

26 Quantitative Remote Sensing Inversion in Earth Science: Theory and Numerical Treatment

Yanfei Wang

Key Laboratory of Petroleum Geophysics, Institute of Geology and Geophysics,
Chinese Academy of Sciences, Beijing, People's Republic
of China

1	Introduction	787
2	Typical Inverse Problems in Earth Science	789
2.1	Land Surface Parameter Retrieval Problem.....	789
2.2	Backscatter Cross-Section Inversion with Lidar.....	790
2.3	Aerosol Inverse Problems.....	791
3	Regularization	792
3.1	What Causes Ill-Posedness.....	792
3.2	Imposing a Priori Constraints on the Solution.....	794
3.3	Tikhonov/Phillips–Twomey’s Regularization.....	794
3.3.1	Choices of the Scale Operator D	795
3.3.2	Regularization Parameter Selection Methods.....	796
3.4	Direct Regularization.....	797
3.5	Statistical Regularization.....	798
4	Optimization	799
4.1	Sparse/Nonsmooth Inversion in l_1 Space.....	799
4.2	Optimization Methods for l_2 Minimization Model.....	801
4.2.1	Newton-Type Methods.....	801
4.2.2	Gradient-Type Methods.....	802
5	Practical Applications	803
5.1	Kernel-Based BRDF Model Inversion.....	803
5.1.1	Inversion by NTSVD.....	803
5.1.2	Tikhonov Regularized Solution.....	803
5.1.3	Land Surface Parameter Retrieval Results.....	803

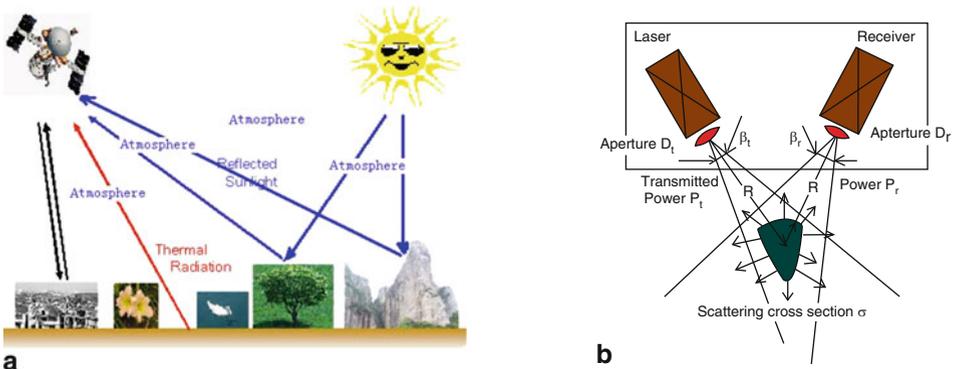
5.2	Inversion of Airborne Lidar Remote Sensing.....	805
5.3	Particle Size Distribution Function Retrieval.....	806
6	<i>Conclusion</i>	809
	<i>References</i>	810

Abstract Quantitative remote sensing is an appropriate way to estimate structural parameters and spectral component signatures of Earth surface cover type. Since the real physical system that couples the atmosphere, water and the land surface is very complicated and should be a continuous process, sometimes it requires a comprehensive set of parameters to describe such a system, so any practical physical model can only be approximated by a mathematical model which includes only a limited number of the most important parameters that capture the major variation of the real system. The pivot problem for quantitative remote sensing is the inversion. Inverse problems are typically ill-posed. The ill-posed nature is characterized by: (C1) the solution may not exist; (C2) the dimension of the solution space may be infinite; (C3) the solution is not continuous with variations of the observed signals. These issues exist nearly for all inverse problems in geoscience and quantitative remote sensing. For example, when the observation system is band-limited or sampling is poor, i.e., there are too few observations, or directions are poor located, the inversion process would be underdetermined, which leads to the large condition number of the normalized system and the significant noise propagation. Hence (C2) and (C3) would be the highlight difficulties for quantitative remote sensing inversion. This chapter will address the theory and methods from the viewpoint that the quantitative remote sensing inverse problems can be represented by kernel-based operator equations and solved by coupling regularization and optimization methods.

1 Introduction

Both modeling and model-based inversion are important for quantitative remote sensing. Here, modeling mainly refers to data modeling, which is a method used to define and analyze data requirements; model-based inversion mainly refers to using physical or empirically physical models to infer unknown but interested parameters. Hundreds of models related to atmosphere, vegetation, and radiation have been established during past decades. The model-based inversion in geophysical (atmospheric) sciences has been well understood. However, the model-based inverse problems for Earth surface received much attention by scientists only in recent years. Compared to modeling, model-based inversion is still in the stage of exploration (Wang et al. 2009c). This is because that intrinsic difficulties exist in the application of *a priori* information, inverse strategy, and inverse algorithm. The appearance of hyperspectral and multiangular remote sensor enhanced the exploration means, and provided us more spectral and spatial dimension information than before. However, how to utilize these information to solve the problems faced in quantitative remote sensing to make remote sensing really enter the time of quantification is still an arduous and urgent task for remote sensing scientists. Remote sensing inversion for different scientific problems in different branch is being paid more and more attentions in recent years. In a series of international study projections, such as International Geosphere-Biosphere Programme (IGBP), World Climate Research Programme (WCRP), and NASA's Earth Observing System (EOS), remote sensing inversion has become a focal point of study.

Model-based remote sensing inversions are usually optimization problems with different constraints. Therefore, how to incorporate the method developed in operation research field into remote sensing inversion field is very much needed. In quantitative remote sensing, since the real physical system that couples the atmosphere and the land surface is very complicated (see  Fig. 1a) and should be a continuous process, sometimes it requires a comprehensive set of



a
Fig. 1

Remote observing the Earth (a); geometry and parameters for laser scanner (b)

parameters to describe such a system, so any practical physical model can only be approximated by a model which includes only a limited number of the most important parameters that capture the major variation of the real system. Generally speaking, a discrete forward model to describe such a system is in the form

$$y = h(\mathbf{x}, \mathbf{S}), \quad (1)$$

where y is single measurement, \mathbf{x} is a vector of controllable measurement conditions such as wave band, viewing direction, time, Sun position, polarization, and the forth, \mathbf{S} is a vector of state parameters of the system approximation, and h is a function which relates \mathbf{x} with \mathbf{S} , which is generally nonlinear and continuous.

With the ability of satellite sensors to acquire multiple bands, multiple viewing directions, and so on, while keeping \mathbf{S} essentially the same, we obtain the following nonhomogeneous equations

$$\mathbf{y} = h(\mathbf{x}, \mathbf{S}) + \mathbf{n}, \quad (2)$$

where \mathbf{y} is a vector in \mathbb{R}^M , which is an M dimensional measurement space with M values corresponding to M different measurement conditions, $\mathbf{n} \in \mathbb{R}^M$ is the vector of random noise with same vector length M . Assume that there are m undetermined parameters need to be recovered. Clearly, if $M = m$, (2) is a determined system, so it is not difficult to develop some suitable algorithms to solve it. If more observations can be collected than the existing parameters in the model (Verstraete et al. 1996), i.e., $M > m$, the system (2) is over determined. In this situation, the traditional solution does not exist. We must define its solution in some other meaning, for example, the least squares error (LSE) solution. However as Li in (Li et al. 1998) pointed out that, "for physical models with about ten parameters (single band), it is questionable whether remote sensing inversion can be an over determined one in the foreseeable future." Therefore, the inversion problems in geosciences seem to be always underdetermined in some sense. Nevertheless, the underdetermined system in some cases, can be always converted to an overdetermined one by utilizing multiangular remote sensing data or by accumulating some *a priori* knowledge (Li et al. 2001).

Developed methods in literature for quantitative remote sensing inversion are mainly statistical methods with several variations from Bayesian inference. In this chapter, using kernel expression, we analyze from algebraic point of view, about the solution theory and methods for

quantitative remote sensing inverse problems. The kernels mentioned in this chapter mainly refer to integral kernel operators (characterized by integral kernel functions) or discrete linear operators (characterized by finite rank matrices). It is closely related with the kernels of linear functional analysis, Hilbert space theory, and spectral theory. In particular, we present regularizing retrieval of parameters with *a posteriori* choice of regularization parameters, several cases of choosing scale/weighting matrices to the unknowns, numerically truncated singular value decomposition (NTSVD), nonsmooth inversion in l_p space and advanced optimization techniques. These methods, as far as we know, are novel to literature in Earth science.

The outline of this chapter is as follows: in [Sect. 2](#), we list three typical kernel-based remote sensing inverse problems. One is the linear kernel-based bidirectional reflectance distribution function (BRDF) model inverse problem, which is of great importance for land surface parameters retrieval; the other is the backscattering problem for Lidar sensing; and the last one is aerosol particle size distributions from optical transmission or scattering measurements, which is a long time existed problem and still an important topic today. In [Sect. 3](#), the regularization theory and solution techniques for ill-posed quantitative remote sensing inverse problems are described. [Section 3.1](#) introduces the conception of well-posed problems and ill-posed problems; [Sect. 3.2](#) discusses about the constrained optimization; [Sect. 3.3](#) fully extends the Tikhonov regularization; [Sect. 3.4](#) discusses about the direct regularization methods for equality-constrained problem; then in [Sect. 3.5](#), the regularization scheme formulated in the Bayesian statistical inference is introduced. In [Sect. 4](#), the optimization theory and solution methods are discussed for finding an optimized solution of a minimization model. [Section 4.1](#) talks about sparse and nonsmooth inversion in l_1 space; [Sect. 4.2](#) introduces Newton-type and gradient-type methods. In [Sect. 5.1](#), the detailed regularizing solution methods for retrieval of ill-posed land surface parameters are discussed. In [Sect. 5.2](#), the results for retrieval of backscatter cross-sections by Tikhonov regularization are displayed. In [Sect. 5.3](#), the regularization and optimization methods for recovering aerosol particle size distribution functions are presented. Finally, in [Sect. 6](#), some concluding remarks are given.

2 Typical Inverse Problems in Earth Science

Many inverse problems in geophysics are kernel-based, e.g., problems in seismic exploration and gravimetry. I do not introduce these solid Earth problems in the present chapter, instead, I mainly focus on Earth surface problems. The kernel methods can increase the accuracy of remote-sensing data processing, including specific land-cover identification, biophysical parameter estimation, and feature extraction (Camps-Valls 2008; Wang et al. 2009c). I introduce three typical kernel-based inverse problems in geoscience, one belongs to the atmospheric problem, another two belong to the Earth surface problems.

2.1 Land Surface Parameter Retrieval Problem

As is well-known, the anisotropy of the land surface can be best described by the BRDF. With the progress of the multiangular remote sensing, it seems that the BRDF models can be inverted to estimate structural parameters and spectral component signatures of Earth surface cover

type (see Strahler et al. 1994; Roujean et al. 1992). The state of the art of BRDF is the use of the linear kernel-driven models, mathematically described as the linear combination of the isotropic kernel, volume scattering kernel, and geometric optics kernel. The information extraction on the terrestrial biosphere and other problems for retrieval of land surface albedos from satellite remote sensing have been considered by many authors in recent years, see for instance the survey papers on the kernel-based BRDF models by (Pokrovsky and Roujean (2002), Pokrovsky et al. (2003), Pokrovsky and Roujean (2003), and references therein. The computational stability is characterized by the algebraic operator spectrum of the kernel-matrix and the observation errors. Therefore, the retrieval of the model coefficients is of great importance for computation of the land surface albedos.

