Regularizing active set method for retrieval of the atmospheric aerosol particle size distribution function

Yanfei Wang* and Changchun Yang

Institute of Geology and Geophysics, Chinese Academy of Sciences, P.O. Box 9825, Beijing, 100029, China
**Corresponding author: yfwang_ucf@yahoo.com*

Received September 18, 2007; revised November 19, 2007; accepted November 26, 2007;
posted December 4, 2007 (Doc. ID 87213); published January 14, 2008

The determination of the aerosol particle size distribution function by using the particle spectrum extinction equation is an ill-posed integral equation of the first kind [S. Twomey, J. Comput. Phys. **18**, 188 (1975); Y. F. Wang, *Computational Methods for Inverse Problems and Their Applications* (Higher Education Press, 2007)], since we are often faced with limited or insufficient observations in remote sensing and the observations are contaminated. To overcome the ill-posed nature of the problem, regularization techniques were developed. However, most of the literature focuses on the application of Phillips–Twomey regularization and its variants, which are unstable in several cases. As is known, the particle size distribution is always nonnegative, and we are often faced with incomplete data. Therefore, we study the active set method and propose a regularizing active set algorithm for ill-posed particle size distribution function retrieval and for enforcing nonnegativity in computation. Our numerical tests are based on synthetic data for theoretical simulations and the field data obtained with a CE 318 Sun photometer for the Po Yang lake region of Jiang Xi Province, China, and are performed to show the efficiency and feasibility of the proposed algorithms. © 2008 Optical Society of America

OCIS codes: 010.1100, 010.1110, 280.1100, 100.0100, 100.3190, 000.4430.

1. INTRODUCTION

Atmospheric aerosols are suspensions of small solid or liquid particles in the atmosphere, which play an important role in atmospheric and environmental research since they take part in many physical and chemical processes in the atmosphere (see [1–5]). It is well known that characteristics of the aerosol particle size, which can be represented as a size distribution function, say $n(r)$, in the mathematical formalism, play an important role in affecting climate. Thus it is necessary to determine the size distribution function of the aerosol particles. Since the relationship between the size of atmospheric aerosol particles and the wavelength dependence of the extinction coefficient was first suggested by Ångström in 1929, the size distribution began to be retrieved by extinction measurements. First, Ångström inferred that the parameters of a Junge size distribution could be obtained from the aerosol optical thickness (AOT) at multiple wavelengths, and he obtained the useful Ångström empirical formula of Junge size distribution, $\tau_{aero}(\lambda)=\beta\lambda^{-\alpha}$, where τ_{aero} is the measured AOT, β is the turbidity coefficient, and α is the Ångström exponent reflecting the aerosol size distribution (see [6]).

The attenuation of the aerosols can be written as integral equations of the first kind,

$$\tau_{aero}(\lambda)=\int_0^\infty \pi r^2 Q_{ex}(r,\lambda,\eta)n(r)\mathrm{d}r+\varrho(\lambda), \qquad (1)$$

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 particles, λ is the wavelength, $\varrho(\lambda)$ is the error or noise, and $Q_{ex}(r,\lambda,\eta)$ is the extinction efficiency factor from Mie theory. Since the AOT can be obtained from measurements of the solar flux density with Sun photometers, one can retrieve the size distribution by the inversion of AOT measurements through the above equations. 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 (see [7,8]).

To overcome oscillations in recovering the particle size distribution function $n(r)$, various techniques have been developed, such as direct regularization methods (e.g., [9–14]), various iterative methods (e.g., [15–24]), moment methods (e.g., [25,26]), statistical methods (e.g., [27,28]) and computed tomography (e.g., [29]). However, these methods do not consider the constraint of nonnegativity of aerosol particle size distributions and may lead to physically meaningless zero and negative solutions. This paper will address this problem. In addition, the commercial Sun photometer CE 318 can only supply four aerosol channels; i.e., only four observations are obtained, a number insufficient for the retrieval of the particle size distribution function $n(r)$ by solving Eq. (1). Therefore, a numerical difficulty occurs. To overcome the numerical difficulty while keeping the solution nonnegative, we first develop a constrained regularizing quadratic model and then propose an active set method for solving the aerosol particle size distribution retrieval problem.

The paper is organized as follows: Section 2 provides

the background for the problem's formulation in infinite space and solution methods. Subsection 2.A formulates the mathematical model as an operator equation of the first kind; Subsection 2.B briefly reviews Phillips–Twomey constrained optimization method and Tikhonov's regularization method. In Section 3, we formulate our model by introducing positive constraints. In Subsection 3.A, we propose a regularizing active set method. Subsections 3.B and 3.C present the numerical implementation details. In Section 4, we first perform theoretical simulations to demonstrate the validity and feasibility of the proposed method; then we use the ground-based remotely sensed measurements to verify the numerical results with our new methods. In Section 5, some concluding remarks are given. Finally, we provide two appendices to state the fundamentals for the constrained quadratic programming problem and to give a realistic algorithm for solving the problem.

Throughout the paper, we use the following notation: "$:=$" denotes "defined as;" K denotes an operator, and $\mathcal{K}$ denotes a discrete operator (i.e., matrix form) in finite dimensional space; "$\vec{n}$" denotes the discretization of a continuous 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; "A^*" and "A^T" denote the adjoint and the transpose of matrix A, respectively, and "s.t." denotes "subject to."

