An inverse problem in optical remote sensing of atmospheric aerosols

Ghislain R. Franssens

Belgian Institute for Space Aeronomy (BISA)
Ringlaan 3, B-1180 Brussels, BELGIUM
E-mail: ghislain.franssens@oma.be

ABSTRACT

A key problem in atmospheric aerosol research is to obtain the particle size density distribution from the optical extinction spectrum. This amounts to solving a Fredholm integral equation of the first kind, but with a very complicated kernel. The kernel represents the electromagnetic scattering function of a single particle (which for simplicity is usually taken to be a homogeneous sphere). This scattering kernel is known analytically, but in the form of an infinite series of complicated terms, involving spherical Bessel functions.

There are two approaches to solve this inverse problem. Numerically, by applying some discretization to the integral equation and then solving the resulting linear system. The other option is to substitute a simple approximation for the scattering kernel, in order to be able to perform the inversion analytically. Neither approach is really straightforward. The numerical route suffers from problems with ill-conditioned linear systems, while the analytical route is hampered by the fact that, the construction of a sufficiently accurate but still analytical tractable approximation, seems not to be easily derivable. Both problems are due to the complexity of the exact kernel. The form in which it is known makes it very hard to infer its mathematical properties. This prevents the construction of stable numerical methods as well as to find suitable approximations. However, despite the shortcomings of both routes, practical useful results have been obtained in several real-world applications.

The paper starts with the mathematical formulation of the problem. Then an example is given of a typical method for each of the above approaches: (i) Twomey's numerical method and (ii) the analytic inversion in the case of the Anomalous Diffraction Approximation (ADA) to the scattering kernel. Both methods are discussed and the synergy between both attempts is illustrated. In particular, the importance of the mathematical insight, yielded by the ADA model, with respect to the posed-ness of the continuous inverse problem and for constructing stable numerical methods, is emphasized. Finally, some comments are given to stimulate the search for a more complete mathematical understanding of this challenging problem.

1. INTRODUCTION

Atmospheric aerosols are microscopic dust particles or droplets suspended in the air. It is the prime and most noticeable constituent of air pollution. Aerosols have a variety of effects on the environment, such as reduced visibility, health hazards, affection of ozone concentrations, they are catalysts of chemical reactions and cause changes in the radiative fluxes and temperature distributions in the atmosphere, thus making them an important issue in the global warming of our planet.

A practical way to estimate on a global scale aerosol concentrations and sizes is by remote sensing the optical extinction of sunlight when it passes through a cloud of aerosols. When measured as a function of wavelength, one obtains an optical extinction spectrum that has a direct relationship with the particles size density distribution and their composition.

Deriving the aerosol size density distribution from the forward spectral extinction measurements requires the inversion of a Fredholm integral equation of the first kind. This equation gives the extinction spectrum as an integral over the number density distribution, with a kernel determined by scattering theory.

Mie and Debye independently developed an exact description of the scattering of light by spherical particles (Born and Wolf 1987, van de Hulst 1981). The mathematical expression they obtained for the scattered field has the form of an infinite series of complicated terms, containing spherical Bessel functions. This complexity is reflected in the scattering kernel, which occurs in the integral equation, and so prevents a direct analytic inversion of this equation.

There are two approaches to solve this inverse problem. Numerically, by applying some discretization to the integral equation and then solving the resulting linear system. The other option is to substitute a simple approximation for the scattering kernel, in order to be able to perform the inversion analytically. Neither approach is really straightforward. The numerical route suffers from problems with ill-conditioned linear systems, while the analytical route requires the construction of a sufficiently accurate but still analytical tractable approximation. Both approaches are hindered by the complexity of the exact kernel. However, despite the shortcomings of both routes, practical results have been obtained in several real-world applications (e.g. Franssens 2000b,d).