The linear kernel-based BRDF model can be described as follows (Roujean et al. 1992):

$$f_{\text{iso}} + k_{\text{vol}}(t_i, t_v, \phi) f_{\text{vol}} + k_{\text{geo}}(t_i, t_v, \phi) f_{\text{geo}} = r(t_i, t_v, \phi), \quad (3)$$

where r is the bidirectional reflectance; the kernels k_{vol} and k_{geo} are so-called kernels, that is, known functions of illumination and of viewing geometry which describe volume and geometric scattering, respectively; t_i and t_v are the zenith angle of the solar direction and the zenith angle of the view direction respectively; ϕ is the relative azimuth of sun and view direction; and f_{iso} , f_{vol} , and f_{geo} are three unknown parameters to be adjusted to fit observations. Theoretically, f_{iso} , f_{vol} , and f_{geo} are closely related to the biomass such as leaf area index (LAI), Lambertian reflectance, sunlit crown reflectance, and viewing and solar angles. The vital task then is to retrieve appropriate values of the three parameters.

Generally speaking, the BRDF model includes kernels of many types. However, it was demonstrated that the combination of RossThick (k_{vol}) and LiSparse (k_{geo}) kernels had the best overall ability to fit BRDF measurements and to extrapolate BRDF and albedo (see, e.g., Wanner et al. 1995; Li et al. 1999; Privette et al. 1997). A suitable expression for the RossThick kernel k_{vol} was derived by Roujean et al. (1992). It is reported that the LiTransit kernel k_{Transit} , instead of the kernel k_{geo} , is more robust and stable than LiSparse non-reciprocal kernel and the reciprocal LiSparse kernel k_{sparse} (LiSparseR) where the LiTransit kernel and the LiSparse kernel are related by $k_{\text{Transit}} = \begin{cases} k_{\text{sparse}}, & B \leq 2, \\ \frac{2}{B} k_{\text{sparse}}, & B > 2, \end{cases}$ and B is given by $B := B(t_i, t_v, \phi) = -O(t_i, t_v, \phi) + \sec t'_i + \sec t'_v$ in Li et al. (2000). More detailed explanation about O and t' in the definition of k_{Transit} can be found in Wanner et al. (1995).

To use the combined linear kernel model, a key issue is to numerically solve the inverse model in a stable way. However, it is difficult to do in practical applications due to ill-posed nature of the inverse problem. So far, statistical methods and algebraic methods have been developed for solving this inverse problem, e.g., Pokrovsky and Roujean (2002, 2003) and Wang et al. (2007a, 2008). We will describe these methods and introduce recent advances in following paragraphs.

2.2 Backscatter Cross-Section Inversion with Lidar

Airborne laser scanning (ALS) is an active remote sensing technique which is also often referred to as Lidar or laser radar. Due to the increasing availability of sensors, ALS has been receiving increasing attention in recent years (e.g., see Wagner et al. 2006). In ALS a laser emits short infrared pulses toward the Earth's surface and a photodiode records the backscattered echo.

With each scan, measurements are taken of the round trip time of the laser pulse, the received echo power and of the beam angle in the locator's coordinate system. The round-trip time of the laser pulse allows calculating the range (distance) between the laser scanner and the object that generated the backscattered echo. Thereby, information about the geometric structure of the Earth's surface is obtained. The received power provides information about the scattering properties of the targets, which can be exploited for object classification and for modeling of the scattering properties.

The latest generation of ALS systems does not only record a discrete number of echoes but also digitizes the whole waveform of the reference pulse and the backscattered echoes. In this way, besides the range further echo parameters can be determined. The retrieval of the backscatter cross-section is of great interest in full-waveform ALS. Since it is calculated by deconvolution, its determination is an ill-posed problem in a general sense.

ALS utilizes a measurement principle firstly strongly related to radar remote sensing (see [Fig. 1b](#)). The fundamental relation to explain the signal strength in both techniques is the radar equation (Wagner et al. 2006):

$$P_r(t) = \frac{D_r^2}{4\pi R^4 \beta_t^2} P_t \left(t - \frac{2R}{v_g} \right) \sigma, \quad (4)$$

where t is the time, R is the range, D_r is the aperture diameter of the receiver optics, β_t is the transmitter beam width, P_t is the transmitted power of the laser, and σ denotes the scattering cross-section. The time delay is equal to $t' = 2R/v_g$ where v_g is the group velocity of the laser pulse in the atmosphere.

Taking the occurrence of multiple scatterers into account and regarding the impulse response $\Gamma(t)$ of the system's receiver, we get (Wagner et al. 2006)

$$P_r(t) = \sum_{i=1}^N \frac{D_r^2}{4\pi R^4 \beta_t^2} P_t(t) * \sigma'_i(t) * \Gamma(t), \quad (5)$$

where $*$ denotes the convolution operator. Since convolution is commutative, we can set $P_t(t) * \sigma'_i(t) * \Gamma(t) = P_t(t) * \Gamma(t) * \sigma'_i(t) = S(t) * \sigma'_i(t)$, i.e., it is possible to combine both the transmitter and the receiver characteristics to a single term $S(t)$. This term is referred to as the system waveform. Thus, we are able to write our problem in the form (Wang et al. 2009a)

$$h(t) = \sum_{i=1}^N (f * g)(t). \quad (6)$$

where h is the incoming signal recorded by the receiver, f denotes a mapping which specifies the kernel function or point spread function, and g is the unknown cross-section. The problem is how to deconvolve the convolution equation (6) to get the approximation to the actual cross-section.

2.3 Aerosol Inverse Problems

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. 1966). So, the determination of

particle size distribution function becomes a basic task in aerosol research (see, e.g., Twomey 1977; Davies 1974; Bohren and Huffman 1983; Mccartney 1976; Bockmann 2001; Bockmann and Kirsche 2006). Since the relationship between the size of atmospheric aerosol particles and the wavelength dependence of the extinction coefficient was first suggested by Angström (Angström 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

$$\tau_{\text{aero}}(\lambda) = \int_0^{\infty} \pi r^2 Q_{\text{ext}}(r, \lambda, \eta) n(r) dr + \varrho(\lambda), \quad (7)$$

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); $l\eta$ is the complex refractive index of the aerosol particles; λ is the wavelength; $\varrho(\lambda)$ is the error/noise; and $Q_{\text{ext}}(r, \lambda, \eta)$ 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.

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., see Voutilainen and Kaipio (2000), Wang et al. (2006a), Wang (2007, 2008), Nguyen and Cox (1989), Wang and Yang (2008). For the aerosol attenuation problem (7), let us rewrite (7) in the form of the abstract operator equation

$$\begin{aligned} K: \mathcal{X} &\longrightarrow \mathcal{Y}, \\ (Kn)(\lambda) + \varrho(\lambda) &= \int_0^{\infty} k(r, \lambda, \eta) n(r) dr + \varrho(\lambda) = o(\lambda) + \varrho(\lambda) = d(\lambda), \end{aligned} \quad (8)$$

where $k(r, \lambda, \eta) = \pi r^2 Q_{\text{ext}}(r, \lambda, \eta)$; \mathcal{X} denotes the function space of aerosol size distributions; and \mathcal{Y} denotes the observation space. Both \mathcal{X} and \mathcal{Y} are considered to be the separable Hilbert space. Note that τ_{aero} in Eq. 7 is the measured term, it inevitably induces noise/errors. Hence, $d(\lambda)$ is actually a perturbed right-hand side. Keeping in mind operator symbol, Eq. 8 can be written as

$$Kn + \varrho = o + \varrho = dl. \quad (9)$$

3 Regularization

3.1 What Causes Ill-Posedness

From this section till the end of the chapter, unless it is specified, we will denote the operator equation as

$$K(x) = y, \quad (10)$$

which is an appropriate expression for an observing system, with K the response function (linear or nonlinear), x the unknown input and y the observed data. Particularly, if K is a linear mapping, we will denote the response system as

$$Kx = y, \quad (11)$$

which is clearly a special case of (10). We will also use K as a operator in infinite spaces sometimes, and a matrix sometimes. We assume that readers can readily recognize them.

The problem (10) is said to be properly posed or well-posed in the sense that it has the following three properties:

(C₁) There exists a solution of the problem, i.e., existence;

(C₂) There is at most one solution of the problem, i.e., uniqueness;

(C₃) The solution depends continuously on the variations of the right hand side (data), i.e., stability.

The condition (C₁) can be easily fulfilled if we enlarge the solution space of the problem (10). The condition (C₂) is seldom satisfied for many indirectly measurement problems. This means more than one solution may be found for the problem (10) and the information about the model is missing. In this case, *a priori* knowledge about the solution must be incorporated and built into the model. The requirement of stability is the most important one. If the problem (10) lacks the property of stability, then the computed solution has nothing to do with the true solution since the practically computed solution is contaminated by unavailable errors. Therefore, there is no way to overcome this difficulty unless additional information about the solution is available. Again, *a priori* knowledge about the solution should be involved.

If problem (10) is well-posed, then K has a well-defined, continuous inverse operator K^{-1} . In particular, $K^{-1}(K(x)) = x$ for any $x \in \mathcal{X}$ and $\text{Range}(K) = \mathcal{Y}$. In this case both the algebraic nature of the spaces and the topologies of the spaces are ready to be employed.

The particle size distribution model (9) is a linear model in infinite spaces. The operator K is compact. The ill-posedness is self-evident because that at least one of the three items for well-posed problems is violated. Note that (3) is a linear model in finite spaces, therefore it is easy to rewrite it into a finite rank operator equation

$$K\mathbf{x} = \mathbf{y}, \quad (12)$$

by setting $\mathbf{x} = [f_{\text{iso}}, f_{\text{vol}}, f_{\text{geo}}]^T$ and $\mathbf{y} = [y_j]$ with the entries $y_j = r_j(t_i, t_v, \phi)$, where \mathbf{y} is the measurement data. The inverse problem is how to recover the model parameters \mathbf{x} given the limited measurement data \mathbf{y} . For Lidar backscatter cross-section inversion, one needs to solve a deconvolution problem. The ill-posedness is due to the ill-conditioning of the spectrum of the operator and noisy data.

Numerically, the discrete ill-posedness for the above examples is because that their operators may be inaccurate (can only be approximately calculated), their models are usually underdetermined if there are too few observations or poor directional range, or the observations are highly linearly dependent and noisy. For example, a single angular observation may lead to a under determined system whose solutions are infinite (the null space of the kernel contains nonzero vectors) or the system has no solution (the rank of the coefficient matrix is not equal to the augmented matrix). In practice, random uncertainty in the reflectances sampled translates into uncertainty in the BRDF and albedo. We note that noise inflation depends on the sampling geometry alone. For example, for MODIS and MISR sampling, they vary with latitude and time of year; but for kernel-based models, they do not depend on wavelength or the type of BRDF viewed. Therefore, the random noise in the observation (BRDF) and the small singular values of K control the error propagation.

3.2 Imposing a Priori Constraints on the Solution

For effective inversion of the ill-posed kernel driven model, we have to impose an *a priori* constraint to the interested parameters. This leads to solving a constrained LSE problem

$$\min J(\mathbf{x}), \text{ s.t. } K\mathbf{x} = \mathbf{y}, \quad \Delta_1 \leq c(\mathbf{x}) \leq \Delta_2, \quad (13)$$

where $J(\mathbf{x})$ denotes an object functional, which is a function of \mathbf{x} , $c(\mathbf{x})$ is the constraint to the solution \mathbf{x} , Δ_1 and Δ_2 are two constants which specify the bounds of $c(\mathbf{x})$. Usually, $J(\mathbf{x})$ is chosen as the norm of \mathbf{x} with different scale. If the parameter \mathbf{x} comes from a smooth function, then $J(\mathbf{x})$ can be chosen as a smooth function, otherwise, $J(\mathbf{x})$ can be nonsmooth.

