{3}$$

where ϱ_i is the unknown observation noise. Therefore, the inverse problem is to solve a perturbed moment problem to get the aerosol particle size distribution $n(y)$.

For the aerosol attenuation problem (1), let us rewrite (1) in the form of the abstract operator equation

$$K : F \longrightarrow O, \\ (Kn)(\lambda) + \varrho(\lambda) = \int_0^\infty k(r, \lambda, \eta)n(r) dr + \varrho(\lambda) = d(\lambda), \tag{4}$$

where $k(r, \lambda, \eta) = \pi r^2 Q_{\text{ex}}(r, \lambda, \eta)$; F denotes the function space of aerosol size distributions; and O denotes the observation space. Both F and O are considered to be the separable Hilbert spaces. Note that τ_{aero} is the measured term, it inevitably induces noise/errors. Hence, $d(\lambda)$ is actually a perturbed right-hand side. Keep in mind operator symbol, (4) can be written as

$$Kn + \varrho = d. \tag{5}$$

If we rewrite (4) and (5) in each wavelength $\lambda_i, i = 1, \dots, m$, we obtain the moment problem

$$(Kn)_i := \int_0^\infty k(r, \lambda_i, \eta)n(r) dr \tag{6}$$

and

$$\tau_{\text{aero}}^i = (Kn)_i + \varrho_i = d_i, \tag{7}$$

where $\varrho_i = \varrho(\lambda_i), d_i = d(\lambda_i), \tau_{\text{aero}}^i = \tau_{\text{aero}}(\lambda_i)$.

2.2. Standard regularization methods

Standard regularization methods refer to solving the following optimization model (Engl, Hanke, & Neubauer, 2000; Tikhonov & Arsenin, 1977; Xiao, Yu, & Wang, 2003):

$$\min J[n] := \frac{1}{2} \|Kn - d\|^2 + \nu \Omega[n], \tag{8}$$

where $\nu > 0$ is the so-called regularization parameter; $\Omega[n]$ is the Tikhonov stabilizer which assigns the smoothness of the function n . Various methods have been developed both directly and iteratively, we refer to Wang et al. (2006) and the references therein for details. There are a large amount of methods for choosing the Tikhonov stabilizer $\Omega[n]$ (Wang, 2007). In the next section, we consider a particular useful one, the maximum entropy regularization model, i.e., we will choose $\Omega[n]$ as the nonnegative entropy $-H(n)$.

3. Theoretical development

3.1. Regularization by maximum entropy constraint and ellipsoidal constraint

It is clear that the moment problem is a special case of Fredholm integral equations of the first kind, which were known as ill-posed problems. Therefore, the aforementioned methods can be used to tackle the ill-posedness. However, notice that the particle size distribution function $n(r)$ is always nonnegative and the practical incomplete observation data, the use of the concept of maximum entropy from information theory and statistic mechanics is quite natural for this purpose, i.e., the entropy of a probability distribution can be considered as a measure of the uncertainty of the experimental outcomes. The idea is by maximizing the Shannon–Jaynes entropy (Engl et al., 2000)

$$H(n) = - \int n(r) \log n(r) dr \tag{9}$$

to solve some kind of moment problem. In which, $n(r)$ is the function to be determined (in the current paper, it refers to the particle size distribution function). The solution is determined or selected from the many sets of functions that can fit the data. The virtues of using the maximum entropy concept are that it provides an unbiased way of obtaining information from incomplete data and at the same time it implicitly possesses the nonnegativity (positivity) constraint to the probability distribution. Using this concept, we formulate a constrained optimization problem for particle size distribution

$$\min M(n) := \frac{1}{2} \|Kn - d\|^2 \quad \text{s.t.} \quad -H(n) \leq \Delta_1, \quad \|Ln\|^2 \leq \Delta_2, \quad 0 \leq n < \infty, \quad (10)$$

where L is a smooth operator to the function $n(r)$, e.g., L can be chosen as a differential operator. In our paper, it comes from a discretization of the Sobolev norm $W^{1,2}$ of $n(r)$, which yields a tridiagonal matrix (see Wang et al., 2006 for details). The first constraint in Eq. (10) is referred as the maximum entropy constraint; the second constraint in Eq. (10) is referred as the ellipsoidal constraint. Therefore, Eq. (10) is similar to a trust region model and may be solved by the corresponding trust region methods (Wang, 2007; Wang & Yuan, 2005; Yuan, 1993). We do not consider the issue in this paper.

Solving the minimization problem (10) is equivalent to solving the following unconstrained problem:

$$\begin{aligned} \min \Psi(n) &:= \frac{1}{2} \|Kn - d\|^2 + v \|Ln\|^2 + \mu(-H(n)) \\ &= \frac{1}{2} \|Kn - d\|^2 + v \|Ln\|^2 + \mu \int n(r) \log n(r) dr \end{aligned} \quad (11)$$

within the feasible set $S = \{n : n \text{ is square integrable and } 0 \leq n < \infty\}$, where v and μ are two regularization parameters which govern the approximation and smoothness of the solution.

Numerically, (11) needs to be reformulated in finite dimensional space

$$\min \tilde{\Psi}(\vec{n}) := \frac{1}{2} \|\mathcal{K}\vec{n} - \vec{d}\|^2 + v \|\mathcal{L}\vec{n}\|^2 + \mu \sum_i n_i \log(n_i) \quad (12)$$

and (12) is solved within the feasible set $S = \{\vec{n} : \vec{n} \in l^2, 0 < \vec{n} < \infty\}$, where \mathcal{K} is a rectangular matrix; $\vec{n} = [n_1, n_2, \dots]^T$ and $\vec{d} = [d_1, d_2, \dots]^T$ are corresponding compatible vectors; \mathcal{L} is a discretization of L .

From a practical viewpoint, it often involves an *a priori* estimation of the solution \vec{n} . Instead of using Eq. (12) directly, it has been suggested that one uses a modified form of the entropy in the functional to be minimized

$$\min \tilde{\Psi}(\vec{n}) := \frac{1}{2} \|\mathcal{K}\vec{n} - \vec{d}\|^2 + v \|\mathcal{L}\vec{n}\|^2 + \mu \sum_i n_i \log(w_{n_i} \cdot n_i), \quad (13)$$

where $w_{\vec{n}} := \{w_{n_1}, w_{n_2}, \dots\}$ is the *a priori* probability distribution (or the weight function) about \vec{n} . A quick look through the literature conveys the impression that this *a priori* distribution is generally set equal to 1, describing complete ignorance (Jaynes, 1968). However, proper choice of the vector $w_{\vec{n}}$ can suppress the difficulty in the numerical realization of (13) that one encounters when simultaneously minimizing the residual term to zero and the penalty to minus infinity.

For choice of the *a priori* $w_{\vec{n}}$, we adopt the method developed in Wang et al. (2007), i.e., we get it from searching for an interior point solution by solving a constrained linear programming problem

$$\min_{\vec{n}} e^T \vec{n} \quad \text{s.t.} \quad \mathcal{K}\vec{n} = \vec{\tau}_{\text{aero}}, \quad \vec{n} \geq 0 \quad (14)$$

within the solution set $\{\vec{n} : \mathcal{K}\vec{n} = \vec{\tau}_{\text{aero}}, \vec{n} \geq 0\}$, where e is a vector with all components 1. Refer to Wang et al. (2007) for an example of specific calculation.

Note that Eq. (13) is a nonlinear function, therefore, iterative approaches should be used. We develop gradient methods for solving this problem in the next subsection.

3.2. A gradient method for solving a regularized solution

Since the nonlinearity of the function $\tilde{\Psi}$ and the gradient can be easily obtained, so it is convenient to use gradient methods for the minimization problem (13). In fact, the gradient of $\tilde{\Psi}$ can be evaluated as

$$g(\vec{n}) = \mathcal{K}^T(\mathcal{K}\vec{n} - \vec{d}) + \nu \mathcal{L}^T(\mathcal{L}\vec{n}) + \mu \text{vec}(1 + \log(w_{\vec{n}} \cdot \vec{n})),$$

where $\text{vec}(1 + \log(w_{\vec{n}} \cdot \vec{n})) := [1 + \log(w_{n_1} \cdot n_1), 1 + \log(w_{n_2} \cdot n_2), \dots, 1 + \log(w_{n_p} \cdot n_p)]^T$, the subindex p refers to the discrete node number.