An interesting existing approximation is the Anomalous Diffraction Approximation (ADA), introduced by van de Hulst, (van de Hulst 1981). It is valid for particles which are large as compared to the wavelength and which have a small refractive index contrast relative to their surroundings. Under these circumstances, optical ray tracing can be used to describe the interaction of light with the particle and reflection and refraction effects can be neglected. The presence of a particle then only produces a change in the complex phase front of an incident monochromatic wave over its geometrical shadow area. As a result of these simplifications a much simpler scattering kernel is obtained, which makes it possible to solve the integral equation analytically (Franssens 2000a). The inverse problem in the ADA can be given an elegant mathematical formulation and at the same time has practical value for retrieving the larger particles. Unfortunately, the ADA is not sufficiently accurate to be used as a substitute for Mie theory in general, but it is an instructive model to gain insight in the physical properties of the problem.

The paper starts with the mathematical formulation of the problem. Then an example is given of a typical method for each of the above approaches: (i) Twomey's numerical method and (ii) the analytic inversion in the case of the Anomalous Diffraction Approximation (ADA) to the scattering kernel. Both methods are discussed and the synergy between both attempts is illustrated. In particular, the importance of the mathematical insight, yielded by the ADA model, with respect to the posed-ness of the continuous inverse problem and for constructing stable numerical methods, is emphasized. Finally, some comments are given to motivate the search for a more complete mathematical understanding of this challenging problem.

2. THE INVERSE PROBLEM

Let $k \triangleq 2\pi/\lambda$, with λ the wavelength of the light, and r the particle radius. The extinction spectrum $\tau(k)$ [1/m], caused by a cloud of spherical particles with number density distribution $N(r)$ [1/(μm cm^3)], is given by (van de Hulst 1981)

$$\tau(k) = \int_0^{+\infty} C(k,r) N(r) dr, \tag{1}$$

where $C(k,r)$ [μm^2] is the scattering cross-section kernel determined by the scattering model. The exact scattering cross-section kernel, for a spherical particle with complex refractive index $n(k)$ immersed in air, was derived by Mie and is given by the infinite series (van de Hulst 1981, Born and Wolf 1987)

$$C(k,r) = \pi r^2 \frac{2}{x^2} \Re\left[\sum_{l=1}^{+\infty} (2l+1)\left(a_l(x,y) + b_l(x,y)\right) \right], \tag{2}$$

with terms

$$a_l(x,y) \triangleq \frac{\dfrac{d\psi_l(x)}{dx} y\psi_l(y) - \psi_l(x) x \dfrac{d\psi_l(y)}{dy}}{\dfrac{d\zeta_l(x)}{dx} y\psi_l(y) - \zeta_l(x) x \dfrac{d\psi_l(y)}{dy}},$$

$$b_l(x,y) \triangleq \frac{\dfrac{d\psi_l(x)}{dx} x\psi_l(y) - \psi_l(x) y \dfrac{d\psi_l(y)}{dy}}{\dfrac{d\zeta_l(x)}{dx} x\psi_l(y) - \zeta_l(x) y \dfrac{d\psi_l(y)}{dy}},$$

containing the modified spherical Bessel and Hankel functions

$$\psi_l(z) \triangleq z j_l(z) = \sqrt{\frac{\pi z}{2}} J_{l+1/2}(z), \quad \chi_l(z) \triangleq z y_l(z) = \sqrt{\frac{\pi z}{2}} Y_{l+1/2}(z),$$

$$\zeta_l(z) \triangleq z h_l^{(2)}(z) = \sqrt{\frac{\pi z}{2}} H_{l+1/2}^{(2)}(z) = \psi(z) - i\chi(z) \,,$$

and where $x \triangleq kr \in \mathbb{R}$, $y \triangleq n(k)x \in \mathbb{C}$.

The inverse problem consists in solving the Fredholm equation of the first kind (1) for $N(r)$, given $\tau(k)$.