The constraint $c(\mathbf{x})$ can be smooth (e.g., Sobolev stabilizer) or nonsmooth (e.g., total variation or l_q norm ($q \neq 2$) based stabilizer). A generically used constraint is the smoothness. It assumes that physical properties in a neighborhood of space or in an interval of time present some coherence and generally do not change abruptly. Practically, we can always find regularities of a physical phenomenon with respect to certain properties over a short period of time (Wang et al. 2007a, 2008). The smoothness *a priori* has been one of the most popular *a priori* assumptions in applications. The general framework is the so-called regularization which will be explained in the next subsection.

3.3 Tikhonov/Phillips–Twomey’s Regularization

Most of inverse problems in real environment are generally ill-posed. Regularization methods are widely-used to solve such ill-posed problems. The complete theory for regularization was developed by Tikhonov and his colleagues (Tikhonov and Arsenin 1977). For the discrete model (12), we suppose \mathbf{y} is the true right-hand side, and denote \mathbf{y}_n the measurements with noise which represents the bidirectional reflectance. The Tikhonov regularization method is to solve a regularized minimization problem

$$J^\alpha(\mathbf{x}) := \|K\mathbf{x} - \mathbf{y}_n\|_2^2 + \alpha \|D^{1/2}\mathbf{x}\|_2^2 \longrightarrow \min \quad (14)$$

instead of solving

$$J(\mathbf{x}) = \|K\mathbf{x} - \mathbf{y}_n\|_2^2 \longrightarrow \min. \quad (15)$$

In (14), α is the regularization parameter and D is a positively (semi-)definite operator. By a variational process, the minimizer of (14) satisfies

$$K^T K\mathbf{x} + \alpha D\mathbf{x} = K^T \mathbf{y}_n. \quad (16)$$

The operator D is a scale matrix which imposes smoothness constraint to the solution \mathbf{x} . The scale operator D and the regularization parameter α can be considered as some kind of *a priori* information, which will be discussed next.

Phillips–Twomey’s regularization is based on solving the problem (Phillips 1962; Twomey 1975)

$$\min_{\mathbf{x}} Q(\mathbf{x}), \text{ s.t. } \|K\mathbf{x} - \mathbf{y}_n\| = \Delta, \quad (17)$$

where $Q(\mathbf{x}) = (D\mathbf{x}, \mathbf{x})$, where D is a preassigned scale matrix and $\Delta > 0$. It is clear that Phillips–Twomey’s regularization shares similarity with Tikhonov’s regularization and can be written in consistent form.

3.3.1 Choices of the Scale Operator D

To regularize the ill-posed problem discussed in [Sect. 3.1](#), the choice of the scale operator D has great impact to the performance to the regularization. Note that the matrix D plays the role in imposing a smoothness constraint to the parameters and in improving the condition of the spectrum of the adjoint operator $K^T K$. Therefore, it should be positively definite or at least positively semi-definite. One may readily see that the identity may be a choice. However this choice does not fully employ the assumption about the continuity of the parameters.

In Wang et al. (2007a), we assume that the operator equation (12) is the discretized version of a continuous physical model

$$K(x(\tau)) = y(\tau) \quad (18)$$

with K the linear/nonlinear operator, $x(\tau)$ the complete parameters describing the land surfaces and y the observation. Most of the kernel model methods reported in literature may have the above formulation. Hence instead of establishing regularization for the operator equation (12) in the Euclidean space, it is more convenient to perform the regularization to the operator equation (18) on an abstract space. So from *a priori* considerations we suppose that the parameters x is a smooth function, in the sense that x is continuous on $[a, b]$, is differentiable almost everywhere and its derivative is square-integrable on $[a, b]$. By Sobolev's imbedding theorem (see, e.g., Tikhonov and Arsenin 1977; Xiao et al. 2003), the continuous differentiable function x in $W^{1,2}$ space imbeds into integrable continuous function space L_2 automatically. The inner product of two functions $x(\tau)$ and $y(\tau)$ in $W^{1,2}$ space is defined by

$$(x(\tau), y(\tau))_{W^{1,2}} := \int_{\Omega} \left(x(\tau)y(\tau) + \sum_{i=1}^n \frac{\partial x}{\partial \tau_i} \frac{\partial y}{\partial \tau_j} \right) d\tau_1 d\tau_2 \dots d\tau_n, \quad (19)$$

where Ω is the assigned interval of the definition.

Now we construct a regularizing algorithm that an approximate solution $x^\alpha \in W^{1,2}[a, b]$ which converges, as error level approaching zero, to the actual parameters in the norm of space $W^{1,2}[a, b]$, precisely we construct the functional

$$J^\alpha(x) = \rho_F[Kx, y] + \alpha L(x), \quad (20)$$

where $\rho_F[Kx, y] = \frac{1}{2} \|Kx - y\|_{L_2}^2$, $L(x) = \frac{1}{2} \|x\|_{W^{1,2}}^2$.

Assume that the variation of $x(\tau)$ is flat and smooth near the boundary of the integral interval $[a, b]$. In this case, the derivatives of x are zeros at the boundary of $[a, b]$. Let h_r be the step size of the grids in $[a, b]$, which could be equidistant or adaptive. Then after discretization of $L(x)$, D is a tridiagonal matrix in the form

$$D := D_1 = \begin{bmatrix} 1 + \frac{1}{h_r^2} & -\frac{1}{h_r^2} & 0 & \dots & 0 \\ -\frac{1}{h_r^2} & 1 + \frac{2}{h_r^2} & -\frac{1}{h_r^2} & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & -\frac{1}{h_r^2} & 1 + \frac{2}{h_r^2} & -\frac{1}{h_r^2} \\ 0 & \dots & 0 & -\frac{1}{h_r^2} & 1 + \frac{1}{h_r^2} \end{bmatrix}.$$

For the linear model (3), after the kernel normalization, we may consider $[a, b] = [-1, 1]$. Thus, D is in the above form with $h_r = 2/(N - 1)$.

There are many kinds of techniques for choosing the scale matrix D appropriately. In Phillips–Twomey's formulation of regularization (see, e.g., Wang et al. 2006), the matrix D is

created by the norm of the second differences, $\sum_{i=2}^{N-1} (x_{i-1} - 2x_i + x_{i+1})^2$, which leads to the following form of matrix D

$$D := D_2 = \begin{bmatrix} 1 & -2 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ -2 & 5 & -4 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 1 & -4 & 6 & -4 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 1 & -4 & 6 & -4 & 1 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 & -4 & 6 & -4 & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 & -4 & 6 & -4 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 1 & -4 & 5 & -2 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 1 & -2 & 1 \end{bmatrix}.$$

However, the matrix D is badly conditioned and thus the solution to minimize the functional $J^\alpha[x]$ with D as the smooth constraint is observed to have some oscillations (Wang et al. 2006). Another option is the negative Laplacian (see, e.g., Wang and Yuan 2003; Wang 2007): $Lx := -\sum_{i=1}^n \frac{\partial^2 x}{\partial \tau_i^2}$, for which the scale matrix D for the discrete form of the negative Laplacian Lx is

$$D := D_3 = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix}.$$

Where we assume that the discretization step length as 1. The scale matrix D_3 is positive semi-definite but not positive definite and hence the minimization problem may not work efficiently for severely ill-posed inverse problems. Another option of the scale matrix D is the identity, i.e., $D := D_4 = \text{diag}(e)$, where e is the components of all ones, however this scale matrix is too conservative and may lead to over regularization.

3.3.2 Regularization Parameter Selection Methods

As noted above, the choice of the regularization parameter α is important to tackle the ill-posedness. *A priori* choice of the parameter α allows $0 < \alpha < 1$. However the *a priori* choice of the parameter does not reflect the degree of approximation that may lead to either over estimate or under estimate of the regularizer.

We will use the widely-used discrepancy principle (see, e.g., Tikhonov and Arsenin 1977; Tikhonov et al. 1995; Xiao et al. 2003) to find an optimal regularization parameter. In fact, the optimal parameter α^* is a root of the nonlinear function

$$\Psi(\alpha) = \|K\mathbf{x}_\alpha - \mathbf{y}_n\|^2 - \delta^2, \quad (21)$$

where δ is the error level to specify the approximate degree of the observation to the true noiseless data, \mathbf{x}_α denotes the solution of the problem in Eq. (16) corresponding to the value α of the related parameter. Noting $\Psi(\alpha)$ is differentiable, fast algorithms for solving the optimal parameter α^* can be implemented. In this chapter we will use the cubic convergent algorithm developed in (Wang and Xiao 2001):

$$\alpha_{k+1} = \alpha_k - \frac{2\Psi(\alpha_k)}{\Psi'(\alpha_k) + (\Psi'(\alpha_k))^2 - 2\Psi(\alpha_k)\Psi''(\alpha_k)}^{\frac{1}{2}}. \quad (22)$$

In the above cubic convergent algorithm, the function $\Psi'(\alpha)$ and $\Psi''(\alpha)$ have the following explicit expression:

$$\Psi'(\alpha) = -\alpha\beta'(\alpha), \quad \Psi''(\alpha) = -\beta'(\alpha) - 2\alpha \left[\left\| \frac{d\mathbf{x}_\alpha}{d\alpha} \right\|^2 + \left(\mathbf{x}_\alpha, \frac{d^2\mathbf{x}_\alpha}{d\alpha^2} \right) \right],$$

where $\beta(\alpha) = \|\mathbf{x}_\alpha\|^2$, $\beta'(\alpha) = 2 \left(\frac{d\mathbf{x}_\alpha}{d\alpha}, \mathbf{x}_\alpha \right)$, and \mathbf{x}_α , $d\mathbf{x}_\alpha/d\alpha$ and $d^2\mathbf{x}_\alpha/d\alpha^2$ can be obtained by solving the following equations:

$$(K^T K + \alpha D)\mathbf{x}_\alpha = K^T \mathbf{y}_n, \quad (23)$$

$$(K^T K + \alpha D) \frac{d\mathbf{x}_\alpha}{d\alpha} = -D\mathbf{x}_\alpha, \quad (24)$$

$$(K^T K + \alpha D) \frac{d^2\mathbf{x}_\alpha}{d\alpha^2} = -2D \frac{d\mathbf{x}_\alpha}{d\alpha}. \quad (25)$$

To solve the linear matrix-vector equations (23–25), we use the Cholesky (square root) decomposition method. A remarkable characteristic of the solution of (23–25) is that the Cholesky decomposition of the coefficient matrix $K^T K + \alpha D$ needs only once, then the three vectors \mathbf{x}_α , $d\mathbf{x}_\alpha/d\alpha$, $d^2\mathbf{x}_\alpha/d\alpha^2$ can be obtained cheaply.

In the case of perturbation of operators, the above method can be applied similarly. Note that in such case, the discrepancy equation becomes

$$\tilde{\Psi}(\alpha) = \|\tilde{K}\mathbf{x}_\alpha - \mathbf{y}_n\|^2 - (\delta + \tilde{\delta})^2 - (\mu_\eta^\kappa(\mathbf{y}_n, \tilde{K}))^2 = 0, \quad (26)$$

where $\tilde{\delta}$ is the error level of \tilde{K} approximating the true operator, $\eta = (\delta, \tilde{\delta})$ and $\mu_\eta(\mathbf{y}_n, \tilde{K})$ is the incompatibility measure of the equation $K\mathbf{x} = \mathbf{y}$ and $\kappa > 0$. Equation 26 is called a generalized discrepancy equation and is an one-dimensional nonlinear equation which can be solved by Newton's or cubic convergent method. For more information about generalized discrepancy, we refer to Wang (2007), Tikhonov et al. (1995) for details.