Perhaps the simplest and the easiest gradient method to code is the steepest descent method

$$\vec{n}_{k+1} = \vec{n}_k + \omega_k s_k, \tag{15}$$

where $s_k = -g_k$, $g_k = g(\vec{n}_k)$, ω_k is the steplength which can be obtained by line search, i.e., optimal ω_k^* satisfies $\omega_k^* = \text{argmin}_{\omega} \tilde{\Psi}(\vec{n}_k + \omega_k s_k)$. However, the steepest descent method is slow in convergence and zigzagging after several iterations (Yuan, 1993). If we restrict the steplength ω_k as a constant in $(0, \|\mathcal{K}\|^{-2})$ in each iteration, we obtain a special gradient method which is known as the Landweber iteration

$$\vec{n}_{k+1} = \vec{n}_k + \omega s_k. \tag{16}$$

However, as is reported that this method is quite slow in convergence and is seldom used for practical problems (Engl et al., 2000; Wang, 2007).

We consider the nonmonotone gradient method. This method was first proposed by Barzilai and Borwein for solving the unconstrained optimization problem, hence is called BB method (Barzilai & Borwein, 1988). The iteration formula is the same as steepest descent method except that the stepsize is controlled by $\alpha_k^{\text{BB1}} = s_{k-1}^T s_{k-1} / s_{k-1}^T y_{k-1}$ or $\alpha_k^{\text{BB2}} = s_{k-1}^T y_{k-1} / y_{k-1}^T y_{k-1}$, where $y_k = g_{k+1} - g_k$, $s_k = \vec{n}_{k+1} - \vec{n}_k$ (see also Appendix A for details).

The method is initially designed for convex quadratic programming problem. But it can be also used for general nonlinear minimization problem provided that the deviation of the object function from a quadratic functional is small and some nonmonotone line search techniques are given (Fletcher, 2001). Since the regularized functional $\tilde{\Psi}(\vec{n})$ must be deviated very small from the norm of the residual $M(\vec{n})$, so we can solve it via BB method. By entropy, the iteration points \vec{n}_k ($k = 1, 2, \dots$) must be positive. Therefore, to avoid negative values in iterations, some safeguard technique should be provided. It is necessary to first modify the negative and zero values of \vec{n}_k obtained in the former step to be positive and nearly zero in order to insert the iteration point into the argument of the natural logarithm in Eq. (13). Therefore, for negative and zero values of \vec{n}_k , we reset the corresponding indices of the weight function $w_{\vec{n}}$ as 1, i.e., complete ignorance of these values and modify the negative and zero values of \vec{n}_k obtained in the former step to be positive and nearly zero and insert the modified \vec{n}_k to the argument of the natural logarithm.

To safeguard the decrease of the sequence $\tilde{\Psi}(\vec{n}_k)$ in each iteration, it is reasonable to add the line search strategy. The widely used line search method is the Wolfe inexact line search strategy, i.e., an acceptable step $s_k = (\vec{n}_k - \alpha_k g_k) - \vec{n}_k$ should satisfy the following conditions for some argument $\beta > 0$ (Yuan, 1993):

$$\tilde{\Psi}(\vec{n}_k) - \tilde{\Psi}(\vec{n}_k + \beta s_k) \geq -\beta \lambda_1 s_k^T g_k, \tag{17}$$

$$s_k^T g(\vec{n}_k + \beta s_k) \geq \lambda_2 s_k^T g_k, \tag{18}$$

where $\lambda_1 \leq \lambda_2$ are two positive parameters in $(0, 1)$. Details of Wolfe inexact line search are given in Wang et al. (2007). However, as is pointed out in Dai and Fletcher (2003), the nonmonotonic behavior of the BB method is lost. Considering these shortcomings, we adopt the adaptive nonmonotone line search method developed in Dai and Fletcher (2003) (similar technique can be also found in Toint, 1997): let $\tilde{\Psi}_{\text{least}}$ be the current least value of the objective functional over all past iterates, i.e., at the k th iteration,

$$\tilde{\Psi}_{\text{least}} = \min_i \tilde{\Psi}(\vec{n}_i) \quad \text{for all } 1 \leq i \leq k. \tag{19}$$

The number of iterations since the value of $\tilde{\Psi}_{\text{least}}$ was found is denoted by l . Define a candidate function value $\tilde{\Psi}_c$ to be the maximum value of the objective functional since the value of $\tilde{\Psi}_{\text{least}}$ was found. Then find a reference function $\tilde{\Psi}_r$ such that

$$\tilde{\Psi}_r - \tilde{\Psi}(\vec{n}_k + \beta s_k) \geq -\beta \gamma_1 s_k^T g_k. \quad (20)$$

In each iteration, the reference value $\tilde{\Psi}_r$ should be improved. The method involves a preassigned small integer parameter $L_r > 0$, and $\tilde{\Psi}_r$ is reduced if the method fails to improve on the previous least value of $\tilde{\Psi}(\vec{n})$ in at most L_r iterations. Therefore, this method does not require the sufficient reduction in $\tilde{\Psi}$, it is much more adaptive.

In our algorithm, the initial search direction is the negative gradient direction. To be sure that the first step is a decreasing step, we use the Wolfe inexact line search strategy instead of the BB step. And, we use the following stopping condition (Wang & Ma, 2007) for ill-posed problems (also recommended in Dai & Fletcher, 2003; Moré & Toraldo, 1991) in our numerical tests:

$$\|\tilde{g}_k\|_2 \leq \varepsilon \|g_1\|_2, \quad (21)$$

where ε is a preassigned tolerance and \tilde{g}_k is defined as

$$(\tilde{g}_k)_i = \begin{cases} (g_k)_i & \text{if } (\vec{n}_k)_i > 0, \\ \min\{(g_k)_i, 0\} & \text{if } (\vec{n}_k)_i = 0. \end{cases}$$

Now, we give the realistic algorithm as follows:

Algorithm 3.1 (*BB method for maximum entropy regularization*).

Step 1: Initialization: Given initial point \vec{n}_0 , tolerance $\varepsilon > 0$ and set $k := 1$.

Step 2: Check whether the stopping condition holds: If (21) is satisfied, stop; Otherwise, set $s_k = -g_k$.

Step 3: If $k = 1$, compute the step size α_k by Wolfe inexact line search strategy (17)–(18). Otherwise compute α_k by α_k^{BB1} or α_k^{BB2} .

Step 4: Update the current iteration point by setting $\vec{n}_{k+1} = \vec{n}_k + \alpha_k s_k$.

Step 5: Find a reference function $\tilde{\Psi}_r$ by Algorithm 3.2 and check whether Eq. (20) is satisfied. If yes, compute a new search direction $s_{k+1} = -g_{k+1}$ and go to Step 7; Otherwise, go to Step 6.

Step 6: Nonmonotone line search: Implement Algorithm 3.2 and update the iteration point and compute a new search direction:

$$\vec{n}_{k+1} = \vec{n}_k + \beta s_k, \quad s_{k+1} = -g_{k+1}.$$

Step 7: Loop: $k := k + 1$ and go to Step 2.

Step 5 in Algorithm 3.1 is used for accepting the next search direction s_{k+1} or not by checking a loose condition (20). Initially, we set $\tilde{\Psi}_r = \infty$. This choice of $\tilde{\Psi}_r$ allows $\tilde{\Psi}(\vec{n}_k) \geq \tilde{\Psi}(\vec{n}_1)$ on early iterations. If the method can find a better functional value in finite iterations $k_{\text{finite}} \leq L_r$, then the values $\tilde{\Psi}_r$ remain unchanged. Otherwise, if the iterations exceed L_r , the value of $\tilde{\Psi}_r$ and $\tilde{\Psi}_c$ need to be reset, i.e., $\tilde{\Psi}_r := \tilde{\Psi}_c$, $\tilde{\Psi}_c := \tilde{\Psi}(\vec{n}_k)$. The algorithm is outlined in Algorithm 3.2.