2. ILL-POSED NATURE OF MODEL INVERSION AND REGULARIZATION

A common feature for all particle size distribution measurement systems is that the relation between noiseless observations and the size distribution function can be expressed as a first-kind Fredholm integral equation (e.g., [16,23,30]). For the aerosol attenuation problem (1), let us rewrite Eq. (1) in the form of the abstract operator equation

$$K:F \to O,$$

$$(Kn)(\lambda) + \varrho(\lambda) = \int_0^\infty k(r,\lambda,\eta)n(r)\mathrm{d}r + \varrho(\lambda) = o(\lambda) + \varrho(\lambda) = d(\lambda), \tag{2}$$

where $k(r,\lambda,\eta) = \pi r^2 Q_{ex}(r,\lambda,\eta)$; F denotes the function space of aerosol size distributions, and O denotes the observation space. Both F and O are considered to be separable Hilbert spaces. Note that τ_{aero} in Eq. (1) is the measured term; it inevitably induces noise or errors. Hence, the right-hand side of Eq. (2), $d(\lambda)$, is actually perturbed. Keep in mind that the operation of Eq. (2) can be written as

$$Kn + \varrho = o + \varrho = d. \tag{3}$$

A. Ill-Posedness

Ill-posedness is a basic problem for inverse problems in various applications (see, e.g., [16,31–34]). For the aerosol particle size distribution function retrieval problem, the ill-posed nature arises because (1) the model operator is compact, and hence the inverse of small singular values of the operator leads to unexpected huge values; (2) the observations contain noise; and (3) the number of observations is insufficient. These ill-posed characteristics produce a kind of jump in the solution space; i.e., instead of being centered around the true solution, the results may spread over the whole parameter space.

B. Regularization

Regularization is a necessary way to tackle the ill-posed nature of the inversion process.

Both Phillips–Twomey regularization (see [9,10]) and Tikhonov regularization (see [31]) belong to standard smooth regularization methods. The general form is given by

$$\min \frac{1}{2}\|Kn - d\|^2 + \nu\Omega[n], \tag{4}$$

where $\Omega[n]$ is the Tikhonov stabilizer that assigns the smoothness of the function n; $\nu > 0$ is the regularization parameter balancing the ill-posedness and smoothness. In Phillips–Twomey regularization, $\Omega[n]$ is chosen as a quadratic form $\Omega[n] = (Dn,n)$, where D is a preassigned scale operator and is usually chosen as the sums of squares of the second differences. In standard Tikhonov regularization, $\Omega[n]$ is chosen as a Sobolev norm function of the form $\Omega[n] = \|n\|^2_{W^{1,2}}$ (see [12] for details).

With the regularization model, developing suitable solution methods is very important.

3. THEORETICAL DEVELOPMENT

A. Regularizing Active Set Method

We consider the regularizing minimization problem

$$\min J_0[n] := \frac{1}{2}\|Kn - d\|^2 + \frac{\nu}{2}(Dn,n) \tag{5}$$

under the constraint $l \leqslant n \leqslant u$, where $\nu > 0$ is to be assigned and D is a positive (semi)definite operator. Recall that a positive definite and (semi)definite operator T refers to $(Tx,x) > 0$ and $(Tx,x) \geqslant 0$, respectively. We solve the problem in the feasible set $S_0 := \{n : l \leqslant n \leqslant u\}$.

Using the inner product, the objective functional can be rewritten as a simple quadratic form

$$J_0[n] := \frac{1}{2}[(K^*K + \nu D)n,n] - (d,Kn) + \frac{1}{2}\|d\|^2 \tag{6}$$

subject to $n \in S_0$, where K^* is the adjoint of K defined by $(x,Ky) = (K^*x,y)$. Note that n is always nonnegative and upper bounded; therefore, problem (5) is equivalent to

$$\min J[n] := \frac{1}{2}[(K^*K + \nu D)n,n] - (d,Kn) \tag{7}$$

subject to $n \in S_1 := \{n : n \geqslant 0\}$.

Thus, the goal of an optimization algorithm is solution of the quadratic problem presented in Eq. (7), that is, the

search for a point, n^*, in $S_1 \subset F$ such that the objective function $J[n]$ is minimized in addition to satisfying the set of constraints.

Before generating the algorithm, we explain a few terms used in solving the quadratic program in Eq. (7).

Feasible point and feasible set: Any point n in F that satisfies all constraints in S_1 is said to be a feasible point. The set of feasible points is referred to as the feasible region. If the constraints are inconsistent, then the problem will be infeasible (e.g., minimize $J[n]$ subject to $n<l$ or $n>u$).

Active constraint: The constraint $n \in S_1$ is said to be active at $\tilde{n}$ if $\tilde{n}$ lies on the boundary of the feasible region and this boundary is formed by the constraints whose indices are members of set S. Set S is referred to as the active set. During the search process, some inequality constraints may become active, and their indices will also be included in S.

Working set: The working set W_k is a prediction of the active constraints at the solution; i.e., a subset of the constraints in S is imposed as a set of equalities.

An active method is an implicit Newton-type method for solving the constrained quadratic programming problem (7), which describes a method for identifying a correct set of active inequality constraints and temporarily giving up the remaining inequality constraints. Originally, the method was designed for a well-posed quadratic programming problem (e.g., [35–37]). We apply it to an ill-posed aerosol particle size distribution function retrieval problem and solve a regularizing problem.

Given an iterate n_k and the working set W_k, we first need to test whether n_k minimizes the quadratic functional $J[n]$ in the subspace defined by the working set. If not, we compute a step s by solving an equality-constrained quadratic programming subproblem in which the constraints corresponding to the working set W_k are treated as equalities and all other constraints are temporarily ignored. So, given the iteration point n_k and the working set $W_k := \{j \in S : n_k^j = 0\}$, the subproblem in terms of the step $s_k = n - n_k$ can be expressed as

$$\min J_{s_k}[s_k + n_k] = \frac{1}{2}(Gs_k, s_k) + (g_k, s_k) + c,$$

$$\text{s.t. } s_k^j = 0, \quad j \in W_k \tag{8}$$

with $G = K^*K + \nu D$, $g_k = Gn_k - K^*d$, and $c = \frac{1}{2}(Gn_k, n_k) - (Kn_k, d)$. Since c is a constant, at the kth iterative step, we actually solve the equation

$$\min Q[s_k] = \frac{1}{2}(Gs_k, s_k) + (g_k, s_k),$$

$$\text{s.t. } s_k^j = 0, \quad j \in W_k. \tag{9}$$

We denote the solution of Eq. (9) by s_k^*. Note that the constraints in W_k were satisfied at n_k; they are also satisfied at $n_k + \alpha s_k^*$ for any value α. It is clear that there is a trivial solution $s_k^* = 0$. Therefore, we suppose for the moment that the optimal s_k^* is nonzero. We need to decide how far to move along the direction s_k^*. The strategy is that if $n_k + s_k^*$ is feasible with regard to all constraints, we set $n_{k+1} = n_k + s_k^*$; otherwise, a line search is made in the direction s_k^* to find the best feasible point; i.e., we set $n_{k+1} = n_k + \alpha_k s_k^*$, where α_k is the step size that satisfies

$$\alpha_k := \min\left\{1, \min_{j \notin W_k, s_k^j < 0} \frac{-n_k^j}{s_k^j}\right\}. \tag{10}$$