3.4 Direct Regularization

Instead of Tikhonov regularization, our goal in this section is to solve an equality constrained l_2 problem

$$\|\mathbf{x}\|_2 \longrightarrow \min, \quad \text{s.t. } \tilde{K}\mathbf{x} + \mathbf{n} = \mathbf{y}_n, \quad (27)$$

where $\tilde{K} \in \mathbb{R}^{M \times N}$ is a perturbation of K (i.e., if we regard K as an accurate operator, then \tilde{K} is an approximation to K which may contain error or noise), $\mathbf{x} \in \mathbb{R}^N$, \mathbf{n} , $\mathbf{y}_n \in \mathbb{R}^M$.

As is mentioned already, the ill-posedness is largely due to the small singular values of the linear operator. Let us denote the singular value decomposition of \tilde{K} as $\tilde{K} = U_{M \times N} \Sigma_{N \times N} V_{N \times N}^T = \sum_{i=1}^N \sigma_i u_i v_i^T$, where both $U = [u_i]$ and $V = [v_i]$ are orthonormal matrices, i.e., the products of U with its transpose and V with its transpose are both identity matrices; Σ is a diagonal matrix whose nonzero entries consist of the singular values of \tilde{K} . The traditional LSE solution \mathbf{x}_{lse} of the constrained optimization system (27) can be expressed by the singular values and singular vectors in the form

$$\mathbf{x}_{lse} = \sum_{i=1}^N \frac{1}{\sigma_i} \left(u_i^T \mathbf{y}_n \right) v_i. \quad (28)$$

If the rank of \tilde{K} is $p \leq \min\{M, N\}$, then the above solution form inevitably encounters numerical difficulties, since the denominator contains numerically infinitesimal values. Therefore,

to solve the problem by the SVD, we must impose *a priori* information. As we have noted, Tikhonov regularization solves a variation problem by incorporating *a priori* information into the solution. In this section, we consider another way of incorporating *a priori* information to the solution. The idea is quite simple: instead of filtering the small singular values by replacing the small singular values with small positive numbers, we just make a truncation of the summation, i.e., the terms containing small singular values are replaced by zeroes. In this way, we obtain a regularized solution of the least squares problem (27) of minimal norm

$$\mathbf{x}_{\text{lse}}^{\text{trunc}} = \sum_{i=1}^p \frac{1}{\sigma_i} \left(\mathbf{u}_i^T \mathbf{y}_n \right) \mathbf{v}_i \quad (29)$$

and $\min_{\mathbf{x}} \|\tilde{K}\mathbf{x} - \mathbf{y}_n\|_2^2 = \sum_{i=p+1, \dots} |\mathbf{u}_i^T \mathbf{y}_n|^2$. We wish to examine the truncated singular value decomposition more. Note that in practice, \tilde{K} may not be exactly rank deficient, but instead be numerically rank deficient, i.e., it has one or more small but nonzero singular values such that $p_\delta < \text{rank}(\tilde{K})$. Here, p_δ refers to the numerical δ -rank of a matrix, see, e.g., Wang et al. (2006b) for details. It is clear from Eq. 29 that the small singular values inevitably give rise to difficulties. The regularization technique for SVD means some of the small singular values are truncated when in computation, and is hence called the NTSVD. Now assume that K is corrupted by the error matrix B_δ . Then, we replace K by a matrix $K_{\tilde{p}}$ that is close to K and mathematically rank deficient. Our choice of $K_{\tilde{p}}$ is obtained by replacing the small nonzero singular values $\sigma_{\tilde{p}+1}, \sigma_{\tilde{p}+2}, \dots$ with exact zeros, i.e.,

$$K_{\tilde{p}} = \sum_{i=1}^{\tilde{p}} \sigma_i \mathbf{u}_i \mathbf{v}_i^T \quad (30)$$

where \tilde{p} is usually chosen as p_δ . We call (30) the NTSVD of K . Now, we use (30) as the linear kernel to compute the least squares solutions. Actually, we solve the problem $\min_{\mathbf{x}} \|K_{\tilde{p}}\mathbf{x} - \mathbf{y}_n\|_2$, and obtain the approximate solution $\mathbf{x}_{\text{lse}}^{\text{appr}}$ of the minimal-norm

$$\mathbf{x}_{\text{lse}}^{\text{appr}} = K_{\tilde{p}}^\dagger \mathbf{y}_n = \sum_{i=1}^{\tilde{p}} \frac{1}{\sigma_i} \left(\mathbf{u}_i^T \mathbf{y}_n \right) \mathbf{v}_i, \quad (31)$$

where $K_{\tilde{p}}^\dagger$ denotes the Moore–Penrose generalized inverse.

Let us explain in more details the NTSVD for the underdetermined linear system. In this case, the number of independent variables is more than the number of observations, i.e., $M < N$. Assume that the δ -rank of \tilde{K} is $\tilde{p} \leq \min\{M, N\}$. It is easy to augment \tilde{K} to be an $N \times N$ square matrix \tilde{K}_{aug} by padding zeros underneath its M nonzero rows. Similarly, we can augment the right-hand side vector \mathbf{y}_n with zeros. The singular decomposition of \tilde{K} can be rewritten as $\tilde{K}_{\text{aug}} = U\Sigma V^T$, where $U = [u_1 u_2 \dots u_N]_{N \times N}$, $V = [v_1 v_2 \dots v_N]_{N \times N}$ and $\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_{\tilde{p}}, 0, \dots, 0)$. From this decomposition, we find that there are $N - \tilde{p}$ theoretical zero singular values of the diagonal matrix Σ . These $N - \tilde{p}$ zero singular values will inevitably induce high numerical instability.

3.5 Statistical Regularization

Bayesian statistics provides a conceptually simple process for updating uncertainty in the light of evidence. Initial beliefs about some unknown quantity are represented by a prior distribution.

Information in the data is expressed by the likelihood function $L(\mathbf{x}|\mathbf{y})$. The *a priori* distribution $p(\mathbf{x})$ and the likelihood function are then combined to obtain the posterior distribution for the quantity of interest. The *a posteriori* distribution expresses our revised uncertainty in light of the data, in other words, an organized appraisal in the consideration of previous experience.

The role of Bayesian statistics is very similar to the role of regularization. Now, we establish the relationship between the Bayesian estimation and the regularization. A continuous random vector \mathbf{x} is said to have a Gaussian distribution if its joint probability distribution function has the form

$$p_{\mathbf{x}}(\mathbf{x}; \mu, C) = \frac{1}{\sqrt{(2\pi)^N \det(C)}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^T C^{-1}(\mathbf{x} - \mu)\right), \quad (32)$$

where \mathbf{x} , $\mu \in \mathbb{R}^N$, C is an n -by- n symmetric positive definite matrix, and $\det(\cdot)$ denotes the matrix determinant. The mean is given by $E(\mathbf{x}) = \mu$ and the covariance matrix is $\text{cov}(\mathbf{x}) = C$.

Suppose $\mathbf{y} = K\mathbf{x} + \mathbf{n}$ is a Gaussian distribution with mean $K\mathbf{x}$ and covariance C_n , where C_n is the noise covariance of the observation noise and model inaccuracy. Then by (32) we obtain

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^M \det(C_n)}} \exp\left(-\frac{1}{2}(\mathbf{y} - K\mathbf{x})^T C_n^{-1}(\mathbf{y} - K\mathbf{x})\right). \quad (33)$$

From (32), the prior probability distribution is given by $p(\mathbf{x}) = \frac{\exp(-\frac{1}{2}\mathbf{x}^T C_x^{-1}\mathbf{x})}{\sqrt{(2\pi)^N \det(C_x)}}$. By Bayesian statistical inference and the above two equations, we obtain an *a posteriori* log likelihood function

$$L(\mathbf{x}|\mathbf{y}) = \log p(\mathbf{x}|\mathbf{y}) = -\frac{1}{2}(\mathbf{y} - K\mathbf{x})^T C_n^{-1}(\mathbf{y} - K\mathbf{x}) - \frac{1}{2}\mathbf{x}^T C_x^{-1}\mathbf{x} + \zeta, \quad (34)$$

where ζ is constant with respect to \mathbf{x} . The maximum a posteriori estimation is obtained by maximizing (34) with respect to \mathbf{x} ,

$$\mathbf{x} = \left(K^T C_n^{-1} K + C_x^{-1}\right)^{-1} K^T C_n^{-1} \mathbf{y}. \quad (35)$$

The easiest way of choosing C_n and C_x is by letting $C_n = \sigma_n^2 I_M$, $C_x = \sigma_x^2 I_N$, and then (35) becomes

$$\mathbf{x} = \left(K^T K + \xi I_M\right)^{-1} K^T \mathbf{y}, \quad (36)$$

where $\xi = \sigma_n^2 / \sigma_x^2$, which is the noise-to-signal ratio. It is clear that the solution obtained by maximum *a posteriori* estimation has the same form as the solution of the Tikhonov regularization.

4 Optimization

4.1 Sparse/Nonsmooth Inversion in l_1 Space

It deserves attention that the ill-posedness is the intrinsic feature of the inverse problems. Unless some additional information/knowledge such as monotonicity, smoothness, boundedness or the error bound of the raw data are imposed, the difficulty is hardly to be solved. Generally speaking, the kernel-driven BRDF model is semiempirical, the retrieved parameters \mathbf{x} are mostly considered as a kind of weight function though it is a function of leaf area index (LAI),

Lambertian reflectance, sunlit crown reflectance, and viewing and solar angles. Therefore, \mathbf{x} is not necessarily positive. However, since it is a weight function, an appropriate arrangement of the components of \mathbf{x} can yield the same results. That is to say, \mathbf{x} can be “made” to be nonnegative. The problem remaining is to develop some proper methods to solve the “artificial” problem. Our new meaning to the solution \mathbf{x}^* is related to the l_1 norm problem

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1, \text{ s.t. } K\mathbf{x} = \mathbf{y}, \mathbf{x} \geq 0, \quad (37)$$

which automatically imposes *a priori* information by considering the solution in l_1 space. Because of the limitations of the observation system, one may readily see that the recovered land surface parameters are discrete and sparse. Therefore, if an inversion algorithm is not robust, the outliers far from the true solution may occur. In this situation, the priori constrained l_1 minimization may work better than the conventional regularization techniques. The model (37) can be reduced to a linear programming problem (see Yuan 2001; Ye 1997; Wang et al. 2005), hence linear programming methods can be used for solving the inverse problem.

The l_1 norm solution method is seeking for a feasible solution within the feasible set $S = \{\mathbf{x} : K\mathbf{x} = \mathbf{y}, \mathbf{x} \geq 0\}$. So it is actually searching for an interior point within the feasible set S , hence is called the interior point method. The dual standard form of (37) is in the form

$$\max \mathbf{y}^T \mathbf{g}, \text{ s.t. } \mathbf{s} = \mathbf{e} - K^T \mathbf{g} \geq 0, \quad (38)$$

where \mathbf{e} is a vector with all components equaling to 1. Therefore, the optimality conditions for $(\mathbf{x}, \mathbf{g}, \mathbf{s})$ to be a primal-dual solution triplet are that

$$K\mathbf{x} = \mathbf{y}, K^T \mathbf{g} + \mathbf{s} = \mathbf{e}, \tilde{S}\tilde{F}\mathbf{e} = 0, \mathbf{x} \geq 0, \mathbf{s} \geq 0, \quad (39)$$

where $\tilde{S} = \text{diag}(s_1, s_2, \dots, s_N)$, $\tilde{F} = \text{diag}(x_1, x_2, \dots, x_N)$ and s_i , x_i are components of vectors \mathbf{s} and \mathbf{x} , respectively. The notation $\text{diag}(\cdot)$ denotes the diagonal matrix whose only nonzero components are the main diagonal line.