3. ANALYTIC INVERSION USING AN APPROXIMATE KERNEL

On physical grounds one can derive a, crude but sometimes practical, approximation to the exact Mie kernel (2), called the Anomalous Diffraction Approximation (ADA) kernel (van de Hulst 1981). The ADA scattering cross-section kernel has the much simpler expression

$$C(k,r) = \pi r^2 \Re\left[2 - 4i \frac{e^{-i\kappa(k)r}}{\kappa(k)r} + 4 \frac{1 - e^{-i\kappa(k)r}}{(\kappa(k)r)^2} \right], \tag{3}$$

where $\kappa(k)r \triangleq 2(y-x) = 2\big(n(k)-1\big)kr$. It describes both scattering and absorption of light by spherical particles with radius r and complex refractive index $n = n_r - in_i$, under the conditions that $|n-1| \ll 1 \ll kr$.

Figs. 1a,b show a plot of both kernels and their difference for transparent ($n(k) \in \mathbb{R}$) and absorbing ($n(k) \in \mathbb{C}$) particles respectively, tabulated over wavelength and for a fixed particle radius $r = 1$ µm. A constant refractive index was used over the whole wavelength region, in order to better visualize the differences in scattering properties between both models, without them being obscured by the absorption peaks present in a more physical refractive index spectrum $n(k)$.

Title:
Graphics produced by IDL
Creator:
IDL Version 5.2 (hp-ux hp_pa)
Preview:
This EPS picture was not saved
with a preview included in it.
Comment:
This EPS picture will print to a
PostScript printer, but not to
other types of printers.

Title:
Graphics produced by IDL
Creator:
IDL Version 5.2 (hp-ux hp_pa)
Preview:
This EPS picture was not saved
with a preview included in it.
Comment:
This EPS picture will print to a
PostScript printer, but not to
other types of printers.

Fig. 1. Comparison ADA-MIE kernel for transparent water drops, (r = 1.0 µm, n = 1.33).

Fig. 2. Comparison ADA-MIE kernel for absorbing water drops, (r = 1.0 µm, n = 1.33 + i0.1).

From Figs. 1 and 2 one sees that the ADA contains the correct geometrical limit (value 2, for $\lambda \to 0$). In addition, it also reproduces with good accuracy the positions of the extrema of the MIE kernel. It systematically under estimates the values at these extrema, amounting to a maximum error of about 20% at the largest peak. Another serious defect of the ADA is that it does not contain the Rayleigh limit of the extinction curve at large wavelengths (i.e. $\tau(k) \sim k^4$), but instead falls off with the square of the wavelength ($\tau(k) \sim k^2$). The fine structure, present in the MIE kernel is obviously missing in the ADA. However, this fine structure is never transferred into the extinction spectrum, because it is averaged out by the integration with any smooth number density distribution in (1).

A more detailed numerical comparison between the ADA and MIE kernels can be found in (van de Hulst 1981, Kerker 1969, Sharma 1992).

Because the ADA correctly contains the geometrical limit, an inversion based on the ADA will correctly reproduce the particles total surface area density A_{tot} (because $\lim_{k \to +\infty} \tau(k) = A_{tot}/2$). Also, the good approximation of the position of the overall maximum in the extinction curve by the ADA will result in a reasonable good estimate of the main particle size, at least for not too wide distributions $N(r)$.

Under the ADA the inverse problem can be solved analytically. To this end we need to extend the kernel $C(k,r)$ in (3) to a complex analytic function of κ and this produces a complex extinction spectrum $E(k)$. The complex equivalent of relation (1), relating $N(r)$ and $\tau(k)$ with kernel (3), can now be written as (Franssens 2000a):

$$F(k) = \int_0^{+\infty} K(k,r)P(r)dr. \tag{4}$$

Herein we defined:

$$F(\kappa) \triangleq \kappa E(\kappa), \tag{5a}$$

$$E(\kappa) \triangleq a(\kappa) - i(b(\kappa) - A_{tot}/4), \tag{5b}$$

$$K(\kappa,r) \triangleq -e^{-i\kappa r} + \frac{1-e^{-i\kappa r}}{i\kappa r}, \tag{5c}$$

$$P(r) \triangleq 2\pi r N(r), \tag{5d}$$

$$A_{tot} \triangleq \int_0^{+\infty} 4\pi r^2 N(r)dr. \tag{5e}$$

The real (measurable) extinction spectrum $\tau(k)$ is related to the complex extinction function $E(k)$ as $t(k) = 2b(k(k))$. The spectral function $a(\kappa(k))$ gives the phase advance of a plane wave when passing through a cloud of particles under incoherent scattering and is related to $b(\kappa(k)) - A_{tot}/4$ by causality (see Franssens 2000a). Finally, A_{tot} represents the (total) surface area density. The relation (4) is called the forward complex ADA transform. It relates the modified complex extinction function $F(k)$ with the perimeter density function $P(r)$, which turn out to be the canonical variables of the problem in the ADA.