Algorithm 3.2 (*Nonmonotone line search*).

Step 1: Given the small integer parameter $L_r > 0$, $\tilde{\Psi}_{\text{least}}$, $\tilde{\Psi}_c$ and set $\tilde{\Psi}_r = \infty$.

Step 2: Adaptive nonmonotone line search:

- If $\tilde{\Psi}(\vec{n}_k) < \tilde{\Psi}_{\text{least}}$, set $\tilde{\Psi}_{\text{least}} := \tilde{\Psi}(\vec{n}_k)$, $\tilde{\Psi}_c := \tilde{\Psi}(\vec{n}_k)$, $l := 0$.
- Otherwise, set $\tilde{\Psi}_c := \max\{\tilde{\Psi}_c, \tilde{\Psi}(\vec{n}_k)\}$, $l := l + 1$. If $l = L_r$, $\tilde{\Psi}_r := \tilde{\Psi}_c$, set $\tilde{\Psi}_c := \tilde{\Psi}(\vec{n}_k)$, set $l := 0$.
- Update $\tilde{\Psi}_r$ until (20) is satisfied.

3.3. Aerosol particle size distribution function retrieval

To retrieve the aerosol particle size distribution function $n(r)$, we need to solve the linear equations (6) for AOT at different wavelengths. We are interested with the particle size in the interval $[0.1, 10] \mu\text{m}$. Note that a coarse difference

gridding ($N \leq 20$) induces large quadrature errors, therefore we choose a relatively large difference gridding size, $N = 200$.

To perform the numerical computations, we apply the technique developed in King, Byrne, Herman, and Reagan (1978), i.e., we assume that the actual aerosol particle size distribution function consists of the multiplication of two functions $h(r)$ and $f(r)$: $n(r) = h(r)f(r)$, where $h(r)$ is a rapidly varying function of r , while $f(r)$ is more slowly varying. The function $h(r)$ takes the form of a Junge size distribution (Junge, 1955), $h(r) = r^{-(\nu^*+1)}$, with ν^* assumed to be in the value of 3.0. In this way we have

$$\tau_{\text{aero}}(\lambda) = \int_a^b [k(r, \lambda, \eta)h(r)]f(r) \, dr, \tag{22}$$

where $k(r, \lambda, \eta) = \pi r^2 Q_{\text{ex}}(r, \lambda, \eta)$ and we denote $k(r, \lambda, \eta)h(r)$ as the new kernel function which corresponding to a new operator \mathcal{E} :

$$(\mathcal{E}f)(\lambda) = \tau_{\text{aero}}(\lambda). \tag{23}$$

For simplicity of notation, the discretization of \mathcal{E} is again denoted by the matrix \mathcal{K} . Now, we can perform our algorithm to the problem

$$\min \frac{1}{2} \|\mathcal{K} \vec{f} - \vec{\tau}_{\text{aero}}\|^2 + \nu \|\mathcal{L} \vec{f}\|^2 + \mu \sum_i f_i \log(w_{f_i} \cdot f_i), \tag{24}$$

where $w_{\vec{f}} := [w_{f_1}, w_{f_2}, \dots]^T$ is the weight function about \vec{f} obtained by searching for an interior point solution by solving a constrained linear programming problem

$$\min_{\vec{f}} e^T \vec{f} \quad \text{s.t. } \mathcal{K} \vec{f} = \vec{\tau}_{\text{aero}}, \quad \vec{f} \geq 0 \tag{25}$$

within the solution set $\{\vec{f} : \mathcal{K} \vec{f} = \vec{\tau}_{\text{aero}}, \vec{f} \geq 0\}$, where e is a vector with components all equalling to 1. To safeguard the nonnegativity of iterative steps, $w_{\vec{f}}$ must be modified according to the descriptions in Section 3.2 when encountering negative and zeros values of \vec{f} .

Numerically, choice of the regularization parameters ν and μ is also a major issue. In theory, both ν and μ should approach zero to get the best approximation. A trade-off must be found to balance the ill-posed nature of the discrete matrix \mathcal{K} . There are two types of parameter selection methods: *a priori* way and *a posteriori* way. Since the error or noise level δ is not always to be estimated, we will not apply the *a posteriori* technique developed in Wang et al. (2006). Practically, for a *a priori* choice of the regularization parameter, both ν and μ should be limited to within (0, 1). For our problem, we restrict ν to be an *a priori* and choose μ in a geometric manner:

$$\mu_k = \mu_0 \cdot \xi^{k-1}, \quad 0 < \xi < 1, \tag{26}$$

where $\mu_0 \in (0, 1)$ which can be provided by users, for example, $\mu_0 = 0.5$, k is the k th iteration. It is quite natural to use this parameter selection rule since in theory, μ_k should approach zero as k approaching infinity.

4. Numerical experiments

4.1. Theoretical simulation

To verify the feasibility of our inversion method, we test it by computer simulations. The simulation consists of two steps. First, a simulated extinction signal (input signal) is generated by computer according to Eq. (3) for a given particle number size distribution $n_{\text{true}}(r)$ (input distribution) and for a given complex refractive index η . Then, the input signal is processed through our algorithm, and the retrieved distribution is compared with input one.

In practice, an exact discretization form of the right-hand side $\vec{\tau}$ of Eq. (2) cannot be gotten accurately, instead a perturbed version \vec{d} is obtained. Numerically, a vector \vec{d} should contain different kind of noises. Here, for simplicity,

we assume that the noise is additive and is mainly the Gaussian random noise, i.e., we have

$$\vec{d} = \vec{\sigma} + \delta \cdot \text{rand}(\text{size}(\vec{\sigma})),$$

where δ is the noise level in $(0, 1)$, $\text{rand}(\text{size}(\vec{\sigma}))$ is the Gaussian random noise with the same size as $\vec{\sigma}$.

The precision of the approximation is characterized by the root mean-square error (rmse)

$$\text{rmse} = \sqrt{\frac{1}{m} \sum_{i=1}^m \frac{(\tau_{\text{comp}}(\lambda_i) - \tau_{\text{meas}}(\lambda_i))^2}{(\tau_{\text{comp}}(\lambda_i))^2}},$$

which describes the average relative deviation of the retrieved signals from the true signals. In which, τ_{comp} refers to the retrieved signals, τ_{meas} refers to the measured signals.

In our example, the size distribution function $n_{\text{true}}(r)$ is given by (Wang et al., 2007)

$$n_{\text{true}}(r) = 10.5r^{-3.5} \exp(-10^{-12}r^{-2}).$$

The particle size radius interval of interest is $[0.1, 2] \mu\text{m}$.

Now we give specifications of the initial input values in Algorithms 3.1 and 3.2 for our theoretical simulation: $\lambda_1 = 1.0 \times 10^{-4}$, $\lambda_2 = 0.9$, $\varepsilon = 1.0 \times 10^{-6}$, regularization parameters $\nu = 0.001$, $\mu_0 = 0.55$, the factor of proportionality $\xi = \frac{1}{10}$, then each μ_k is iteratively calculated by the iteration formula (26), $L_r = 7$, $\Psi_r = 1.0 \times 10^{10}$, weight function $w_{\vec{n}}$ is computed from Eq. (14) using simulated data (see Wang et al., 2007 for detailed algorithm description), $\vec{n}_0 = w_{\vec{n}}$, the number of discretization nodes is $N = 200$.