The interior point method generates iterates $\{\mathbf{x}_k, \mathbf{g}_k, \mathbf{s}_k\}$ such that $\mathbf{x}_k > 0$ and $\mathbf{s}_k > 0$. As the iteration index k approaches infinity, the equality-constraint violations $\|\mathbf{y} - K\mathbf{x}\|$ and $\|K^T \mathbf{g}_k + \mathbf{s}_k - \mathbf{e}\|$ and the duality gap $\mathbf{x}_k^T \mathbf{s}_k$ are driven to zero, yielding a limiting point that solves the primal and dual linear problems. For the implementation procedures and examples about using the algorithm, please refer to Wang et al. (2007b, 2009d) for details.

A more general regularization model is recently proposed in Wang et al. (2009b), where the authors considered a regularization model in general form

$$\min J[\mathbf{f}] := \frac{1}{2} \|\mathcal{K}\mathbf{f} - \mathbf{h}_n\|_{l_p}^p + \frac{\nu}{2} \|L(\mathbf{f} - \mathbf{f}^0)\|_{l_q}^q, \quad (40)$$

where $p, q > 0$ which are specified by users, $\nu > 0$ is the regularization parameter, L is the scale operator, and \mathbf{f}^0 is an *a priori* solution of the original model. This formulation includes most of the developed methods. Particularly, for $p = 2$ and $q = 1$ or $q = 0$, the model represents nonsmooth and sparse regularization, which represents a quite important and hot topic in present, compressive sensing for decoding in signal processing. A regularizing active set method was proposed both for quadratic programming and non-convex problems, we refer Wang et al. (2009b) for details.

4.2 Optimization Methods for l_2 Minimization Model

4.2.1 Newton-Type Methods

The conventional Tikhonov regularization method is equivalent to constrained l_2 minimization problem

$$\min_{\mathbf{x}} \|\mathbf{x}\|_2, \quad \text{s.t. } K\mathbf{x} = \mathbf{y}. \quad (41)$$

This reduces to solve an unconstrained optimization problem

$$\mathbf{x} = \operatorname{argmin}_{\mathbf{x}} J^\alpha(\mathbf{x}), \quad J^\alpha(\mathbf{x}) = \frac{1}{2} \|K\mathbf{x} - \mathbf{y}\|_2^2 + \frac{\alpha}{2} \|\mathbf{x}\|_2^2. \quad (42)$$

The gradient and Hessian of $J^\alpha(\mathbf{x})$ are given by $\operatorname{grad}_{\mathbf{x}}[J^\alpha(\mathbf{x})] = (K^T K + \alpha I)^{-1} \mathbf{x} - K^T \mathbf{y}$ and $\operatorname{Hess}_{\mathbf{x}}[J^\alpha(\mathbf{x})] = K^T K + \alpha I$, respectively. Hence at the k -th iterative step, the gradient and Hessian of $J^\alpha(\mathbf{x}_k)$ can be expressed as $\operatorname{grad}_k[J^\alpha]$ and $\operatorname{Hess}_k[J^\alpha]$, which are evaluated by $\operatorname{grad}_{\mathbf{x}_k}[J^\alpha(\mathbf{x}_k)]$ and $\operatorname{Hess}_{\mathbf{x}_k}[J^\alpha(\mathbf{x}_k)]$ respectively.

Newton-type methods are based on Gauss–Newton method and its various variations. We only supply the algorithm for Gauss–Newton method in this subsection. The Gauss–Newton method is an extension of Newton method in one-dimensional space to higher dimensional space. The iteration formula reads as

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \tau_k (\operatorname{Hess}_k[J^\alpha])^{-1} \operatorname{grad}_k[J^\alpha], \quad (43)$$

where τ_k , a damping parameter, that can be solved by line search technique, is used to control the direction $(\operatorname{Hess}_k[J^\alpha])^{-1} \operatorname{grad}_k[J^\alpha]$. One may also apply a more popular technique, called the trust region technique, to control the direction $(\operatorname{Hess}_k[J^\alpha])^{-1} \operatorname{grad}_k[J^\alpha]$ within a reliable generalized ball in every iteration (see Wang and Yuan 2005; Wang 2007). We recall that the inverse of $\operatorname{Hess}_k[J^\alpha]$ should be avoided for saving the amount of computation. Instead, linear algebraic decomposition methods can be applied to solve $(\operatorname{Hess}_k[J^\alpha])^{-1} \operatorname{grad}_k[J^\alpha]$.

There are different variations of the Gauss–Newton method, which are based on the approximation of the explicit Hessian matrix $\operatorname{Hess}_{\mathbf{x}}[J^\alpha]$, e.g., DFP, BFGS, L-BFGS, and trust region methods. For these extensions to well-posed and ill-posed problems, please refer to Kelley (1999), Yuan (1993), Nocedal (1980), Dennis and Schnable (1983), Wang and Yuan (2005), Yuan (1994), and Wang (2007) for details. We mention briefly a global convergence method, the trust region method. The method solves an unconstrained non-quadratic minimization problem $\min_{\mathbf{x} \in \mathbb{R}^n} \Gamma(\mathbf{x})$. For the problem (42), the trust region method requires solving a trust region subproblem

$$\min_s Y(s) := (\operatorname{grad}_{\mathbf{x}}[J^\alpha], s) + \frac{1}{2} (\operatorname{Hess}_{\mathbf{x}}[J^\alpha] s, s), \quad \text{s.t. } \|s\| \leq \Delta, \quad (44)$$

where $\Delta > 0$ is the trust region radius. In each step, a trial step s is computed and decided whether it is acceptable or not. The decision rule is based on the ratio ρ between the actual reduction in the objective functional and the predicted reduction in the approximate model. And the trust region iterative step remains unchanged if $\rho \leq 0$, where $\rho = \frac{\operatorname{Ared}(\mathbf{x})}{\operatorname{Pred}(\mathbf{x})}$, and $\operatorname{Ared}(\mathbf{x})$ and $\operatorname{Pred}(\mathbf{x})$ are defined by $J^\alpha(\mathbf{x}) - J^\alpha(\mathbf{x} + s)$ and $Y(0) - Y(s)$, respectively. For the model in (42), since it is in a quadratic form, the ratio ρ is always equal to 1. This means the trial step s , no matter it is good or not, will be always accepted. We note that the approximate accuracy is characterized by the discrepancy between the observation and the true data; therefore variations of the norm

of the discrepancy may reflect the degree of approximation. Based on these considerations, we propose to accept or reject the trial step s_k at the k th step by the ratio

$$\rho_k = \frac{J^\alpha(\mathbf{x}_{k+1})}{J^\alpha(\mathbf{x}_k)} = \frac{J^\alpha(\mathbf{x}_k + \mathbf{s}_k)}{J^\alpha(\mathbf{x}_k)},$$

where $J^\alpha(\mathbf{x}_{k+1})$ and $J^\alpha(\mathbf{x}_k)$ are the reductions in norm of the discrepancy at $(k+1)$ -th and k -th steps, respectively. For the convergence and regularizing properties, we refer to Wang (2007) and Wang and Ma (2009) for details.

4.2.2 Gradient-Type Methods

The gradient method does not need the Hessian information. For the linear operator equation $K\mathbf{x} = \mathbf{y}$, where K , \mathbf{x} and \mathbf{y} are with the same meaning as before, we first recall the well-known fixed point iteration method in standard mathematical textbook: the fixed point iteration formula for solving the above linear operator equation is as

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \tau(\mathbf{y} - K\mathbf{x}_k), \quad (45)$$

where $\tau \in (0, 2/\|K\|)$ and K is linear, bounded and nonnegative. One may readily see that this method is very similar to the method of successive-approximations, where a very simple way to introduce the method is the following. Consider the operator $T(\mathbf{x}) = \mathbf{x} + \tau(\mathbf{y} - K\mathbf{x})$, where τ is the so-called relaxation parameter. Any solution of the linear operator equation is equivalent to finding a fixed point of the operator T , i.e., solve for \mathbf{x} from $\mathbf{x} = T(\mathbf{x})$. Assuming that T is a contraction mapping, then by the method of successive approximations, we obtain the following iterative scheme $\mathbf{x}_{k+1} = T(\mathbf{x}_k)$, i.e., iterative formula (45). The method converges if and only if $K\mathbf{x} = \mathbf{y}$ has a solution.

Now we introduce a very simple gradient method, the steepest descent method, the iteration formula reads as

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \tau_k \cdot K^T(\mathbf{y} - K\mathbf{x}_k), \quad (46)$$

where τ_k is obtained by line search, i.e., $\tau_k = \operatorname{argmin}_{\tau>0} J(\mathbf{x}_k - \tau \operatorname{grad}_k[J])$. If we restrict stepsize τ_k to be fixed in $(0, 2/\|K^T K\|)$, then the steepest descent method reduces to the famous Landweber–Fridman iteration method. More extensions include nonmonotone gradient method, truncated conjugate gradient method with trust region techniques and different applications in applied science and can be found in, e.g., Brakhage (1987), Wang and Yuan (2002), Wang (2007), Fletcher et al. (2001), Barzilai and Borwein (1988), and Wang and Ma (2007).

Finally we want to mention that for underdetermined ill-posed problems, regularization constraints or *a priori* constraints should be incorporated into the minimization model then we may apply the aforementioned gradient methods. Application examples on aerosol particle size distribution function retrieval problems and nonmonotone gradient method are include in Wang (2008).

5 Practical Applications

5.1 Kernel-Based BRDF Model Inversion

5.1.1 Inversion by NTSVD

Consider the linear combination of three kernels k_{geo} , k_{vol} and the isotropic kernel

$$\hat{f}_{\text{iso}} + \hat{f}_{\text{geo}} k_{\text{geo}}(t_i, t_v, \phi) + \hat{f}_{\text{vol}} k_{\text{vol}}(t_i, t_v, \phi) = \hat{r}$$

for each observation. Considering the smoothing technique in l_2 space, we solve the following constrained optimization problem

$$\min \|\hat{f}_{\text{iso}}, \hat{f}_{\text{geo}}, \hat{f}_{\text{vol}}\|^T, \text{ s.t. } \hat{f}_{\text{iso}} + \hat{f}_{\text{geo}} k_{\text{geo}} + \hat{f}_{\text{vol}} k_{\text{vol}} = \hat{r}. \quad (47)$$

Let us just consider an extreme example for kernel-based BRDF model: i.e., if only a single observation is available at one time, then it is clear that the above equation has infinitely many solutions. If we denote $K = [1 \quad k_{\text{geo}}(t_i, t_v, \phi) \quad k_{\text{vol}}(t_i, t_v, \phi)]_{1 \times 3}$, then the singular decomposition of the zero augmented matrix K_{aug} leads to $K_{\text{aug}} = U_{3 \times 3} \Sigma_{3 \times 3} V_{3 \times 3}^T$ with $U = [u_1 \ u_2 \ u_3]$, $\Sigma = \text{diag}(\sigma_1, \sigma_2, \sigma_3)$, $V = [v_1 \ v_2 \ v_3]$, where each u_i , v_i , $i = 1, 2, 3$, are the 3-by-1 columns. Our *a priori* information is based on searching for a minimal norm solution within the infinite set of solutions, i.e., the solution $f^* = [\hat{f}_{\text{iso}}^*, \hat{f}_{\text{geo}}^*, \hat{f}_{\text{vol}}^*]^T$ satisfies $\hat{f}_{\text{iso}}^* + \hat{f}_{\text{geo}}^* k_{\text{geo}}(t_i, t_v, \phi) + \hat{f}_{\text{vol}}^* k_{\text{vol}}(t_i, t_v, \phi) = \hat{r}$ and at the same time $\|f^*\| \rightarrow \text{minimum}$.