The complex inverse of (4) is (Franssens 2000a)

$$P(r) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} K(-\kappa, r) F(\kappa) d\kappa, \; r \geq 0. \tag{6}$$

The inverse complex ADA transform (6) contains the complex extinction function of which only the part $b(\kappa(k))$ is measured. So to be of any practical use, one has to eliminate $a(\kappa(k))$ from $F(\kappa)$ in (6). This is possible because light propagation is a causal process. Mathematically, this is equivalent to the fact that $E(\kappa)$ is complex analytic over $\Im(\kappa) \leq 0$. On the real κ axis, the real and imaginary parts of $E(\kappa)$ therefore form a Hilbert transform pair. This argument allows us to eliminate $a(\kappa)$ from $F(\kappa)$ in (6) and this yields the following equivalent real form of (6). It now contains only the measured real spectrum $t(k)$ and its small wavelength limit $\tau_\infty \triangleq \lim_{k \to +\infty} \tau(k) = A_{tot}/2$, (Franssens 2000a):

$$N(r) = \frac{1}{2\pi r} \frac{1}{2\pi} \int_{0}^{+\infty} \Im\left[K\left(-\kappa(k), r\right) \frac{d\kappa^2(k)}{dk} \right] \left(\tau(k) - \tau_\infty\right) dk. \tag{7}$$

A direct numerical integration of (7) is only feasible when $t(k)$ is sufficiently dense measured over a sufficiently broad wavelength interval and when the measurement errors are small.

In (Franssens 2000a) it is shown that the forward and inverse ADA transforms (4) and (6) are well-posed. The real valued inverse formula (7) was found to be well-posed, provided the correct physical limit behavior of the refractive index spectrum $n(k)$ is used.

4. NUMERICAL SOLUTION BY TWOMEY'S KERNEL COVARIANCE METHOD

Twomey's kernel covariance method (Twomey 1967) is a numerical method to retrieve the size distribution from the extinction spectrum, which is widely used in atmospheric aerosol research. It is not to be confused with Phillips' method with added regularization (Phillips 1962), an approach that was also addressed by Twomey, a few years earlier (Twomey 1963).

Twomey's kernel covariance method can be summarized as follows. Recall the forward transform (1), but now evaluated only at M measured wavelengths:

$$\int_0^R Q(k_m,r)\pi r^2 N(r)dr = \tau_m, \quad m=1,M \,, \qquad Q(k,r) \triangleq \frac{C(k,r)}{\pi r^2} \tag{8}$$

and for a truncated interval of integration. We replaced the scattering cross-section kernel $C(k,r)$ in terms of the scattering efficiency kernel $Q(k,r)$.

Twomey represents the unknown $N(r)$ as a linear combination of basis functions, formed by the spectrally fixed efficiency kernels $Q(k_n,r)$, as

$$\pi r^2 N(r) = \sum_{n=1}^{M} c_n Q(k_n,r) \,. \tag{9}$$

Substituting (9) in (8) leads to the following linear system for the expansion coefficients c_n:

$$\sum_{n=1}^{M}\left(\int_0^R Q(k_m,r)Q(k_n,r)dr\right)c_n = \tau_m, \quad m=1,M \,. \tag{10}$$

The system matrix is just the (efficiency) kernel covariance matrix, with elements

$$A_{mn} \triangleq \int_0^R Q(k_m,r)Q(k_n,r)dr \,. \tag{11}$$

There are two problems with this method, which are inherent to the way the integral equation is discretized and which are independent of the scattering model used or the extinction data fed into.