If $\alpha_k < 1$ in Eq. (10), then a new working set W_{k+1} is constructed by adding one active constraint. This constraint is defined by the index, say l, that achieves the minimum in Eq. (10), and this index is added to the active set W_k. The procedure for adding constraints to W_k is continued until a point n_k^* is reached that minimizes the quadratic functional over its current working set W_k^*. It is easy to recognize that this point is given at the trivial solution $s_k^j = 0$ and satisfies the optimality conditions for Eq. (9),

$$\sum_{j \in W_k^*} \lambda_j^* = Gn_k^* - K^*d, \tag{11}$$

for some vector Lagrangian multiplier $\lambda_j^*, j \in W_k^*$.

With the above preparation, we focus on the numerical procedure for solving the problem. The first-order necessary condition for Eq. (9) at W_k^* yields

$$Gs_k^* + g_k - \sum_{j \in W_k^*} \lambda_j^* = 0, \tag{12}$$

$$s_k^{*j} = 0, \quad j \in W_k^*, \tag{13}$$

$$\lambda_j^* \geqslant 0, \quad j \in W_k^*. \tag{14}$$

If we define the multipliers corresponding to the inequality constraints that are not in the working set to be zero, then n_k^* and λ_j^* satisfy the Karush–Kuhn–Tucker (KKT) conditions for Eq. (7) with the constraint $n \in S$, i.e.,

$$Gn_k^* - K^*d - \sum_{j \in W_k} \lambda_j^* = 0, \tag{15}$$

$$n_k^{*j} = 0, \quad j \in W_k, \tag{16}$$

$$n_k^{*j} \geqslant 0, \quad j \in S, \quad j \notin W_k, \tag{17}$$

and the multiplier λ_j^* is adjusted to update the model. If $\lambda_j^* \geqslant 0$ for all $j \in W_k$, then n_k^* is a strict minimizer; otherwise, if one of the multipliers $\lambda_j^* < 0$, the objective functional $Q[s_k]$ may be decreased by dropping this constraint; i.e., we remove an index j corresponding to one of the negative multipliers from the working set and solve a new subproblem (9) for the new step.

Based on the above preparation, the regularizing active set algorithm for the aerosol particle size distribution function retrieval problem is given as follows:

Algorithm 3.1 (a regularizing active set algorithm).

Step 1. Compute a feasible starting point n_0; set W_0 to be a subset of the active constraints at n_0; give the initial regularization parameter $\nu_0 > 0$ and the positive (semi)definite matrix D; set $k := 0$ and compute $G = K^*K + \nu D$.

Step 2. Solve Eq. (9) to find s_k; If $s_k \neq 0$, GOTO Step 3; Otherwise, GOTO Step 4.

Step 3. Compute α_k from Eq. (10); Set $n_{k+1}=n_k+\alpha_k s_k$; If $\alpha_k=1$, GOTO Step 5; Otherwise, find $l\notin W_k$ such that $n_k^l+\alpha_k s_k^l=0$ and set $W_k:=W_k\cup\{l\}$;
Step 4. Compute the Lagrangian multipliers λ_k^j that satisfy Eq. (11); set $W_k^*=W_k$; If $\lambda_k^j\geqslant 0$ for all $j\in W_k$, STOP; output the solution $n^*=n_k$; Otherwise, set $j=\text{argmin}_{j\in W_k}\lambda_k^j$; $n_{k+1}=n_k$; set $W_k:=W_k\setminus\{j\}$; GOTO Step 5;
Step 5. Set $W_{k+1}:=W_k$, $k:=k+1$ and update regularization parameter ν_k; GOTO Step 2.

In Step 1, the computation of $G=K^*K+\nu D$ is not necessary if the solution of Eq. (9) is by an iterative method, say, the conjugate gradient method (see Appendix B), since only matrix–vector multiplication is performed. The numerical procedure for solving Eq. (9) is given in Appendix B.

Remark. Note that our model is formulated in a regularizing form and that the object functional $J[n]$ is strictly convex for proper choice of D and ν; therefore, it is indeed a regularizing algorithm with active set solution. Convergence can be shown to follow similarly to the proof of the well-posed case (see [35–37]).

B. Choosing the Scale Matrix D and the Regularization Parameter ν

To ensure the convexity of the quadratic programming problem (7) and (9), it is necessary to choose the appropriate regularization parameter ν and the scale matrix D. There are several ways to choose the matrix D; however, it is pointed out in [12] that the following form of D possesses good numerical stability:

$$D=\begin{bmatrix} 1+\frac{1}{h_r^2} & -\frac{1}{h_r^2} & 0 & \cdots & 0 \\ -\frac{1}{h_r^2} & 1+\frac{2}{h_r^2} & -\frac{1}{h_r^2} & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & -\frac{1}{h_r^2} & 1+\frac{2}{h_r^2} & -\frac{1}{h_r^2} \\ 0 & \cdots & 0 & -\frac{1}{h_r^2} & 1+\frac{1}{h_r^2} \end{bmatrix}$$

with step size $h_r=(b-a)/(N-1)$.