For the unconstrained minimization problem (24), the smooth operator \mathcal{L} is a triangular matrix in the form

$$\mathcal{L} = \begin{bmatrix} 1 + \frac{1}{s_r^2} & -\frac{1}{s_r^2} & 0 & \cdots & 0 \\ -\frac{1}{s_r^2} & 1 + \frac{2}{s_r^2} & -\frac{1}{s_r^2} & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & -\frac{1}{s_r^2} & 1 + \frac{2}{s_r^2} & -\frac{1}{s_r^2} \\ 0 & \cdots & 0 & -\frac{1}{s_r^2} & 1 + \frac{1}{s_r^2} \end{bmatrix}.$$

In which, $s_r = (2 - 0.1)/(N - 1)$ is the step size of the grids in $[a, b]$. The matrix \mathcal{L} imposes some weights on the vector \vec{f} , which performs better than other scale operator in regularizing ill-posedness (see Wang et al., 2006).

The experiments were performed on a personal Pentium (R) computer with CPU 1.0GHz. In the first place, the complex refractive index η is assumed to be $1.45 - 0.00i$ and $1.50 - 0.00i$. Then we invert the same data, supposing η has an imaginary part. The complex refractive index η is assumed to be $1.45 - 0.03i$ and $1.50 - 0.02i$, respectively. Numerical illustrations are plotted in Figs. 1, 3 and 5 using log–log scale with noise levels $\delta = 0.005, 0.01$ and 0.05 for different refractive indices, respectively. The linear plots of these figures in above cases are illustrated in Figs. 2, 4 and 6, respectively. The rmses for each case are shown in Table 1. In our simulations, it is clear that μ_k approaches zero after successful iterations in every measurement, which fulfills the regularization theory (Tikhonov & Arsenin, 1977; Xiao et al., 2003). To shown the computational efficiency, we also record the iterative steps and CPU time, which is given in Table 2.

To show the nonmonotonicity of the BB iteration step, we record the values of the relative error Err_{rel} in each step:

$$\text{Err}_{\text{rel}}^k = \frac{\|\mathcal{K}\vec{n}_k - \vec{\sigma}\|}{\|\vec{\sigma}\|}.$$

The behavior of the relative error of the BB method for the case $\eta = 1.50 - 0.02i$ and $\delta = 0.05$ in every iteration is shown in Fig. 7. It can be clearly seen from Fig. 7 that the relative errors do not decrease monotonically due to the nonmonotonicity of the BB method.

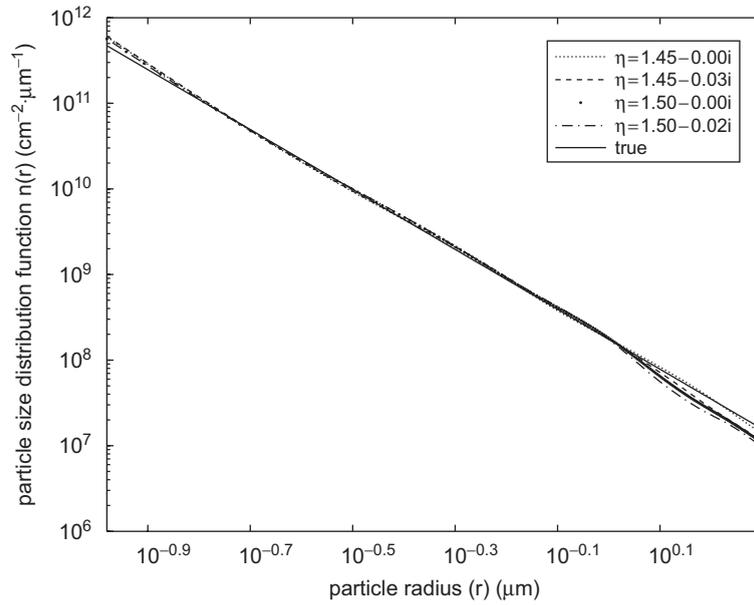


Fig. 1. Log–log plots of input and retrieved results with our inversion method in the case of error level $\delta = 0.005$ and different complex refractive indices.

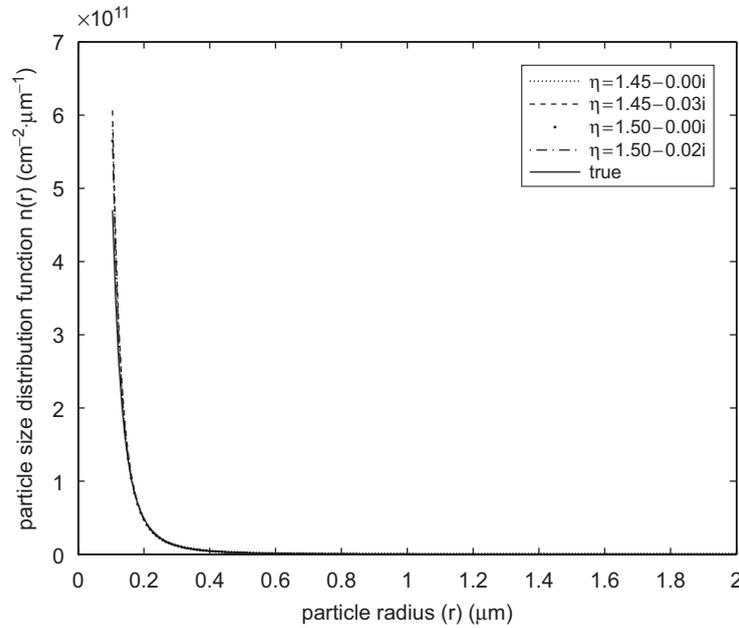


Fig. 2. Linear plots of input and retrieved results with our inversion method in the case of error level $\delta = 0.005$ and different complex refractive indices.

Next, to show the well fitness of our algorithm, we set the slowly varying function $f(r)$ as a Gaussian function with two peaks

$$f(r) = c_1 \exp\left(-\frac{1}{2}\left(\frac{r - p_1}{\sigma_1}\right)^2\right) + c_2 \exp\left(-\frac{1}{2}\left(\frac{r - p_2}{\sigma_2}\right)^2\right),$$

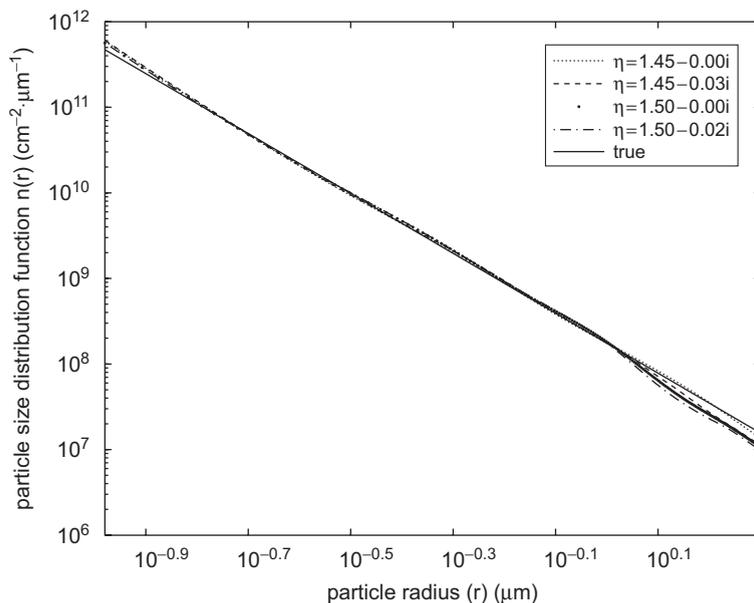


Fig. 3. Log–log plots of input and retrieved results with our inversion method in the case of error level $\delta = 0.01$ and different complex refractive indices.