(1) The matrix elements diverge when the integration limit R tends to ∞:

$$\lim_{R\to+\infty}\int_0^R Q(k_m,r)Q(k_n,r)dr = \infty \,. \tag{12}$$

This is due to the fact that the scattering efficiency kernel $Q(k,r)$ tends to the geometrical limit for $R \to +\infty$ (i.e. $\lim_{r\to+\infty} Q(k,r) = 2, \; 0 \le k < +\infty$). It could be avoided by introducing an additional weighting function to make the integral converge or, perhaps more meaningful, by extracting the geometrical limit from the efficiency kernel.

(2) The condition number of the linear system tends to infinity when the number of measurement points M increases:

$$\lim_{M\to\infty} cond\left[\int_0^R Q(k_m,r)Q(k_n,r)dr\right] \to \infty \,. \tag{13}$$

This is caused by the fact that the basis is not ortho-normal. When the separation of the measurement wavelengths tends to zero, adjacent columns tend to become linear dependent. Consequently, Twomey's method is a bad discretization of the continuous integral equation, because the obtained discrete solution fails to converge to the continuous solution with increasing measurement resolution.

Surprisingly however, although the method is conceptually wrong, it can be made to numerically work for certain combinations of the number of measurements M and upper limit of integration R . It appears that for these combinations both problems tend to cancel each other out. This fact has not been noticed in the literature, until recently (Franssens 2000c).

In practice, one usually resorts to regularization methods to make the method work for a given M and often arbitrarily chosen R . Regularization however should not become a standard practice, but a means of last resort to extract some 'inspired' solution out of a bad quality data set. It cannot be justified to use regularization, as a means of repair for a diverging solution method.

The problems associated with Twomey's method can easily be avoided, at least in the ADA, as the following reformulation shows. For clarity, only the case of transparent particles is considered.

Starting over again, but now with the forward ADA transform in its new formulation as given in (Franssens 2000a), we get the equations

$$\int_0^{+\infty} K_i(\kappa_m, r) P(r) dr = \kappa_m(\tau_m - \tau_\infty)/2, \quad m = 1, M , \tag{14}$$

and we used the transparent ADA kernel

$$K_i(\kappa, r) \triangleq -\sin(\kappa r) + \frac{1-\cos(\kappa r)}{\kappa r} . \tag{15}$$

Represent the unknown $P(r)$ as the linear combination:

$$P(r) = \frac{2}{\pi} \sum_{n=1}^{M} c_n K_i(\kappa_n, r) . \tag{16}$$

Substituting (16) in (14) gives the linear system for the expansion coefficients

$$\sum_{n=1}^{M} \underbrace{\left(\frac{2}{\pi} \int_0^{+\infty} K_i(\kappa_n, r) K_i(\kappa_m, r) dr \right)}_{=\delta(\kappa_n - \kappa_m)} c_n = \kappa_m(\tau_m - \tau_\infty)/2, \quad m = 1, M . \tag{17}$$

Because of the ortho-normality of the basis $\{K_i(\kappa_n, r), \; n = 1, M\}$ (Franssens 2000a), the kernel covariance matrix is now proportional to the unit matrix and we can explicitly solve the linear system, yielding

$$c_m = \kappa_m(\tau_m - \tau_\infty)\Delta\kappa_m / 2 \,. \tag{18}$$

Back substitution of (18) in (16) gives the explicit expression for $P(r)$ and hence,

$$N(r) = \frac{1}{2\pi r}\frac{1}{\pi}\sum_{m=1}^{M} K_i(\kappa_m, r)\kappa_m(\tau_m - \tau_\infty)\Delta\kappa_m \,. \tag{19}$$

This is just a staircase approximation to the inverse ADA transform integral (7) in the case of transparent particles (Franssens 2000a).

This shows that, Twomey's kernel covariance method, when implemented properly, is equivalent to a staircase discretization of the inverse solution. This result therefore shows that, when used with an ortho-normal basis, Twomey's method becomes a good discretization of the continuous integral equation, because the staircase approximation (19) indeed converges to the inverse integral when the spacing between the wavelengths tends to zero.