For choosing the regularization parameter ν, we employ the *a posteriori* technique developed in [24]; i.e., we choose the regularization parameter ν iteratively in a geometric manner:

$$\nu_k=\nu_0\cdot\xi^{k-1},$$

where $\nu_0\in(0,1)$ and is provided by the user: for example, $\nu_0=0.5$; $\xi\in(0,1)$ is the factor of proportionality, and k is the kth iteration. It is quite natural to use this parameter selection rule, since in theory ν_k should approach zero as k approaches infinity.

C. Aerosol Particle Size Distribution Function Retrieval

To retrieve the aerosol particle size distribution function $n(r)$, we need to solve linear system (3) for the AOT at different wavelengths. We are interested in the particle size in the interval $[0.1,10]$ μm. Note that a coarse difference gridding ($N\leqslant 20$) 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 [38]; that is, we assume that the actual aerosol particle size distribution function consists of the multiplication of two functions $h(r)$ and $f(f)$: $n(r)=h(r)f(r)$, where $h(r)$ is a rapidly varying function of r that takes the form of a Junge size distribution, while $f(r)$ is more slowly varying. In this way we have

$$\tau_{aero}(\lambda)=\int_a^b [k(r,\lambda,\eta)h(r)]f(r)\mathrm{d}r, \tag{18}$$

where $k(r,\lambda,\eta)=\pi r^2 Q_{ex}(r,\lambda,\eta)$ and we designate $k(r,\lambda,\eta)h(r)$ as the new kernel function that corresponds to a new operator Ξ:

$$(\Xi f)(\lambda)=\tau_{aero}(\lambda). \tag{19}$$

In the following, we consider the numerical performance for solving the abstract model problem (18). For simplicity of notation, the discretization of Ξ is denoted by the matrix $\mathcal{K}$. We omit the discretization details (see [12]) and assume that the original model is already in the discrete form

$$\mathcal{K}\vec{f}=\overrightarrow{\tau_{aero}}. \tag{20}$$

The explanations for the notation of $\mathcal{K}$, $\vec{f}$, and others were given at the end of Section 1.

Now, we can perform our algorithm for the problem

$$\min \frac{1}{2}\|\mathcal{K}\vec{f}-\overrightarrow{\tau_{aero}}\|^2+\nu\|D^{1/2}\vec{f}\|^2 \tag{21}$$

subject to $\vec{f}\geqslant 0$, where $\nu>0$ is the regularization parameter and D represents the data to impose smoothness. Using the technique given in Subsection 3.B, the well-posedness of the computational model is established.

4. NUMERICAL EXPERIMENTS

A. Synthetic Simulation

To verify the feasibility of our inversion method, we have tested it by computer simulation. The simulation consists of two steps. First, a synthetic extinction signal (input signal) is generated by computer according to Eq. (2) for a given value of the particle size distribution $n_{true}(r)$ (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 the input one.

In practice, an exact discretization of the right-hand side $\vec{o}$ of Eq. (2) cannot be obtained accurately; instead a perturbed version is obtained. Numerically, a vector $\vec{d}$ should contain different kinds of noise. Here, for simplicity, we assume that the noise is additive and is mainly Gaussian random noise; that is, we have

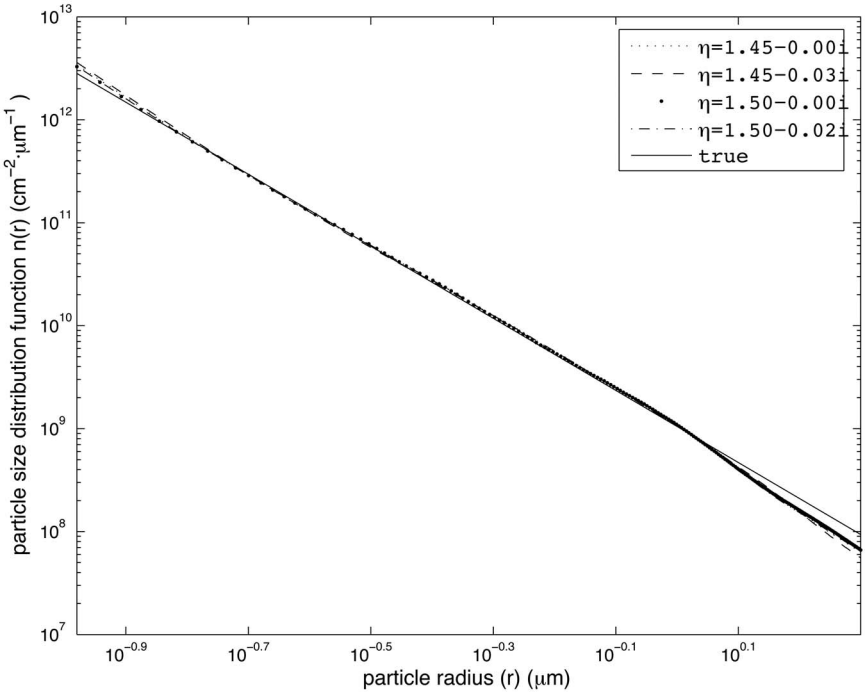

Fig. 1. Input and results retrieved with our inversion method for error level $\delta=0.005$ and different complex refractive indices.

$$\vec{d} = \vec{o} + \delta \times \text{rand}(\text{size}(\vec{o})),$$

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

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

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

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

In our example, the size distribution function $n_{true}(r)$ is given by (see [24])

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

The particle size radius interval of interest is $[0.1, 2]$ μm. The distribution function corresponds to a rapidly changing function $h(r)$ times a slowly varying function $f(r)$. Since most measurements of the continental aerosol particle size distribution reveal that these functions follow a Junge distribution (see [39,40])

$$h(r) = Cr^{-(\nu^*+1)},$$

where ν^* is a shaping constant with typical values in the range 2.0–4.0; therefore, it is reasonable to use an $h(r)$ of Junge type as the weighting factor of $f(r)$. In this work, we choose $\nu^*=3$ and $f(r)=10.5r^{1/2}\exp(-10^{-12}r^{-2})$. The form of this size distribution function is similar to the one given by [15], where a rapidly changing function $h(r)=Cr^{-3}$ can be identified, but it is more similar to a Junge distribution for $r \geqslant 0.1$ μm. One can also generate other particle size distributions and compare the reconstruction with the input. The algorithm works well.