5.1.2 Tikhonov Regularized Solution

Denote by M the number of measurements in the kernel-based models. Then the operator equation (12) can be rewritten in the following matrix-vector form

$$K\mathbf{x} = \mathbf{y}, \quad (48)$$

$$\text{where } K = \begin{bmatrix} 1 & k_{\text{geo}}(1) & k_{\text{vol}}(1) \\ 1 & k_{\text{geo}}(2) & k_{\text{vol}}(2) \\ \vdots & \vdots & \vdots \\ 1 & k_{\text{geo}}(M) & k_{\text{vol}}(M) \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} f_{\text{iso}} \\ f_{\text{geo}} \\ f_{\text{vol}} \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_M \end{bmatrix}.$$

In which, $k_{\text{geo}}(k)$ and $k_{\text{vol}}(k)$ represent the values of kernel functions $k_{\text{geo}}(t_i, t_v, \phi)$ and $k_{\text{vol}}(t_i, t_v, \phi)$ corresponding to the k -th measurement for $k = 1, 2, \dots$; r_k represents the k -th observation for $k = 1, 2, \dots$

By Tikhonov Regularization, we solve for a regularized solution \mathbf{x}^α from minimizing the functional

$$J(\mathbf{x}, \alpha) = \frac{1}{2} \|K\mathbf{x} - \mathbf{y}\|^2 + \frac{1}{2} \alpha \|D\mathbf{x}\|^2. \quad (49)$$

Choices of the parameter α and the scale operator D are discussed in [Sect. 3.3](#).

5.1.3 Land Surface Parameter Retrieval Results

We use the combination of RossThick kernel and LiTransit kernel in the numerical tests. In practice, the coefficient matrix K cannot be determined accurately, and a perturbed version \tilde{K} is

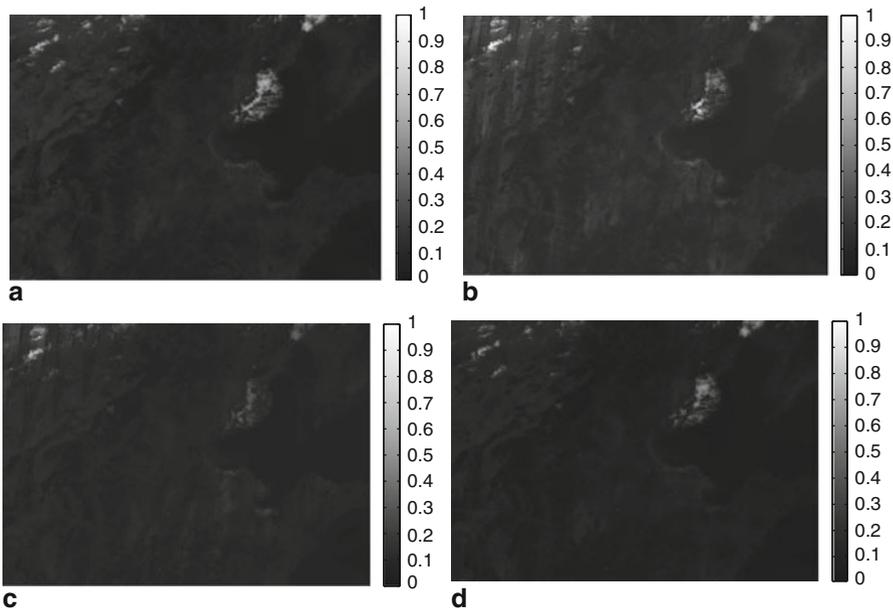
obtained instead. Also instead of the true measurement \mathbf{y} , the observed measurement $\mathbf{y}_n = \mathbf{y} + \mathbf{n}$ is the addition of the true measurement \mathbf{y} and the noise \mathbf{n} , which for simplicity is assumed to be additive Gaussian random noise. Therefore it suffices to solve the following operator equation with perturbation

$$\tilde{K}\mathbf{x} = \mathbf{y}_n,$$

where $\tilde{K} := K + \delta B$ for some perturbation matrix B and δ denotes the noise level (upper bound) of \mathbf{n} in $(0,1)$. In our numerical simulation, we assume that B is a Gaussian random matrix, and also that $\|\mathbf{y}_n - \mathbf{y}\| \leq \delta < \|\mathbf{y}_n\|$. The above assumption about the noise can be interpreted as that the signal-to-noise ratio (SNR) should be greater than 1. We make such an assumption as we believe that observations (BRDF) are not trustable otherwise. It is clear that (48) is an underdetermined system if $M \leq 2$ and an overdetermined system if $M > 3$. Note that for satellite remote sensing, because of the restrictions in view and illumination geometries, $\tilde{K}^T \tilde{K}$ needs not have bounded inverse (see Wang et al. 2007a; Wang et al. 2008; Verstraete et al. 1996; Li et al. 2001). We believe that the proposed regularization method can be employed to find an approximate solution $\mathbf{x}_\alpha^\dagger$ satisfies $\|\tilde{K}\mathbf{x}_\alpha^\dagger - \mathbf{y}_n\| \rightarrow \min$.

We use atmospherically corrected moderate resolution imaging spectroradiometer (MODIS) 1B product acquired on a single day as an example of single observation BRDF at certain viewing direction. Each pixel has different view zenith angle and relative azimuth angle. The data MOD021KM.A2001135-150 with horizontal tile number (26) and vertical tile number (4) were measured covers Shunyi county of Beijing, China. The three parameters are retrieved by using this 1B product.  Figure 2a plots the reflectance for band 1 of a certain day (DOY=137). In MODIS AMBRALS algorithm, when insufficient reflectances or a poorly representative sampling of high quality reflectances are available for a full inversion, a database of archetypal BRDF parameters is used to supplement the data and a magnitude inversion is performed (see Verstraete et al. 1996; Strahler et al. 1999). We note that the standard MODIS AMBRALS algorithm cannot work for such an extreme case, even for MODIS magnitude inversion since it is hard to obtain seasonal data associated with a dynamic land cover in a particular site. But our method still works for such an extreme case because that smoothness constraint is implanted into the model already. We plot the white-sky albedos (WSAs) retrieved by NTSVD, Tikhonov regularization and sparse inversion for band 1 of one observation (DOY=137) in  Fig. 2b-d, respectively. From  Fig. 2b-d, we see that the albedo retrieved from insufficient observations can generate the general profile. We observe that most of the details are preserved though the results are not perfect. The results are similar to the one from NTSVD method developed in Wang et al. (2007a). Hence, we conclude that these developed methods can be considered useful methods for retrieval of land surface parameters and for computing land surface albedos. Thus these developed algorithms can be considered as supplement algorithms for the robust estimation of the land surface BRDF/albedos.

We want to emphasize that our method can generate smoothing data for helping retrieval of parameters once sufficient observations are unavailable. As we have pointed out in Wang et al. (2007a, 2008), we do not suggest discarding the useful history information (e.g., data that is not too old) and the multiangular data. Instead, we should fully employ such information if it is available. The key to why our algorithm outperforms previous algorithms is because that our algorithm is adaptive, accurate and very stable, which solves kernel-based BRDF model of any order, which may be a supplement for BRDF/albedo retrieval product.



■ Fig. 2

(a) Reflectance for band 1 of MOD021KM.A2001137; (b) white-sky albedo retrieved by Tikhonov regularization method; (c) white-sky albedo retrieved by NTSVD method; and (d) WSA retrieved by l_1 sparse regularization method

For the remote sensor MODIS, which can generate a product by using 16 days different observations data, this is not a strict restriction for MODIS, since it aims at global exploration. For other sensors, the period for their detection of the same area will be longer than 20 days or more. Therefore, for vegetation in the growing season, the reflectance and albedos will change significantly. Hence robust algorithms to estimate BRDF and albedos in such cases are highly desired. Our algorithm is a proper choice, since it can generate retrieval results which quite approximate the true values of different vegetation type of land surfaces by capturing just one time of observation.

Moreover, for some sensors with high spatial resolution, the quasi multiangular data are impossible to obtain. This is why there are not high resolution albedo products. But with our algorithm, we can achieve the results. This is urgently needed in real applications.

5.2 Inversion of Airborne Lidar Remote Sensing

The analytical representation of the transmitted laser pulse and the true cross-sections given by third-order spline functions. For example, we generate the synthetic laser pulse function $f_{lp}(x)$ within the interval $[2, 3]$ by the formula $f_{lp}(x) = -31.25x^3 + 206.25x^2 - 356.25x + 218.75$. The analytical representation of the cross-section function $g_{cs}(x)$ within the interval $[3/8, 1/2]$ is by the formula $g_{cs}(x) = 8x^3 - 10x^2 + 3x + \frac{1}{6}$.

The recorded waveform function $h_{wf}(x)$ (i.e., data) is calculated by a convolution of the splines representing the transmitted laser pulse ($f_{lp}(x)$) and the cross-section ($g_{cs}(x)$) so that

$$h_{wf}(x) = f_{lp}(x) \star g_{cs}(x).$$

Note that h_{wf} represents the observation that means different kinds of noise may be also recorded besides the true signal. Here we only consider a simple case, i.e., we assume that the noise is mainly additive Gaussian noise in $[0, 1]$, i.e., $h_{wf} = h_{wf}^{\text{true}} + \delta \cdot \text{rand}(\text{size}(h_{wf}^{\text{true}}))$, where $\delta > 0$ is the noise level, $\text{rand}(\text{size}(h_{wf}^{\text{true}}))$ is the Gaussian random noise with the same size as h_{wf}^{true} . In our simulation, the Gaussian random noise is generated with mean equaling 0 and standard deviation equaling 2.

We apply Tikhonov regularization algorithm (see [▶ Sect. 3.3](#)) to recover the cross-section and make a comparison. The synthetic laser pulse sampled with 1ns resolution is shown in [▶ Fig. 3a](#). Comparisons of the undistorted cross-sections with the recovered cross-sections are illustrated in [▶ Fig. 3b](#). It is apparent that our algorithm can find stable recoveries to the simulated synthetic cross-sections. We do not list the plot of the comparison results for small noise levels since the algorithm yields perfect reconstructions. We also tested the applicability of the regularization method to LMS-Q560 data (RIEGL LMS-Q560 (www.riegl.co.at)). The emitted laser scanner sensor pulse is shown in [▶ Fig. 3c](#). The recorded waveform of the first echo of this pulse is shown in [▶ Fig. 3d](#) (dotted line). The retrieved backscatter cross-section using regularization method is shown in [▶ Fig. 4a](#). The solid line in [▶ Fig. 3d](#) shows the reconstructed signal derived by the convolution of the emitted laser pulse and this cross-section. One may see from [▶ Fig. 4a](#) that there are several small oscillations in the region [3850, 3860] ns. But note that the amplitude of these oscillations are typically small, we consider they are noise or computational errors induced by noise when performing numerical inversion. To show the necessity of regularization, we plot the result of least squares fitting without regularization in [▶ Fig. 4b](#). The comparison results immediately reveal the importance of acceptance of regularization. More extension about numerical performances and comparisons can be found in Wang et al. (2009a).

5.3 Particle Size Distribution Function Retrieval

We consider retrieving aerosol particle size distribution function $n(r)$ from the attenuation equation (7). But it is an infinite dimensional problem with only a finite set of observations, so it is improbable to implement such a system by computer to get a continuous expression of the size distribution $n(r)$. Numerically, we solve the discrete problem of operator equation (9). Using collocation (Wang et al. 2006), the infinite problem can be written in an finite dimensional form by sampling some grids $\{r_j\}_{j=1}^N$ in the interval of interests $[a, b]$.