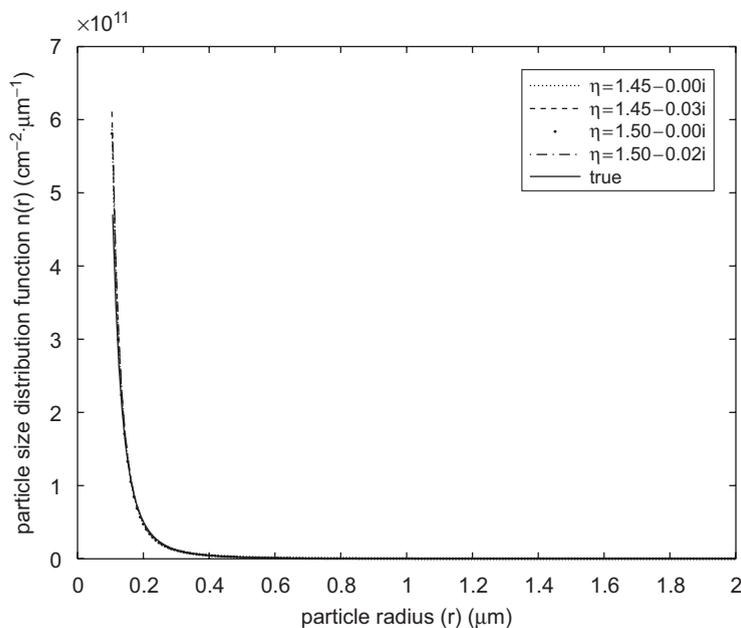


Fig. 4. Linear plots of input and retrieved results with our inversion method in the case of error level $\delta = 0.01$ and different complex refractive indices.

where $c_1 = 10$, $c_2 = 2$, $p_1 = 0.6$, $p_2 = 1.5$, $\sigma_1^2 = 0.025$, $\sigma_2^2 = 0.036$ and to compare the reconstructed signal $f_{\text{comp}}(r)$ with the input signal $f(r)$. The complex refractive index η is again assumed to be $1.45 - 0.00i$, $1.45 - 0.03i$, $1.50 - 0.00i$ and $1.50 - 0.02i$, respectively. Both $f_{\text{comp}}(r)$ and $f(r)$ are normalized in comparison. Numerical illustrations are plotted

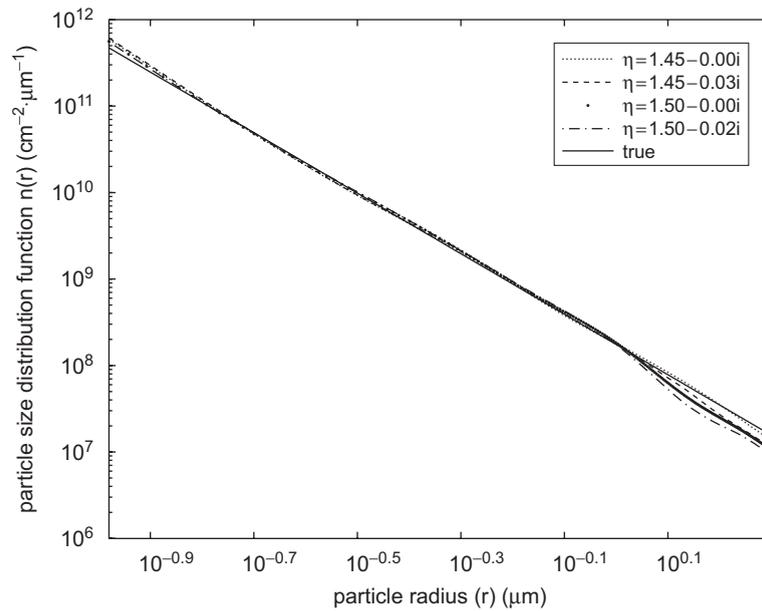


Fig. 5. Log–log plots of input and retrieved results with our inversion method in the case of error level $\delta = 0.05$ and different complex refractive indices.

Table 1
The root mean-square errors for different noise levels

Noise levels	$\eta = 1.45 - 0.00i$	$\eta = 1.45 - 0.03i$	$\eta = 1.50 - 0.00i$	$\eta = 1.50 - 0.02i$
$\delta = 0.005$	1.4501×10^{-4}	8.7595×10^{-5}	9.8996×10^{-5}	1.0632×10^{-4}
$\delta = 0.01$	1.5067×10^{-4}	9.2079×10^{-5}	6.8340×10^{-5}	1.0414×10^{-4}
$\delta = 0.05$	3.1027×10^{-4}	2.5333×10^{-4}	1.8165×10^{-4}	2.1722×10^{-4}

Table 2
The iterative steps (CPU time (seconds)) for different noise levels

Noise levels	$\eta = 1.45 - 0.00i$	$\eta = 1.45 - 0.03i$	$\eta = 1.50 - 0.00i$	$\eta = 1.50 - 0.02i$
$\delta = 0.005$	1995 (2.8130 s)	1492 (2.1400 s)	1130 (1.6090 s)	1133 (1.4060 s)
$\delta = 0.01$	1311 (1.8900 s)	1575 (2.5940 s)	781 (1.0630 s)	1102 (1.6880 s)
$\delta = 0.05$	1348 (1.9210 s)	2000 (2.8590 s)	1329 (1.5470 s)	894 (1.2810 s)

in Figs. 8, 9 and 10 with noise levels $\delta = 0.005, 0.01$ and 0.05 for different refractive indices, respectively. The results show that the slowly varying functions of Gaussian type are well reconstructed.

It indicates from our computer simulation that our method does not affected too much by variation of the complex refractive index and noise. Therefore, our method is stable for retrieving aerosol particle size distribution functions.

4.2. Discussions on numerical results

In this subsection, we choose the ground measured data by the sun-photometer CE 318 (for illustration of the device, experimental site and instrument specifications, refer to Wang et al., 2006) to test the feasibility of the proposed

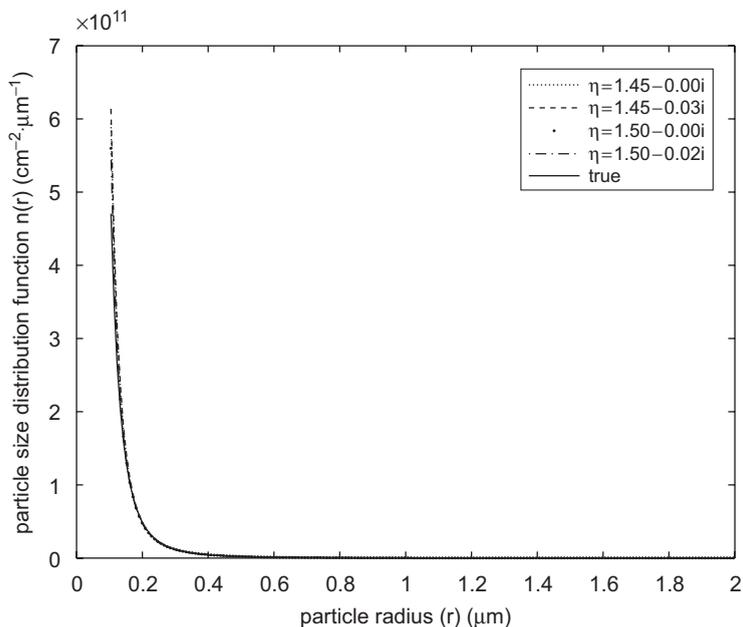


Fig. 6. Linear plots of input and retrieved results with our inversion method in the case of error level $\delta = 0.05$ and different complex refractive indices.