5. REMARKS

The considered inverse problem originated with the theoretical work of Mie and Debye on light scattering by spheres and thus dates back to 1908 (Born and Wolf 1987). Almost a century later, one is still in need for a satisfactory solution method. Attempts to formulate a stable numerical method or attempts to solve it (approximately) by analytical means are hindered by the complexity of the exact scattering kernel.

The solution methods discussed above, reveal a synergy between the analytic and numerical approach. On one hand, mathematical insight proves helpful to formulate a stable and converging numerical solution method. On the other hand, having an explicit analytic inversion formula can turn out to be pretty useless in case the practical data set to be fed into is scarce (spectrum measured with poor resolution or over a too limited spectral interval) and/or of bad quality (large measurement errors). No matter in what way the problem is attacked, both approaches would benefit from a better understanding of the mathematical properties of the Mie scattering kernel.

It is unlikely that an analytical route can be found to solve this problem exactly. An approximate solution, based on an approximate kernel (as the ADA kernel), seems to be the way to go. The ADA is an interesting simplification, but is too crude to be generally applicable. The quest is therefore still open for a less approximate kernel than the ADA kernel, but which still allows analytic inversion of the Fredholm integral equation. This approximation should at least contain the correct geometrical and Rayleigh spectral limits to be acceptable. The insight gained from such a model would no doubt be also helpful to formulate practical inversion algorithms for low quality spectral data.

The complex ADA yielded valuable mathematical insight into the spectral inversion problem. It resulted in a new formulation of the forward and inverse problem that is well-posed. The posed-ness of the inverse problem (1) with the exact kernel (2) is still unclear. The ADA example showed that a mathematical (complex analytic) reformulation was necessary to obtain a well-posed problem.

Twomey's discretization of the integral equation (1) is an example of a numerical method, which convergence depends on a proper formulation of the problem in terms of an ortho-normal kernel. At present, it is not clear how the exact Mie kernel (2) could be turned into an ortho-normal equivalent (as could be done in the ADA).

The author hopes that the above examples may have created an interest in advancing the mathematical understanding of this challenging problem and he remains available for further discussions.

REFERENCES

Born M. and Wolf E. (1987), Principles of Optics, Pergamon Press, Oxford, UK.

Franssens G., De Mazière M. and Fonteyn D. (2000a), Determination of the aerosol size distribution by analytic inversion of the extinction spectrum in the complex anomalous diffraction approximation, *Applied Optics*, 39, pp. 4214-4236.

Franssens G. (2000b), A global climatology of stratospheric aerosol properties from spectral inversion of SAGE II extinction measurements: 1985-1995, Proceedings of the International Radiation Symposium 2000, Saint Petersburg, Russia. To appear.

Franssens G. (2000c), Retrieval of the aerosol size distribution in the complex Anomalous Diffraction Approximation, Atmospheric Environment, Sep., 2000. Submitted.

Franssens G. (2000d), A new method for aerosol size distribution retrieval based on the Anomalous Diffraction Approximation, Proceedings of the EOS/SPIE Symposium on Remote Sensing, Barcelona, Spain, 25-29 Sep., 2000. To appear.

Kerker M. (1969), The scattering of light and other electromagnetic radiation, Academic Press, New York.

Phillips D. L. (1962), A technique for the numerical solution of certain integral equations of the first kind, *Journal for Assoc. Computing Machines*, 9, pp. 84-97.

Sharma S. (1992), On the validity of the anomalous diffraction approximation, *Journal of Modern Optics*, 39, pp. 2355-2361.

Twomey S. (1963), On the numerical solution of Fredholm integral equations of the first kind by the inversion of the linear system produced by quadrature, *Journal for Assoc. Computing Machines*, 10, pp. 97-101.

Twomey S. and Howell H. B. (1967), Some aspects of the optical estimation of microstructure in fog and cloud, *Applied Optics*, 6, pp. 2125-2131.

van de Hulst H. C. (1981), Light scattering by small particles, Dover Publications, New York.