Now we give specifications of the initial input values in Algorithms 3.1 and B.1 (Appendix B) for our theoretical simulation: the initial regularization parameter $\nu_0=0.5$ and the factors of proportionality $\xi=0.5$; then each ν_k is iteratively calculated by the iteration formula given in Subsection 3.B; $\vec{n}_0$ is a vector with components all equaling 0.1; W_0 is chosen by the procedure in Appendix B; the number of discretization nodes is $N=200$. In matrix D, $h_r=(2-0.1)/(N-1)$ is the step size of the grids in $[a,b]$.

First, the complex refractive index η is assumed to be $1.45-0.00i$. Then we invert the same data, supposing that η has an imaginary part. The complex refractive index η is assumed to be $1.45-0.03i$, $1.50-0.00i$, and $1.50-0.02i$. Numerical illustrations are plotted in Figs. 1–3 with noise levels $\delta=0.005$, 0.01, 0.05 for the different complex refractive indices, respectively. Values of rmse for different noise levels and different complex refractive indices are given in Table 1. It can be clearly seen from Table 1 that rmse is of the order of $O(10^{-5})$ to $O(10^{-6})$, which is smaller than that from [24], where rmse values for different noise levels and different complex refractive indices are of the order of $O(10^{-4})$.

Our computer simulation indicates that our method is not too affected by variation of the complex refractive index and noise. The results are comparable with or better than those from [24], where a regularizing damped Gauss–Newton method is used. Therefore we conclude that our method is stable for retrieving aerosol particle size distribution functions.

B. Discussion of Numerical Results

In this subsection, we choose the ground-based data measured by the CE 318 Sun photometer (for illustration of

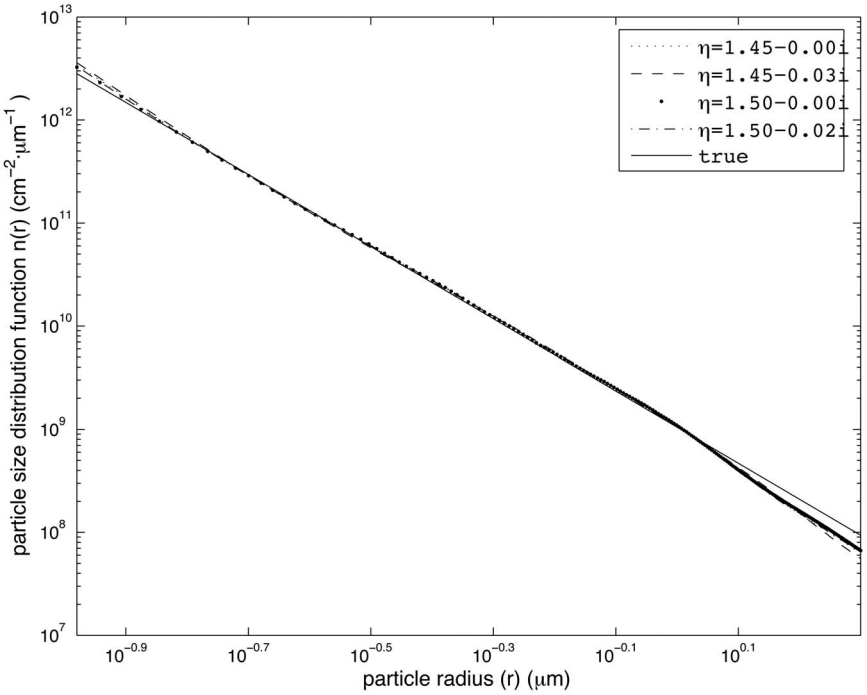

Fig. 2. Input and results retrieved with our inversion method for error level $\delta=0.01$ and different complex refractive indices.

the device, experimental site, and instrument specifications, please refer to [12]) to test the feasibility of the proposed algorithm. We performed successive *in situ* experiments using the CE 318 from October 17–31, 2005. The meteorological data are provided by the Ying Tan Agricultural Ecological Station of CERN (Chinese Ecosystem Research Network).

In this numerical experiment, several days chosen during October 17–31 are used for aerosol inversion. The particle size range $[0.1, 10]$ μm is examined. For an illustration of the air mass history, AOT, and meteorological description, we refer to [24]. By examination of the AOT values in the morning and afternoon on October 17–31, we found that the magnitudes of AOT values on October 17 and 18 are abnormally high. Hence they may not truly reflect the aerosol distribution, and we choose data for other days in this study.

The composition of the atmospheric aerosols from Yin Tan consists of both small and large particles. Both the scattering and the absorption of the particles play a major

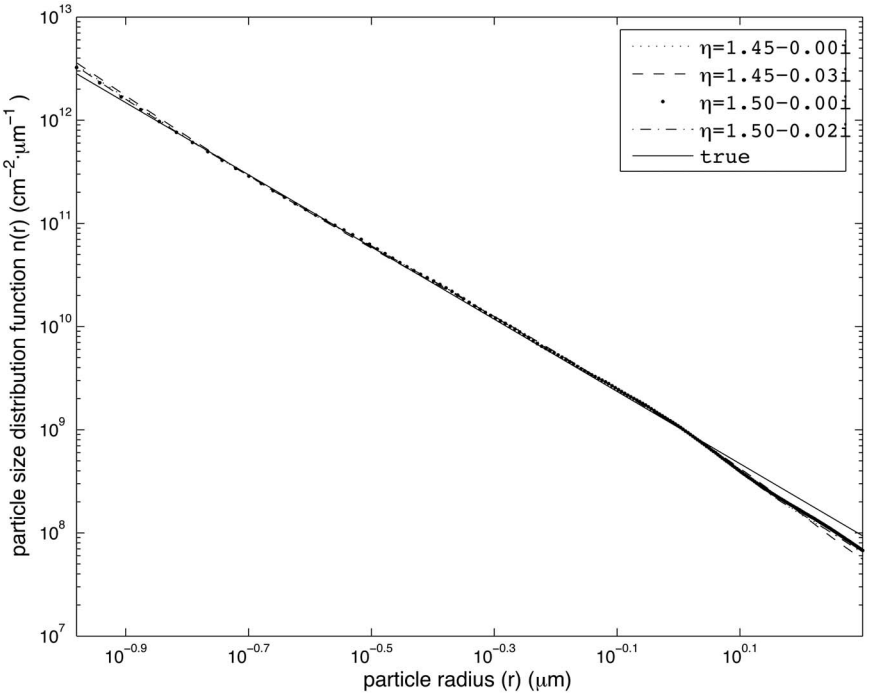

Fig. 3. Input and results retrieved with our inversion method for error level $\delta=0.05$ and different complex refractive indices.