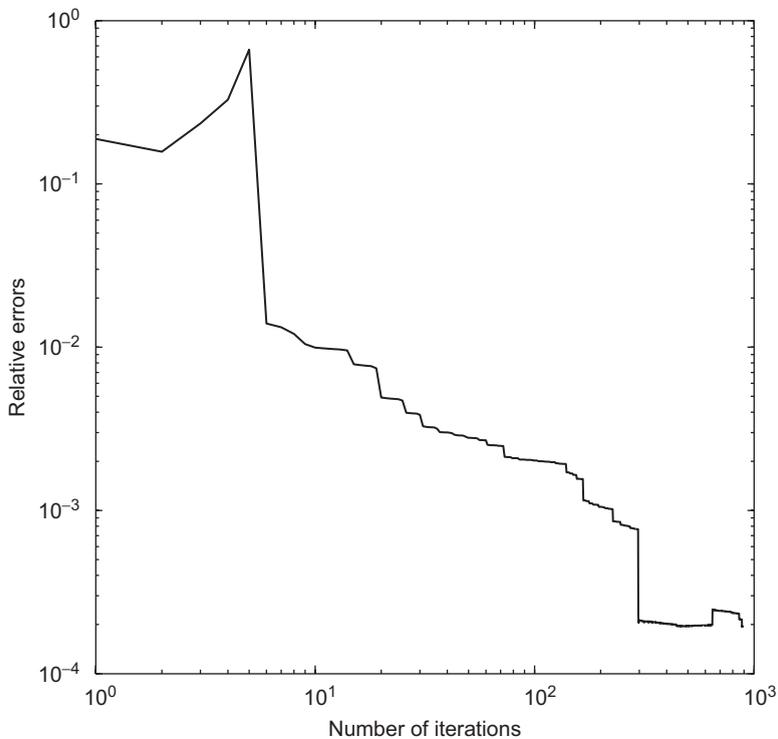


Fig. 7. Nonmonotonicity of the relative errors.

algorithm. We have performed successive *in situ* experiments using CE 318 from October 17 to October 31 in the year 2005. The meteorological data are provided by the Ying Tan Agricultural Ecological Station of CERN (Chinese Ecosystem Research Network).

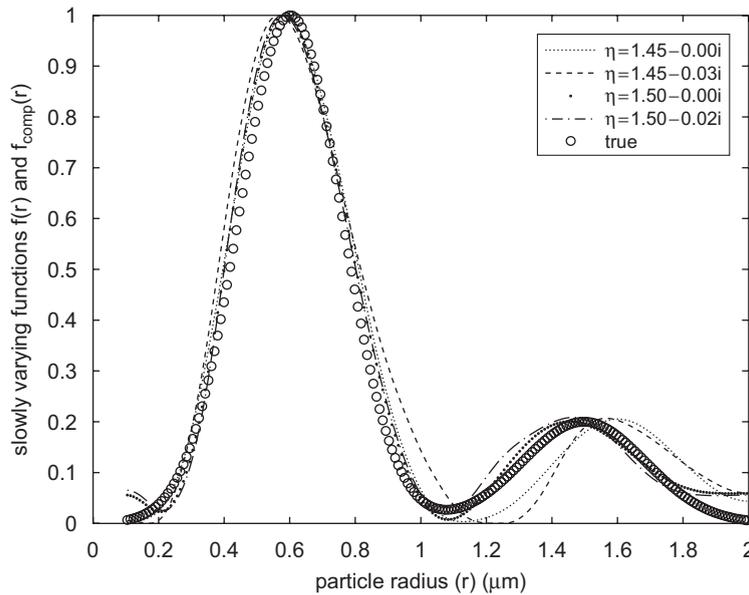


Fig. 8. Input and retrieved slowly varying functions with our inversion method in the case of error level $\delta = 0.005$ and different complex refractive indices.

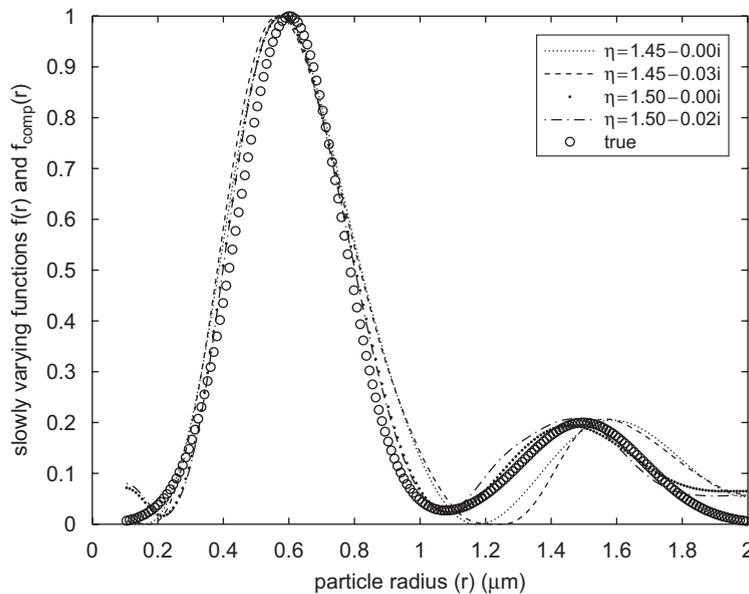


Fig. 9. Input and retrieved slowly varying functions with our inversion method in the case of error level $\delta = 0.01$ and different complex refractive indices.

In this numerical experiment, several chosen days on October 17–31 are used for aerosol inversion. The particle size in the range $[0.1, 10] \mu\text{m}$ is examined. For an illustration of the air mass history, AOT and meteorological description, we refer to Wang et al. (2007). By examination of the AOT values in the morning and afternoon on October 17–31, we find that the magnitude of AOT values on October 17 and 18 are abnormally high. Hence, it may not reflect truly about the aerosol distribution. Therefore, we choose the data at other days in this study.

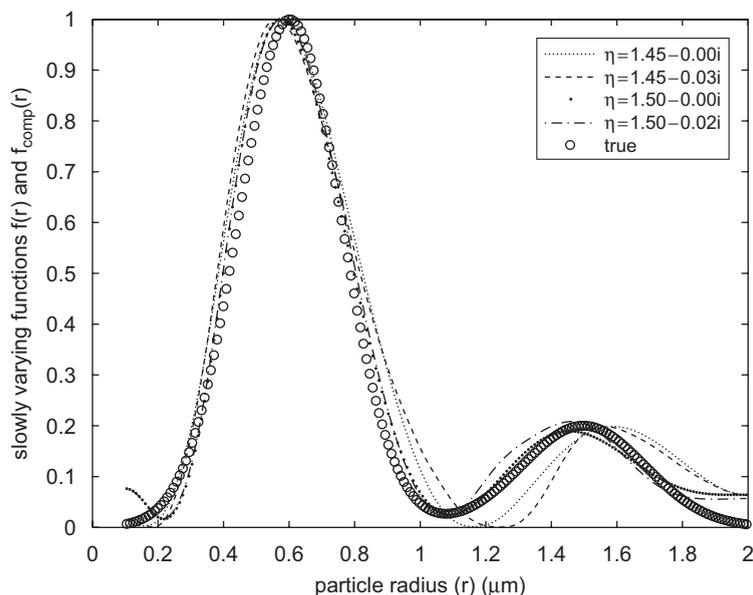


Fig. 10. Input and retrieved slowly varying functions with our inversion method in the case of error level $\delta = 0.05$ and different complex refractive indices.

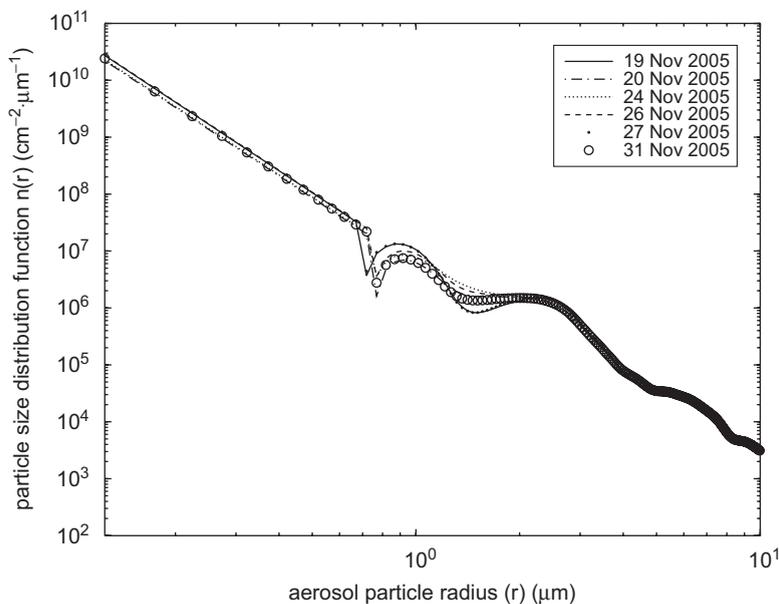


Fig. 11. Particle size distribution in October (PM), 2005.

