AT721 Section 10:

Introduction to Inverse Radiation Problems:

Inspection of the canonical solution of the radiative transfer equation (xx,yy,zz) reveals an integral equation of the form

$$y(z) = \alpha(z)x(\chi, z) + \int_{a(z)}^{b(z)} K(\chi, z, z')x(\chi, z')dz' \qquad (10.1)$$

where $y(z)$ are the 'data' or measurements (typically radiances I_λ), $K(\chi, z, z')$ is the kernel of the solution equation and $\alpha(z)$ is some known function (like the transmittance from the surface to atmosphere). $x(\chi, z')$ is the source function and z, z' might be thought of as coordinates. χ may be considered either a parameter that influences the values of both *K* and the source vector $x(\chi, z')$ (such as the single scatter albedo, asymmetry parameter or other properties that are not retreived) or a vector quantity representing desired information to be retrieved.

The mathematics of the inverse radiation problem can be classified according to:

- $\alpha(z) = 0$ - (10.1) is an integral equation of the first kind
- $\alpha(z) = 1$ (10.1) is an integral equation of the second kind
- *a,b* are constants, (10.1) is of a *Fredholm* type (first or second depending on $\alpha(z)$)
- *a*=constant, *b*=*z,* (10.1) is of a *Volterra* type.

Inverse problems can be further classified depending on the focus of attention. If *x(χ, z')* is the desired information, such as it is for temperature retrieval problems, then the inversion problem is linear. If, however, the vector χ is desired, as it is in the example of constituent retrievals where χ is related to the concentration of the attenuating species, then the problem is non linear. We will be concerned with inversions of (10.1) of both linear and non-linear types of problems.

10.1The nature of the inverse problem

Certain characteristics of both linear and non-linear inverse problems can be highlighted with th following example. We start with the linear form of (10.1)

$$y(z) = \int_a^b K(z, z')x(z)dz \qquad (10.2a)$$

and we can always introduce some type of quadrature to discretize this in the form

$$\mathbf{y} \approx \mathbf{K}\mathbf{x} \qquad (10.2b)$$

where **y** is a column vector of n measurements, **x** is a column vector of n source functions and **K** is an nxn matrix. (10.2b) is now in the form of a linear discrete inverse problem in contrast to its the continuous inverse problem form of (10.2a). The solution of this discrete problem then takes the form

$$\mathbf{x} = \mathbf{K}^{-1}\mathbf{y}$$

In this way, the linear inverse problem apparently reduces to a straight forward matrix inversion but the problem is more complicated than this. We can begin to appreciate some of then issues with the following problem Suppose we make two measurements

$$\boldsymbol{y} = \begin{pmatrix} 2.0 \\ 4.0001 \end{pmatrix}$$

and determine that for our problem the Kernel has the values

$$\mathbf{K} = \begin{pmatrix} 1.0 & 1.0 \\ 2.0 & 2.0001 \end{pmatrix}$$

then we can determine that the solution to the pair of equations implied in (10.2) is

$$\mathbf{x} = \begin{pmatrix} 1.0 \\ 1.0 \end{pmatrix}$$

However, consider the situation where a small error is introduced to one of the measurements

$$\boldsymbol{y} + \varepsilon = \begin{pmatrix} 2.0 \\ 4.0 \end{pmatrix}$$

then the solution now becomes

$$\mathbf{x} = \begin{pmatrix} 2.0 \\ 0.0 \end{pmatrix}$$

Therefore a small error in the data leads to a substantial change in the solution. Readers without practical experience might be left with the impression that there is no fundamental problem here because of a misguided belief that errors in the data can always be made vanishingly small. This belief is misguided for two reasons.

(i) The severity of instabilities in many problems we deal with is so great that the gain in information on x obtained from improvements in data accuracy is small.

(ii) The representation of the observing system in the from of (x) always contains inescapable sources of uncertainty, including the errors associated with discretization of a continuum field.

A related characteristic of the inversion problems we deal with is a consequence of the inherent instability – that is that there are many solutions that represent the data and model used to define the solutions as shown in Fig. 10.x.