Table 1. Root Mean-Square Errors for Various Noise Levels δ

	Complex Refractive Index η			
δ	$1.45-0.00i$	$1.45-0.03i$	$1.50-0.00i$	$1.50-0.02i$
0.005	7.0873×10^{-6}	7.3227×10^{-6}	6.6037×10^{-6}	6.7025×10^{-6}
0.01	1.4175×10^{-5}	1.4646×10^{-5}	1.3208×10^{-5}	1.3405×10^{-5}
0.05	7.0875×10^{-5}	7.3226×10^{-5}	6.6039×10^{-5}	6.7025×10^{-5}

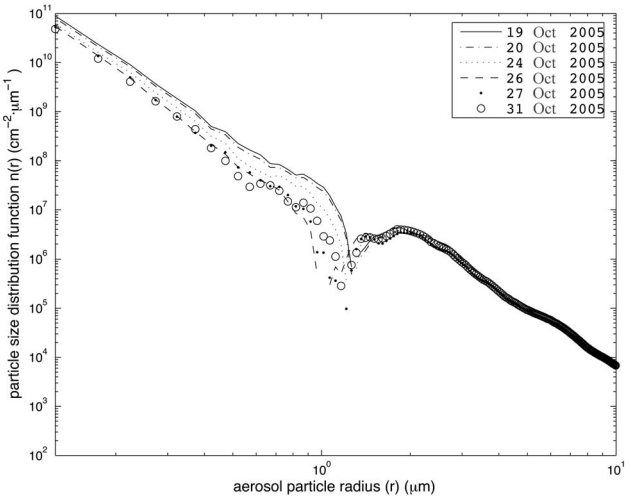

Fig. 4. Particle size distribution in October 2005 (mornings).

part. Therefore, a complex refractive index value of $\eta = 1.50-0.095i$ was used to perform the inversion (see [24]). For theory and methods for determining a complex refractive index η, we refer to [41–43] and the references therein for further information. Now we specify initial input values in Algorithms 3.1 and B.1 for our numerical simulations: all of the initializations are the same as that in the synthetic simulation, except that in matrix D, $h_r = (10-0.1)/(N-1)$ is the step size of the grids in [0.1,10].

The size distribution functions $n(r)$ retrieved by our algorithms are plotted in Fig. 4 for the chosen data from the morning. The figure indicates that the aerosol particles do not decrease rapidly. But it can be seen from the figure that there is a quick jump of the particle size distribution function in the radius interval $[0.8, 2.0]$ μm. Outside that region, the distribution changes smoothly. 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 and that particles of a size near 1.0 μm constitute the primary particles that can be a major source of air contamination. We assume that the large particles of size $[5.0, 10.0]$ μm are composed mainly of clouds 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 nonlinear optimization methods for the solution of the atmospheric aerosol particle size distribution function retrieval problem. We first formulate the regularized quadratic model by imposing nonnegative constraints, then develop an active set method for solving the minimization problem.

We first performed theoretical simulations to verify the feasibility of our inversion method. Our results show that the proposed method is quite 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.

Since the methods developed in this paper are used for an ill-posed aerosol particle size distribution retrieval problem, it may also be suitable for other ill-posed inverse problems. But more investigation is needed because the data and model are different.

APPENDIX A: KARUSH–KUHN–TUCKER CONDITIONS

Consider the general inequality-constrained quadratic programming problem

$$\min Q(x) = \frac{1}{2}x^T G x - p^T x, \tag{A1}$$

$$\text{s.t. } a_i^T x = b_i, \quad i \in E, \tag{A2}$$

$$a_i^T x \geqslant b_i, \quad i \in I, \tag{A3}$$

where $G \in \mathbb{R}^{n\times n}$, E and I are finite sets of indices, and p, x, and $\{a_i\}$ $(i \in E \cup I)$ are vectors with n elements.

The Lagrangian function is defined by

$$L(x,\lambda) = Q(x) - \sum_{i\in E \text{ cup } I} \lambda_i (a_i^T x - b_i). \tag{A4}$$

If we define the active set S at an optimal point x^* as

$$S(x^*) = \{i \in E \cup I : a_i^T x^* = b_i\}, \tag{A5}$$

then the KKT conditions for Eq. (A1) are

$$G x^* - p - \sum_{i\in S(x^*)} \lambda_i^* a_i = 0, \tag{A6}$$

$$a_i^T x^* = b_i, \qquad \text{for all } i \in S(x^*), \tag{A7}$$

$$a_i^T x^* \geqslant b_i, \qquad \text{for all } i \in I \setminus S(x^*), \tag{A8}$$

$$\lambda_i^* \geqslant 0, \qquad \text{for all } i \in I \cap S(x^*). \tag{A9}$$

If there are only $m \leqslant n$ constraints, then Eq. (A1) can be written as

$$\min Q(x) = \frac{1}{2}x^T G x - p^T x, \tag{A10}$$

$$\text{s.t. } Ax = b, \tag{A11}$$

where $A \in \mathbb{R}^{m\times n}$ is defined by $A = [a_i]^T$, $i \in E$. The Lagrangian function is defined by

$$L(x,\lambda) = Q(x) - \lambda^T(Ax - b). \tag{A12}$$

The KKT conditions for Eqs. (A10) and (A11) are given by

$$\begin{bmatrix} G & -A^T \\ A & 0 \end{bmatrix}\begin{bmatrix} x^* \\ \lambda^* \end{bmatrix} = \begin{bmatrix} p \\ b \end{bmatrix}. \tag{A13}$$

The above results are true for nonnegative constraints $x \geqslant 0$ and equality constraints $x = 0$.