The composition of the atmospheric aerosols consists of both small and large particles in Yin Tan. Both the scattering and absorption of the particles play a major part. Therefore, a complex refractive index value of $\eta = 1.50 - 0.095i$ was used to perform the inversion (Wang et al., 2007). Now we give specifications of the initial input values in Algorithms 3.1 and 3.2 for our numerical simulations: $\lambda_1, \lambda_2, \varepsilon, \mu_0, \xi, L_r, \tilde{\Psi}_r$ and the discretization nodes number N are the same as in theoretical simulations; $\nu = 0.005, \vec{n}_0 = w_{\vec{H}}, w_{\vec{H}}$ is computed from Eq. (14) using measured data. The smooth operator \mathcal{L} is the same as in theoretical simulation except that $s_r = (10 - 0.1)/(N - 1)$.

By our algorithms, the retrieval results of the number of size distribution function $n(r)$ are plotted in Fig. 11 for the chosen data in the afternoon. It indicates from the figure that the aerosol particles do not decrease rapidly. In all of the chosen days, there are several oscillations of the particle size distribution function near the radius $1.0 \mu\text{m}$. Outside the region containing $1.0 \mu\text{m}$, they changes smoothly. The retrieval results are consistent with the results in Wang et al. (2007). Therefore, from our experiments and the local environment observation, we conclude that the aerosol particle size distribution of Ying Tan city is mainly small air particles. The large particles of size $[5.0, 10.0] \mu\text{m}$, we consider they are mainly composed of cloud and ferric oxide, iron hydroxide and sulfur depositions. The results are consistent with the local air conditions and our observations.

5. Concluding remarks

In this paper, we investigate regularization and optimization methods for retrieval of the atmospheric aerosol particle size distribution function. We first formulate the model by moment problem, and then develop a constrained maximum entropy regularization model. For solution methods, we propose a gradient method of BB type.

We first did theoretical simulations to verify the feasibility of our inversion method. Our results show that the proposed method is stable and insensitive to complex refractive index η and noise levels. Then, we applied our proposed method to the inversion of real extinction data obtained by adapting a commercial sun-photometer CE 318. The numerical experiments illustrate that our new algorithm works well for the retrieval of aerosol particle size distribution functions.

There are still some problems related to the computational model developed in this paper: we use the gradient type of method in this paper instead of Newton type of method. Actually, the Hessian of the model (13) can be easily obtained as

$$\text{Hess}(\vec{n}) = \mathcal{K}^T \mathcal{K} + \nu \mathcal{L}^T \mathcal{L} + \mu \text{diag} \left(\frac{1}{n_1}, \frac{1}{n_2}, \dots, \frac{1}{n_p} \right).$$

Therefore, it is not difficult to establish some Newton and quasi-Newton methods. However, the Hessian is dependent on the value of \vec{n} . Instability will occur when components of \vec{n} approach zeros. These problems deserve further studying.

Finally, we want to point out that our proposed method aims at ill-posed inversion, therefore it is also applicable to geophysical inverse problems and other research area in quantitative remote sensing (Wang, 2007).

Acknowledgments

I am very grateful to reviewers' very helpful comments and suggestions on this paper. The research was supported by National Science Foundation under Grant No. 10501051 and the Scientific Research Foundation of the Human Resource Ministry. It is also partly supported by China National 973 project under Grant No. 2007CB714400.

Appendix A. BB method

The Barzilai–Borwein (BB) method belongs to the group of nonmonotone gradient methods, which was first proposed by Barzilai and Borwein (1988) for solving the unconstrained quadratic programming problem

$$\min q(x) := \frac{1}{2}(Ax, x) - (b, x), \quad (\text{A.1})$$

where $q : \mathbb{R}^m \rightarrow \mathbb{R}$, A is a symmetric positive definite matrix. Let x_k be the k th iterate and g_k the gradient of q at x_k . A gradient method for solving (A.1) calculates the next point from

$$x_{k+1} = x_k - \alpha_k g_k, \quad (\text{A.2})$$

where α_k is the steplength that depends on the method being used. The key point of Barzilai and Borwein et al.'s method is the computation of the steplength using the former gradient step instead of the current gradient step.

Let us investigate the quasi-Newton equation of Eq. (A.1)

$$y_k = As_k, \quad (\text{A.3})$$

where $y_k = g_{k+1} - g_k$, $s_k = x_{k+1} - x_k$. Replacing A by a diagonal matrix $\alpha^{-1}I$ ($\alpha > 0$) and solving

$$\min \|y_{k-1} - \alpha^{-1}I s_{k-1}\|^2$$

yields

$$\alpha_k^{\text{BB1}} = \frac{s_{k-1}^T s_{k-1}}{s_{k-1}^T y_{k-1}}. \quad (\text{A.4})$$

Similarly, if we use αI to approximate the inverse of A , we can choose α_k as the solution of problem

$$\min \|\alpha I y_{k-1} - s_{k-1}\|^2$$

and obtain

$$\alpha_k^{\text{BB2}} = \frac{s_{k-1}^T y_{k-1}}{y_{k-1}^T y_{k-1}}. \quad (\text{A.5})$$

Note that $s_{k-1} = -\alpha_{k-1} g_{k-1}$ and $y_{k-1} = A s_{k-1}$, we can also obtain the equivalent form of Eqs. (A.4) and (A.5):

$$\alpha_k^{\text{BB1}'} = \frac{g_{k-1}^T g_{k-1}}{g_{k-1}^T A g_{k-1}} \quad (\text{A.6})$$

and

$$\alpha_k^{\text{BB2}'} = \frac{g_{k-1}^T g_{k-1}}{g_{k-1}^T A^T A g_{k-1}}. \quad (\text{A.7})$$

Now it is clear from Eqs. (A.6) and (A.7) that the BB method only employ the former gradient step instead of the current gradient step.

The key point of Barzilai and Borwein et al.'s method is the two choices of the stepsize α_k . Investigation on Eqs. (A.6)–(A.7) reveals that Eq. (A.6) is better than Eq. (A.7), since $\alpha_k^{\text{BB1}'}$ has small jumps than $\alpha_k^{\text{BB2}'}$ when A is ill-conditioning. Therefore, we can alternately use the formulas (A.4)–(A.5) in iterations.

Appendix B. Instructions on implementing the algorithms

To use the algorithms described in this paper, we want to mention some procedures and subroutines.

Wolfe line search subroutine: Eqs. (17)–(18) is equivalent to

$$f(0) - f(\beta_k) \geq -\beta_k \lambda_1 f'(0), \quad (\text{B.1})$$

$$f'(\beta_k) \geq \lambda_2 f'(0), \quad (\text{B.2})$$

where $f(\beta) := \tilde{\Psi}(\vec{n}_k + \beta s_k)$. This subroutine finds β_k by minimizing $f(\beta)$ and β_k satisfies Eqs. (B.1)–(B.2). Now I describe a simple feasible algorithm given by Yuan and Sun (1997).

Algorithm B.1 (*Line search algorithm*).

Step 1: Initialization: Given initial points $a_1 = 0$, $a_2 = +\infty$, $\beta_0 > 0$, $\lambda_1 \in (0, \frac{1}{2})$, $\lambda_2 \in (\lambda_1, 1)$, $t > 1$ and compute $f(0)$, $f'(0)$. Set $k := k + 1$.

Step 2: Compute $f(\beta_k)$ and check condition (B.1) is satisfied or not: If (B.1) is satisfied, goto Step 3; Otherwise, set $a_{k+1} := a_k$, $b_{k+1} := \beta_k$, goto Step 4.