Denoting by $K = (K_{ij})_{N \times N}$, \mathbf{n} , \mathbf{q} and \mathbf{d} the corresponding vectors, we have

$$K\mathbf{n} + \mathbf{q} = \mathbf{d}. \quad (50)$$

This discrete form can be used for computer simulations.

Phillips–Twomey’s regularization is based on solving the problem

$$\min_{\mathbf{n}} Q(\mathbf{n}), \text{ s.t. } \|K\mathbf{n} - \mathbf{d}\| = \Delta, \quad (51)$$

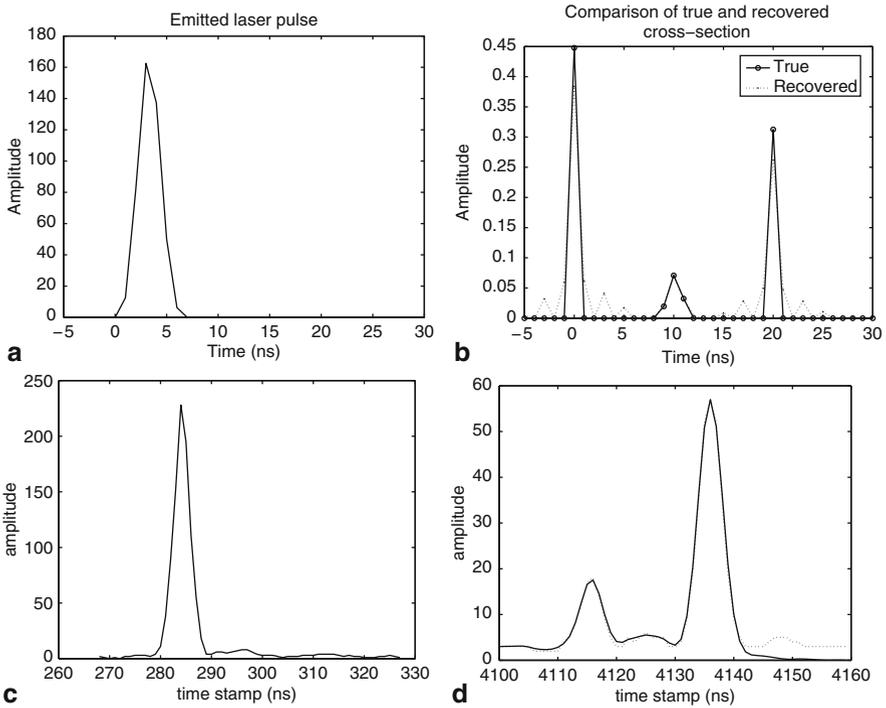


Fig. 3

(a) Synthetic emitted laser pulse; (b) comparison of the true and recovered cross sections in the case of noise of level 1; (c) second emitted laser pulse; and (d) recorded echo waveform of the laser pulse shown in (c) (dotted curve) and its reconstruction using the cross section shown in Fig. 4 (a) (solid curve)

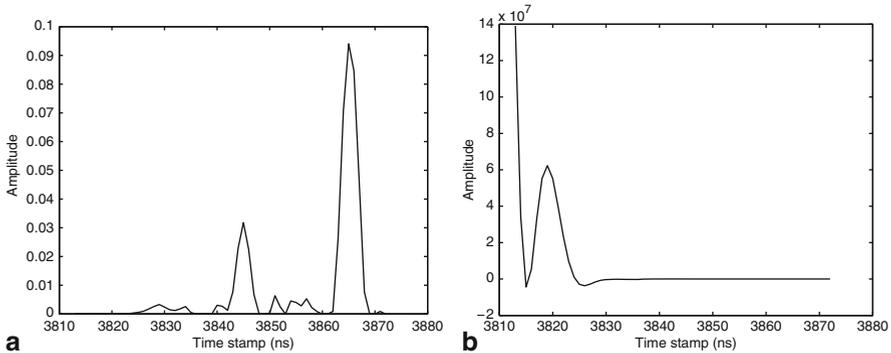


Fig. 4

(a) The retrieved backscatter cross-section using regularization; (b) the retrieved backscatter cross-section using least squares fitting without regularization

where $Q(\mathbf{n}) = (D\mathbf{n}, \mathbf{n})$, where D is a preassigned scale matrix.

In Phillips–Twomey’s formulation of regularization, the choice of the scale matrix is vital. They chose the form of the matrix D by the norm of the second differences, $\sum_{i=2}^{N-1} (n_{i-1} - 2n_i + n_{i+1})^2$, which corresponds to the form of matrix $D = D_2$. However, the matrix D is badly conditioned. For example, with $N = 200$, the largest singular value is 15.998012. The smallest singular value is 6.495571×10^{-17} . This indicates that the condition number of the matrix D defined by the ratio of the largest singular value to the smallest singular value equals 2.462911×10^{17} , which is worse. Hence, for small singular values of the discrete kernel matrix K , the scale matrix D cannot have them filtered even with large Lagrangian multiplier μ . This numerical difficulty encourages us to study a more robust scale matrix D , which is formulated as follows.

We consider the Tikhonov regularization in Sobolev $W^{1,2}$ space as is mentioned in **◆ Sect. 3.3.1**. By variational process, we solve a regularized linear system of equations

$$K^T K\mathbf{n} + \alpha H\mathbf{n} - K^T \mathbf{d} = 0, \quad (52)$$

where H is a triangular matrix in the form of D_1 . For choice of the regularization parameter, We consider the *a posteriori* approach mentioned in **◆ 3.3.2**. Suppose we are interested in the particle size in the interval $[0.1, 4] \mu\text{m}$, the step size is $h_r = \frac{3.9}{N-1}$. Now choosing the discrete nodes $N = 200$, the largest singular value of H is $1.041482176501067 \times 10^4$ by double machine precision, and the smallest singular value of H is 0.99999999999953 by double machine precision. Compared to the scale matrix D of Phillips–Twomey’s regularization, the condition number of H is $1.041482176501554 \times 10^4$, which is better than D in filtering small singular values of the discrete kernel K .

To perform the numerical computations, we apply the technique developed in King et al. (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. In this way we have

$$\tau_{\text{aero}}(\lambda) = \int_a^b [k(r, \lambda, \eta)h(r)]f(r)dr, \quad (53)$$

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

$$(\Xi f)(r) = \tau_{\text{aero}}(\lambda). \quad (54)$$

After obtaining the function $f(r)$, the size distribution function $n(r)$ can be obtained by multiplying $f(r)$ by $h(r)$.

The extinction efficiency factor (kernel function) $Q_{\text{ext}}(r, \lambda, \eta)$ is calculated from Mie theory: by Maxwell’s electromagnetic (E, H) theory, the spherical particle size scattering satisfies

$$\text{curl}H = i\kappa\eta^2 E, \quad \text{curl}E = -i\kappa H, \quad (55)$$

where $\kappa = 2\pi/\lambda$. The Mie solution process is one of finding a set of complex numbers a_n and b_n which give vectors E and H that satisfy the boundary conditions at the surface of the sphere (Bohren and Huffman 1983). Suppose the boundary conditions of the sphere is homogenous, the expressions for Mie scattering coefficients a_n and b_n are related by

$$a_n(z, \eta) = \frac{\eta\psi_n(\eta z)\psi'_n(z) - \psi_n(z)\psi'_n(\eta z)}{\eta\psi_n(\eta z)\xi'_n(z) - \psi'_n(\eta z)\xi_n(z)}, \quad (56)$$

$$b_n(z, \eta) = \frac{\psi_n(\eta z)\psi'_n(z) - \eta\psi_n(z)\psi'_n(\eta z)}{\psi_n(\eta z)\xi'_n(z) - \eta\psi'_n(\eta z)\xi_n(z)}, \quad (57)$$

where $\psi_n(z) = \sqrt{\frac{\pi z}{2}} J_{n+\frac{1}{2}}(z)$, $\xi_n(z) = \sqrt{\frac{\pi z}{2}} J_{n+\frac{1}{2}}(z) - i\sqrt{\frac{\pi z}{2}} N_{n+\frac{1}{2}}(z)$; $J_{n+\frac{1}{2}}(z)$ and $N_{n+\frac{1}{2}}(z)$ are the $(n + \frac{1}{2})$ -th order first kind Bessel function and second kind Bessel function (Neumann function), respectively. These complex-valued coefficients, functions of the refractive index η , $z = \frac{2\pi r}{\lambda}$ and ηz provide the full solution to the scattering problem. Thus the extinction efficiency factor (kernel function) can be written as

$$Q_{\text{ext}}(r, \lambda, \eta) = \frac{2}{z^2} \sum_{n=1}^{\infty} (2n+1) \text{Real}(a_n + b_n). \quad (58)$$

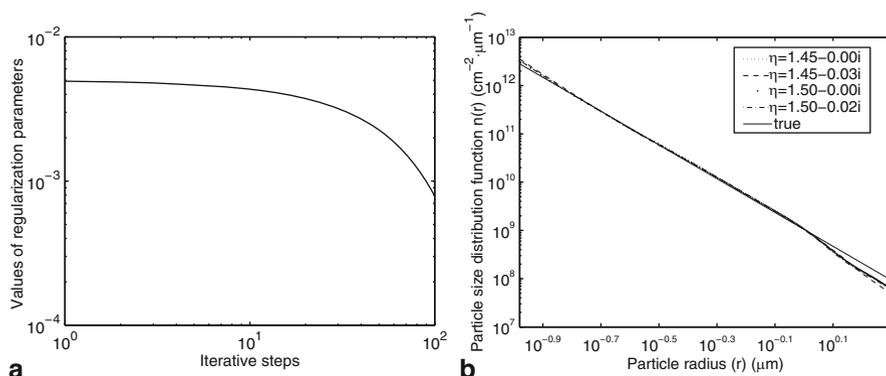
The size distribution function $n_{\text{true}}(r) = 10.5r^{-3.5} \exp(-10^{-12}r^{-2})$ is used to generate synthetic data. The particle size radius interval of interest is $[0.1, 2] \mu\text{m}$. This aerosol particle size distribution function can be written as $n_{\text{true}}(r) = h(r)f(r)$, where $h(r)$ is a rapidly varying function of r , while $f(r)$ is more slowly varying. Since most measurements of the continental aerosol particle size distribution reveal that these functions follow a Junge distribution (Junge 1955), $h(r) = r^{-(\nu^*+1)}$, where ν^* is a shaping constant with typical values in the range 2.0–4.0, therefore it is reasonable to use $h(r)$ of Junge type as the weighting factor to $f(r)$. 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 Twomey (1975), where a rapidly changing function $h(r) = Cr^{-3}$ can be identified, but it is more similar to a Junge distribution for $r \geq 0.1 \mu\text{m}$. One can also generate other particle number size distributions and compare the reconstruction with the input. In the first place, the complex refractive index η is assumed to be $1.45 - 0.00i$ and $1.50 - 0.00i$, respectively. 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. The precision of the approximation is characterized by the root mean-square error (rmse)

$$\text{rmse} = \sqrt{\frac{1}{m} \sum_{i=1}^m \frac{(\tau_{\text{comp}}(\lambda_i) - \tau_{\text{meas}}(\lambda_i))^2}{(\tau_{\text{comp}}(\lambda_i))^2}}, \quad (59)$$

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. Numerical illustrations are plotted in [► Fig. 5b](#) with noise level $\delta = 0.05$ for different refractive indices, respectively. The behavior of regularization parameter is plotted in [► Fig. 5a](#). The rmses for each case are shown in [► Table 1](#).