APPENDIX B: INSTRUCTIONS ON IMPLEMENTING THE REGULARIZING ACTIVE SET ALGORITHM

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

$$\min J[n], \qquad \text{s.t. } n \in S. \tag{B1}$$

$J[n]$ is a nonlinear function, given in Eq. (7), $S = \{n : n \geqslant 0\}$. At the kth iteration, the search direction s_k is computed from

$$\min Q[s_k], \qquad \text{s.t. } s_k^j = 0,\ j \in W_k. \tag{B2}$$

$Q[s_k]$ is a nonlinear function, given in Eq. (9). Then $n_{k+1} = n_k + \alpha_k s_k$. The gradient of $Q[s_k]$ is denoted $\text{grad}_k[Q] = Gs_k + g_k$.

Initialization for choosing W_0. The initial working set W_0 is related to the initial point n_0. Since the constraint is nonnegative, therefore the components of W_0 can be chosen as i if the ith component of n_0 equals zero, and as zero, otherwise. This means that the ith constraint of W_0 is active. The index i is from 1 to N.

Solving quadratic problem (9). Since we are interested in finding the feasible direction s_k, it is unnecessary to solve Eq. (9) accurately. We apply a feasible direction of descent method with conjugate gradient solution. First, we address the basic concept and procedures of feasible direction of descent methods.

The fundamental concept of feasible direction methods is that of the feasible direction of descent. If $n \in S$, then $s \neq 0$ is called a feasible direction of descent for n if there exists α_{upper} such that for all $\alpha \in (0, \alpha_{\text{upper}})$ the following two properties hold: (1) $n + \alpha s \in S$, (2) $J[n + \alpha s] < J[n]$. Note that condition (2) is equivalent to requiring that $\text{grad}[J]^T s < 0$.

The basic steps in feasible direction methods involve solving a nonlinear programming subproblem to find the direction vector and then finding the step size along this direction by performing a constrained one-dimensional line search. After updating the current point, the above steps are repeated until the termination criterion is satisfied.

Based on the above comments, the feasible direction of Eq. (9) is the vector in null space. There are several ways for solving the nonlinear programming subproblem, say, the steepest descent method, conjugate gradient method, and Newton and quasi-Newton method [35]. Because the model is quadratic, we apply the conjugate gradient method, which is fast and efficient.

Algorithm B.1. (Feasible conjugate gradient algorithm).

Step 1 Input s_0 (such that $n_0 \in S$); Compute $\text{grad}_0[Q] := Gs_0 + g_0$ and such that $\text{grad}_0[Q]$ is a feasible direction. Step 2 If $\|\text{grad}_0[Q]\| \leqslant \epsilon$, output $s^* = s_0$, STOP; Otherwise, set $z_0 := -\text{grad}_0[Q]$,

$$\rho_0 := z_0^T z_0 \quad \text{and} \quad \text{set } k := 1.$$

Step 3 Compute the next iteration points:

$$\alpha_k := \rho_{k-1} / [s_{k-1}^T (G s_{k-1})],$$

$$s_k := s_{k-1} + \alpha_k z_{k-1},$$

$$\text{grad}_k[Q] := \text{grad}_{k-1}[Q] + \alpha_k G s_k (\text{such that } \text{grad}_k[Q] \text{ is a feasible direction}),$$

$$\rho_k := \text{grad}_k[Q]^T \text{grad}_k[Q],$$

$$\beta_k := \rho_k / \rho_{k-1},$$

$$z_k := -\text{grad}_{k-1}[Q] + \beta_k z_{k-1}.$$

Step 4 If $\|\text{grad}_k[Q]\| \leqslant \epsilon$ or k exceeds the maximum iterative steps, output m_k, STOP; Otherwise, set $k := k + 1$, GOTO Step 3.

In our calculation, we choose the initial values such that s_0 is a vector with components all equaling 0.1, $\epsilon = 1.0 \times 10^{-8}$. With the above instructions, users can combine Algorithms 3.1 and B.1 and repeat the experiment easily.

ACKNOWLEDGMENTS

First we thank the reviewers for their helpful comments and suggestions for this paper. The research is supported by China National Science Foundation (NSFC) under grant 10501051 and the Scientific Research Foundation of the Human Resource Ministry. It is also partly supported by NSFC-RFBR cooperation research (2008-2009) and China National 973 project under grant 2007CB714400.

REFERENCES

1. S. Twomey, *Atmospheric Aerosols* (Elsevier, 1977).
2. C. E. Junge, *Air Chemistry and Radioactivity* (Academic, 1963).
3. G. F. Bohren and D. R. Huffman, *Absorption and Scattering of Light by Small Particles* (Wiley, 1983).
4. C. N. Davies, "Size distribution of atmospheric aerosol," J. Aerosol Sci. **5**, 293–300 (1974).
5. G. J. Mccartney, *Optics of Atmosphere* (Wiley, 1976).
6. A. Ångström, "On the atmospheric transmission of sun radiation and on dust in the air," Geogr. Ann. **11**, 156–166 (1929).
7. G. E. Shaw, "Sun photometry," Bull. Am. Meteorol. Soc. **64**, 4–10 (1983).
8. J. A. Curcio, "Evaluation of atmospheric aerosol particle size distribution from scattering measurements in the visible and infrared," J. Opt. Soc. Am. **51**, 548–551 (1961).
9. D. L. Phillips, "A technique for the numerical solution of certain integral equations of the first kind," J. Assoc. Comput. Mach. **9**, 84–97 (1962).
10. S. Twomey, "On the numerical solution of Fredholm integral equations of the first kind by the inversion of the