- Step 3:* Check condition (B.2) is satisfied or not: If (B.2) is satisfied, stop, output β_k ; Otherwise, set $a_{k+1} := \alpha_k$, $b_{k+1} := b_k$. If $b_{k+1} < +\infty$, goto Step 4, Otherwise, set $\alpha_{k+1} := t\alpha_k$, $k := k + 1$, goto Step 2.
- Step 4:* Compute a new iteration point by bisection approach, i.e., $\alpha_{k+1} := (a_{k+1} + b_{k+1})/2$ and set $k := k + 1$, goto Step 2.

In our numerical tests, in Step 1, we choose $\lambda_1 = 1.0 \times 10^{-4}$ and $\lambda_2 = 0.9$.

BB subroutine: This subroutine finds iteration point \vec{n}_{k+1} by setting $\vec{n}_{k+1} = \vec{n}_k + s_k$. Note that there are several ways to compute α_k^{BB} , one way is using only α_k^{BB1} by Eq. (A.4), another way is using only α_k^{BB2} by Eq. (A.5). One can also use combination of α_k^{BB1} and α_k^{BB2} , e.g., setting $\alpha_k^{\text{BB}} = \alpha_k^{\text{BB1}}$ in odd steps and $\alpha_k^{\text{BB}} = \alpha_k^{\text{BB2}}$ in even steps. In my opinion, α_k^{BB1} should be mostly used instead of α_k^{BB2} , since by Eq. (A.7), α_k^{BB2} has large jumps for ill-posed problems.

For the *nonmonotone subroutine*, it can be followed directly by Algorithm 3.2.

References

- Ångström, A. (1929). On the atmospheric transmission of sun radiation and on dust in the air. *Geografiska Annaler*, 11, 156–166.
- Barzilai, J., & Borwein, J. (1988). Two-point step size gradient methods. *IMA Journal of Numerical Analysis*, 8, 141–148.
- Bockmann, C. (2001). Hybrid regularization method for the ill-posed inversion of multiwavelength lidar data in the retrieval of aerosol size distributions. *Applied Optics*, 40, 1329–1342.
- Bockmann, C., & Kirsche, A. (2006). Iterative regularization method for lidar remote sensing. *Computer Physics Communications*, 174, 607–615.
- Bohren, G. F., & Huffman, D. R. (1983). *Absorption and scattering of light by small particles*. New York: Wiley.
- Chahine, M. T. (1970). Inverse problems in radiation transfer: Determination of atmospheric parameters. *Journal of Aerosol Science*, 27, 960–967.
- Dai, Y. H., & Fletcher, R. (2003). *Projected Barzilai–Borwein methods for large-scale box-constrained quadratic programming*. University of Dundee Report NA/215.
- Davies, C. N. (1974). Size distribution of atmospheric aerosol. *Journal of Aerosol Science*, 5, 293–300.
- Engl, H. W., Hanke, M., & Neubauer, A. (2000). *Regularization of inverse problems*. Dordrecht: Kluwer Academic Publishers.
- Ferri, F., Bassini, A., & Paganini, E. (1995). Modified version of the Chahine algorithm to invert spectral extinction data for particle sizing. *Applied Optics*, 34, 5829–5839.
- Fletcher, R. (2001). *On the Barzilai–Borwein method*. University of Dundee Report NA/207.
- Grassl, H. (1971). Determination of aerosol size distributions from spectral attenuation measurements. *Applied Optics*, 10, 2534–2538.
- Houghton, J. T., Meira Filho, L. G., Callander, B. A., Harris, N., Kattenberg, A., & Maskell, K. (1996). *Climate change 1995*. Cambridge: Cambridge University Press, (Published for the Intergovernmental Panel on Climate Change).
- Jaynes, E. T. (1968). Prior probabilities. *IEEE Transactions on Systems Science and Cybernetics*, SSC-4, 227–241.
- Junge, C. E. (1955). The size distribution and aging of natural aerosols as determined from electrical and optical data on the atmosphere. *Journal of Meteorology*, 12, 13–25.
- King, M. D., Byrne, D. M., Herman, B. M., & Reagan, J. A. (1978). Aerosol size distributions obtained by inversion of spectral optical depth measurements. *Journal of Aerosol Science*, 35, 2153–2167.
- Lumme, K., & Rahola, J. (1994). Light scattering by porous dust particles in the discrete-dipole approximation. *The Astrophysical Journal*, 425, 653–667.
- Mccartney, G. J. (1976). *Optics of atmosphere*. New York: Wiley.
- Moré, J., & Toraldo, G. (1991). On the solution of large quadratic programming problems with bound constraints. *SIAM Journal on Optimization*, 1, 93–113.
- Phillips, D. L. (1962). A technique for the numerical solution of certain integral equations of the first kind. *Journal of the Association for Computing Machinery*, 9, 84–97.
- Shaw, G. E. (1979). Inversion of optical scattering and spectral extinction measurements to recover aerosol size spectra. *Applied Optics*, 18, 988–993.
- Shifrin, K. S., & Zolotov, L. G. (1996). Spectral attenuation and aerosol particle size distribution. *Applied Optics*, 35, 2114–2124.
- Tikhonov, A. N., & Arsenin, V. Y. (1977). *Solutions of ill-posed problems*. New York: Wiley.
- Toint, Ph. L. (1997). A non-monotone trust region algorithm for nonlinear optimization subject to convex constraints. *Mathematical Programming*, 77, 69–94.
- Twomey, S. (1963). On the numerical solution of Fredholm integral equations of the first kind by the inversion of the linear system produced by quadrature. *Journal of the Association for Computing Machinery*, 10, 97–101.
- Twomey, S. (1975). Comparison of constrained linear inversion and an iterative nonlinear algorithm applied to the indirect estimation of particle size distributions. *Journal of Computational Physics*, 18, 188–200.
- Twomey, S. (1977). *Atmospheric aerosols*. Amsterdam: Elsevier.
- Voutilainenand, A., & Kaipio, J. P. (2000). Statistical inversion of aerosol size distribution data. *Journal of Aerosol Science*, 31(Suppl. 1), 767–768.
- Wang, Y. F. (2007). *Computational methods for inverse problems and their applications*. Beijing: Higher Education Press.
- Wang, Y. F., Fan, S. F., & Feng, X. (2007). Retrieval of the aerosol particle size distribution function by incorporating *a priori* information. *Journal of Aerosol Science*, 38, 885–901.
- Wang, Y. F., Fan, S. F., Feng, X., Yan, G. J., & Guan, Y. N. (2006). Regularized inversion method for retrieval of aerosol particle size distribution function in $W^{1,2}$ space. *Applied Optics*, 45, 7456–7467.

- Wang, Y. F., & Ma, S. Q. (2007). Projected Barzilai–Borwein methods for large scale nonnegative image restorations. *Inverse Problems in Science and Engineering*, 15(6), 559–583.
- Wang, Y. F., & Yuan, Y. X. (2005). Convergence and regularity of trust region methods for nonlinear ill-posed inverse problems. *Inverse Problems*, 21, 821–838.
- Wright, D. L. (2000). Retrieval of optical properties of atmospheric aerosols from moments of the particle size distribution. *Journal of Aerosol Science*, 31, 1–18.
- Wright, D. L., Yu, S. C., Kasibhatla, P. S., McGraw, R., Schwartz, S. E., Saxena, V. K. et al. (2002). Retrieval of aerosol properties from moments of the particle size distribution for kernels involving the step function: Cloud droplet activation. *Journal of Aerosol Science*, 33, 319–337.
- Xiao, T. Y., Yu, S. G., & Wang, Y. F. (2003). *Numerical methods for the solution of inverse problems*. Beijing: Science Press.
- Yamamoto, G., & Tanaka, M. (1969). Determination of aerosol size distribution function from spectral attenuation measurements. *Applied Optics*, 8, 447–453.
- Yuan, Y. X. (1993). *Numerical methods for nonlinear programming*. Shanghai: Shanghai Science and Technology Publication.
- Yuan, Y. X., & Sun, W. Y. (1997). *Theory and methods of optimization*. Beijing: Science Press.