6 Conclusion

In this chapter, we study the regularization and optimization methods for solving the inverse problems in geoscience and quantitative remotely sensing. Three typical kernel-based problems are introduced, including computation of number of aerosol particle size distribution function, estimation of land surface biomass parameter and backscatter cross-section. These problems are formulated in functional space by introducing the operator equations of the first kind. The mathematical models and solution methods in l_1 and l_2 spaces are considered. The regularization strategies and optimization solution techniques are fully described. The equivalence between the Tikhonov regularization and Bayesian statistical inference for solving geoscience inverse problems is established. The general regularization model in l_p - l_q (for $p, q \geq 0$) spaces, which can be convex or non-convex, are introduced. Numerical simulations for these problems are performed and illustrated.



■ Fig. 5

Iterative computational values of regularization parameters when the error level $\delta = 0.05$ (a); input and retrieved results with our inversion method in the case of error level $\delta = 0.05$ and different complex refractive indices (b)

■ Table 1

The rmse for different noise levels

Noise levels	$\eta = 1.45 - 0.00i$	$\eta = 1.45 - 0.03i$	$\eta = 1.50 - 0.02i$
$\delta = 0.005$	1.6443×10^{-5}	1.2587×10^{-5}	2.2773×10^{-5}
$\delta = 0.01$	1.6493×10^{-5}	1.2720×10^{-5}	2.2847×10^{-5}
$\delta = 0.05$	1.6996×10^{-5}	1.3938×10^{-5}	2.3504×10^{-5}

Acknowledgments

This research is supported by National “973” Key Basic Research Developments Program of China under grant numbers 2007CB714400, National Natural Science Foundation of China (NSFC) under grant numbers 10871191, 40974075 and Knowledge Innovation Programs of Chinese Academy of Sciences KZCX2-YW-QN107.

References

- Ångström A (1929) On the atmospheric transmission of sun radiation and on dust in the air. *Geogr Ann* 11:156–166
- Barzilai J, Borwein J (1988) Two-point step size gradient methods. *IMA J Numer Anal* 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. *Appl Opt* 40:1329–1342
- Bockmann C, Kirsche A (2006) Iterative regularization method for lidar remote sensing. *Comput Phys Commun* 174:607–615
- Bohren GF, Huffman DR (1983) Absorption and scattering of light by small particles. Wiley, New York
- Brakhage H (1987) On ill-posed problems and the method of conjugate gradients. In: Engl HW, Groetsch CW (eds) Inverse and ill-posed problems, Academic, Boston, pp 165–175
- Camps-Valls G (2008) New machine-learning paradigm provides advantages for remote sensing. *SPIE Newsroom*, DOI: 10.1117/2.1200806.1100
- Davies CN (1974) Size distribution of atmospheric aerosol. *J Aerosol Sci* 5:293–300

- Dennis JE, Schnable RB (1983) Numerical methods for unconstrained optimization and nonlinear equations. Prentice Hall, Englewood Cliffs
- Fletcher R (2001) On the Barzilai–Borwein method. Numerical Analysis Report NA/207
- Houghton JT, Meira Filho LG, Callander BA, Harris N, Kattenberg A, Maskell K (1966) Climate change 1995. Published for the Intergovernmental Panel on Climate Change, Cambridge University Press
- Junge CE (1955) The size distribution and aging of natural aerosols as determined from electrical and optical data on the atmosphere. *J Meteor* 12:13–25
- Kelley CT (1999) Iterative methods for optimization. SIAM, Philadelphia
- King MD, Byrne DM, Herman BM, Reagan JA (1978) Aerosol size distributions obtained by inversion of spectral optical depth measurements. *J Aerosol Sci* 35:2153–2167
- Li X, Wang J, Hu B, Strahler AH (1998) On utilization of a priori knowledge in inversion of remote sensing models. *Sci China D* 41:580–585
- Li X, Wang J, Strahler AH (1999) Apparent reciprocal failure in BRDF of structured surfaces. *Prog Nat Sci* 9:747–750
- Li X, Gao F, Liu Q, Wang JD, Strahler AH (2000) Validation of a new GO kernel and inversion of land surface albedo by kernel-driven model (1). *J Remote Sens* 4:1–7
- Li X, Gao F, Wang J, Strahler AH (2001) A priori knowledge accumulation and its application to linear BRDF model inversion. *J Geophys Res* 106:11925–11935
- Mccartney GJ (1976) Optics of atmosphere. Wiley, New York
- Nguyen T, Cox K (1989) 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, Cincinnati, pp 330–330
- Nocedal J (1980) Updating quasi-Newton matrices with limited storage. *Math Comput* 95:339–353
- Phillips DL (1962) A technique for the numerical solution of certain integral equations of the first kind. *J Assoc Comput Mach* 9:84–97
- Pokrovsky O, Roujean JL (2002) Land surface albedo retrieval via kernel-based BRDF modeling: I. Statistical inversion method and model comparison. *Remote Sens Environ* 84:100–119
- Pokrovsky OM, Roujean JL (2003) Land surface albedo retrieval via kernel-based BRDF modeling: II. An optimal design scheme for the angular sampling. *Remote Sens. Environ.* 84: 120–142
- Pokrovsky IO, Pokrovsky OM, Roujean JL (2003) Development of an operational procedure to estimate surface albedo from the SEVIRI/MSG observing system by using POLDER BRDF measurements: II. Comparison of several inversion techniques and uncertainty in albedo estimates. *Remote Sens Environ* 87:215–242
- Privette JL, Eck TF, Deering DW (1997) Estimating spectral albedo and nadir reflectance through inversion of simple bidirectional reflectance distribution models with AVHRR/MODIS-like data. *J Geophys Res* 102:29529–29542
- Roujean JL, Leroy M, Deschamps PY (1992) A bidirectional reflectance model of the Earth's surface for the correction of remote sensing data. *J Geophys Res* 97:20455–20468
- Strahler AH, Li XW, Liang S, Muller J-P, Barnsley MJ, Lewis P (1994) MODIS BRDF/albedo product: Algorithm technical basis document. NASA EOS-MODIS Doc. 2.1
- Strahler AH, Lucht W, Schaaf CB, Tsang T, Gao F, Li X, Muller JP, Lewis P, Barnsley MJ (1999) MODIS BRDF/albedo product: Algorithm theoretical basis document. NASA EOS-MODIS Doc. 5.0
- Tikhonov AN, Arsenin VY (1977) Solutions of ill-posed problems. Wiley, New York
- Tikhonov AN, Goncharsky AV, Stepanov VV, Yagola AG (1995) Numerical methods for the solution of ill-posed problems. Kluwer, Dordrecht
- Twomey S (1975) 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
- Twomey S (1977) Atmospheric aerosols. Elsevier Sci. Publ. Company, Amsterdam
- Verstraete MM, Pinty B, Myneny RB (1996) Potential and limitations of information extraction on the terrestrial biosphere from satellite remote sensing. *Remote Sens Environ* 58:201–214
- Voutilainenand A, Kaipio JP (2000) Statistical inversion of aerosol size distribution data. *J Aerosol Sci* 31:767–768
- Wagner W, Ullrich A, Ducic V, Melzer T, Studnicka N (2006) Gaussian decomposition and calibration of a novel small-footprint full-waveform digitizing airborne laser scanner. *ISPRS J Photogram Remote Sens* 60:100–112
- Wang YF (2007) Computational methods for inverse problems and their applications. Higher Education Press, Beijing
- Wang YF (2008) An efficient gradient method for maximum entropy regularizing retrieval of atmospheric aerosol particle size distribution function. *J Aerosol Sci* 39:305–322

- Wang YF, Xiao TY (2001) Fast realization algorithms for determining regularization parameters in linear inverse problems. *Inverse Probl* 17: 281–291
- Wang YF, Yuan YX (2002) On the regularity of a trust region-CG algorithm for nonlinear ill-posed inverse problems. In: Sunada T, Sy PW, Yang L (eds) *Proceedings of the third Asian mathematical conference*, World Scientific, Singapore, pp 562–580
- Wang YF, Yuan YX (2003) A trust region algorithm for solving distributed parameter identification problem. *J Comput Math* 21:759–772
- Wang YF, Yuan YX (2005) Convergence and regularity of trust region methods for nonlinear ill-posed inverse problems. *Inverse Probl* 21:821–838
- Wang YF, Ma SQ (2007) Projected Barzilai-Borwein methods for large scale nonnegative image restorations. *Inverse Probl Sci Eng* 15: 559–583
- Wang YF, Ma SQ (2009) A fast subspace method for image deblurring. *Appl Math Comput* 215:2359–2377
- Wang YF, Yang CC (2008) A regularizing active set method for retrieval of atmospheric aerosol particle size distribution function. *J Opt Soc Am A* 25:348–356
- Wang YF, Li XW, Ma SQ, Yang H, Nashed Z, Guan YN (2005) BRDF model inversion of multiangular remote sensing: ill-posedness and interior point solution method. In: *Proceedings of the 9th international symposium on physical measurements and signature in remote sensing (ISPM-SRS) vol XXXVI*, pp 328–330
- Wang YF, Fan SF, Feng X, Yan GJ, Guan YN (2006a) Regularized inversion method for retrieval of aerosol particle size distribution function in $W^{1,2}$ space. *Appl Opt* 45:7456–7467
- Wang YF, Wen Z, Nashed Z, Sun Q (2006b) Direct fast method for time-limited signal reconstruction. *Appl Opt* 45:3111–3126
- Wang YF, Li XW, Nashed Z, Zhao F, Yang H, Guan YN, Zhang H (2007a) Regularized kernel-based BRDF model inversion method for ill-posed land surface parameter retrieval. *Remote Sens Environ* 111:36–50
- Wang YF, Fan SF, Feng X (2007b) Retrieval of the aerosol particle size distribution function by incorporating *a priori* information. *J Aerosol Sci* 38:885–901
- Wang YF, Yang CC, Li XW (2008) A regularizing kernel-based BRDF model inversion method for ill-posed land surface parameter retrieval using smoothness constraint. *J Geophys Res* 113:D13101
- Wang YF, Zhang JZ, Roncat A, Künzer C, Wagner W (2009a) Regularizing method for the determination of the backscatter cross-section in Lidar data. *J Opt Soc Am A* 26:1071–1079
- Wang YF, Cao JJ, Yuan YX, Yang CC, Xiu NH (2009b) Regularizing active set method for nonnegatively constrained ill-posed multichannel image restoration problem. *Appl Opt* 48:1389–1401
- Wang YF, Yang CC, Li XW (2009c) Kernel-based quantitative remote sensing inversion. In: Camps-Valls G, Bruzzone L (eds) *Kernel methods for remote sensing data analysis*. Wiley, New York
- Wang YF, Ma SQ, Yang H, Wang JD, Li XW (2009d) On the effective inversion by imposing *a priori* information for retrieval of land surface parameters. *Sci China D* 39:360–369
- Wanner W, Li X, Strahler AH (1995) On the derivation of kernels for kernel-driven models of bidirectional reflectance. *J Geophys Res* 100:21077–21090
- Xiao TY, Yu SG, Wang YF (2003) *Numerical methods for the solution of inverse problems*. Science Press, Beijing
- Ye YY (1997) *Interior point algorithms: theory and analysis*. Wiley, Chichester
- Yuan YX (1994) Nonlinear programming: trust region algorithms. In: Xiao ST, Wu F (eds) *Proceedings of Chinese SIAM annual meeting*, Tsinghua University Press, Beijing, pp 83–97
- Yuan YX (1993) *Numerical methods for nonlinear programming*. Shanghai Science and Technology Publication, Shanghai
- Yuan YX (2001) A scaled central path for linear programming. *J Comput Math* 19:35–40