linear system produced by quadrature," J. Assoc. Comput. Mach. **10**, 97–101 (1963).
11. G. E. Shaw, "Inversion of optical scattering and spectral extinction measurements to recover aerosol size spectra," Appl. Opt. **18**, 988–993 (1979).
12. Y. F. Wang, S. F. Fan, X. Feng, G. J. Yan, and Y. N. Guan, "Regularized inversion method for retrieval of aerosol particle size distribution function in $W^{1,2}$ space," Appl. Opt. **28**, 7456–7467 (2006).
13. K. S. Shifrin and L. G. Zolotov, "Spectral attenuation and aerosol particle size distribution," Appl. Opt. **35**, 2114–2124 (1996).
14. C. Böckmann, "Hybrid regularization method for the ill-posed inversion of multiwavelength lidar data in the retrieval of aerosol size distributions," Appl. Opt. **40**, 1329–1342 (2001).
15. S. Twomey, "Comparison of constrained linear inversion and an iterative nonlinear algorithm applied to the indirect estimation of particle size distributions," J. Comput. Phys. **18**, 188–200 (1975).
16. Y. F. Wang, *Computational Methods for Inverse Problems and Their Applications* (Higher Education Press, 2007).
17. M. T. Chahine, "Inverse problems in radiation transfer: determination of atmospheric parameters," J. Aerosol Sci. **27**, 960–967 (1970).
18. F. Ferri, A. Bassini, and E. Paganini, "Modified version of the Chahine algorithm to invert spectral extinction data for particle sizing," Appl. Opt. **34**, 5829–5839 (1995).
19. G. Yamamoto and M. Tanaka, "Determination of aerosol size distribution function from spectral attenuation measurements," Appl. Opt. **8**, 447–453 (1969).
20. H. Grassl, "Determination of aerosol size distributions from spectral attenuation measurements," Appl. Opt. **10**, 2534–2538 (1971).
21. K. Lumme and J. Rahola, "Light scattering by porous dust particles in the discrete-dipole approximation," Astrophys. J. **425**, 653–667 (1994).
22. C. Böckmann and A. Kirsche, "Iterative regularization method for lidar remote sensing," Comput. Phys. Commun. **174**, 607–615 (2006).
23. A. Voutilainenand and J. P. Kaipio, "Statistical inversion of aerosol size distribution data," J. Aerosol Sci. **31**, 767–768 (2000).
24. Y. F. Wang, S. F. Fan, and X. Feng, "Retrieval of the aerosol particle size distribution function by incorporating *a priori* information," J. Aerosol Sci. **38**, 885–901 (2007).
25. J. Heintzenberg, "Properties of log-normal particle size distributions," Aerosol Sci. Technol. **21**, 46–48 (1994).
26. D. L. Wright, "Retrieval of optical properties of atmospheric aerosols from moments of the particle size distribution," J. Aerosol Sci. **31**, 1–18 (2000).
27. M. Ye, S. Wang, Y. Lu, T. Hu, Z. Zhu, and Y. Xu, "Inversion of particle-size distribution from angular light-scattering data with genetic algorithms," Appl. Opt. **38**, 2677–2685 (1999).
28. D. A. Ligon, J. B. Gillespie, and P. Pellegrino, "Aerosol properties from spectral extinction and backscatter estimated by an inverse Monte Carlo method," Appl. Opt. **39**, 4402–4410 (2000).
29. G. Ramachardran, D. Lieth, and L. Todd, "Extraction of spatial aerosol distribution from multispectral light extinction measurements with computed tomography," J. Opt. Soc. Am. A **11**, 144–154 (1994).
30. T. Nguyen and K. Cox, "A method for the determination of aerosol particle distributions from light extinction data," in *Abstracts of the American Association for Aerosol Research Annual Meeting* (American Association of Aerosol Research, 1989), pp. 330.
31. A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems* (Wiley, 1977).
32. T. Y. Xiao, Sh. G. Yu, and Y. F. Wang, *Numerical Methods for the Solution of Inverse Problems* (Science, 2003).
33. C. R. Vogel, *Computational Methods for Inverse Problems* (SIAM, 2002).
34. A. Bakushinsky and A. Goncharsky, *Ill-Posed Problems: Theory and Applications* (Kluwer Academic, 1994).
35. Y. X. Yuan, *Numerical Methods for Nonlinear Programming* (Shanghai Science and Technology, 1993).
36. J. Nocedal and S. J. Wright, *Numerical optimization* (Springer, 1999).
37. R. Fletcher, *Pratical Methods of Optimization* (Wiley, 1987).
38. M. D. King, D. M. Byrne, B. M. Herman, and J. A. Reagan, "Aerosol size distributions obtained by inversion of spectral optical depth measurements," J. Atmos. Sci. **35**, 2153–2167 (1978).
39. C. E. Junge, "The size distribution and aging of natural aerosols as determined from electrical and optical data on the atmosphere," J. Meteorol. **12**, 13–25 (1955).
40. W. E. Clark and K. T. Whitby, "Concentration and size distribution measurements of atmospheric aerosols and a test of the theory of self-preserving size distributions," J. Atmos. Sci. **24**, 677–687 (1967).
41. A. J. Reagan, D. M. Byrne, D. M. King, J. D. Spinhirne, and B. M. Herman, "Determination of complex refractive index and size distribution of atmospheric particles from bistatic-monostatic lidar and solar radiometer measurements," J. Geophys. Res. **85**, 1591–1599 (1980).
42. M. Weindisch and W. von Hoyningen-Huen, "Possibility of refractive index determination of atmospheric aerosol particles by ground-based solar extinction and scattering measurements," Atmos. Environ. **28**, 785–792 (1994).
43. F. Q. Yan, H. L. Hu, and J. Zhou, "Measurements of number density distribution and imaginary part of refractive index of aerosol particles," Acta Opt. Sin. **23**, 854–859 (2003).