10.2 Further characteristics of inverse problems

Underdetermined problems: When the equation $\mathbf{y} = \mathbf{Kx}$ fails to provide enough information to determine uniquely (although not stably) all the elements of **x,** the problem is said to be *underdetermined*. These problems arise where there are more unknowns that data, i.e. when **y** is an *n* column vector of data and **x** is a *m* column vector of the unknowns where $m>n$.

Even determined problems: In this case there is exactly enough information to determine the required information (as in our simple 2x2 problem introduced above).

Overdetermined problems: When there is too much information contained in $\mathbf{y} = \mathbf{Kx}$ for it to possess an exact solution, the problem is said to be *overdetermined.* These problems typically have more equations than unknowns, $m<n$

Determining whether a problem is under or overdetermine, is however, not quite as obvious suggested by these comments. Many problems that appear to be over-determined in fact are underdetermined owing to that fact that the data (such as radiances measured at different wavelengths) do no contain independent information and thus are not independent of one another. This is illustrated in the simple 2x2 example above where we note that closeness of the Kernel values representing each of the 'measurements'. This is a sign of lack of independence and is the root cause for the solution instability.

Therefore most problems that arise in practice are neither completely overdetermined or underdetermined. (give example). These are referred to as *mixed determined* problems and ideally we would like to sort the unkown model parameters into two groups – those that are overdetermined and those that are underdetermined (TBD later).

10.3 Properties of vectors and matrices

We have seen by this example how small changes in the measurements lead to large changes in the resultant solution of a very simple 2x2 system (2 equations and 2 unknowns). However, many of our problems are posed for much larger systems and it is crucial that we have some way of understanding how uncertainties behave for these larger systems. These systems are also not always even determined and thus not necessarily square – i.e. the **K** matrix is often non-square.

The introduction of the matrix transpose allows us to apply methods to non square matrices that are only valid for square matrices. The transpose $\mathbf{K}^{\mathbf{T}}$ of a matrix **K** is obtained by interchanging rows of matrices with columns of the matrix. The diagonal elements are thus unaffected. A symmetric matrix has the property $\mathbf{K}^{\mathbf{T}} = \mathbf{K}$ and the product $\mathbf{K}^{\mathbf{T}}\mathbf{K}$ is referred to as the symmetric product since it results in a symmetric and square matrix even when **K** is asymmetric and non square. This product is encountered frequently in following chapters. The inverse of the square matrix $\mathbf{K}^{\mathbf{T}}\mathbf{K}$

$$(\mathbf{K}^{\mathbf{T}}\mathbf{K})^{-1} = \mathbf{K}^{-1}(\mathbf{K}^{\mathbf{T}})^{-1}$$

so

$$\mathbf{K}^{-1} = (\mathbf{K}^{\mathbf{T}}\mathbf{K})^{-1}\mathbf{K}^{\mathbf{T}} \qquad (10.3)$$

that is a simple post-multiplication of $(\mathbf{K}^{\mathbf{T}}\mathbf{K})^{-1}$ by $\mathbf{K}^{\mathbf{T}}$ yields $\mathbf{K}^{-1}$. Since **K** can be non-square, this procedure enables any real matrix to be inverted. The inverse of **K** as given by (10.3) will also be frequently encountered below.

Length and square norm of a vector: (length of $\mathbf{x})^2 = \mathbf{x}^{\mathbf{T}}\mathbf{x}$ which is also referred to as the dot product of vectors, or scalar product. $\mathbf{x}^{\mathbf{T}}\mathbf{x}$ is the square norm of the vector **x**.

Orthogonality of two vectors: The property of orthogonality is important for many applications. Two arbitrary vectors **u** and **v** are orthogonal when $\mathbf{u}^{\mathrm{T}}\mathbf{v}=0$

The quadratic form of **K**: Any square (general) matrix **K** can be represented uniquely as a sum of a symmetric matrix and a skew-symmetric matrix – the latter has the elements $k_{ij} = -k_{ji}$; $k_{ii} = 0$. This representation follows as

$$\mathbf{K} = \frac{1}{2}(\mathbf{K} + \mathbf{K}^{\mathbf{T}}) + \frac{1}{2}(\mathbf{K} - \mathbf{K}^{\mathbf{T}})$$

where the first term is the symmetric part and the second is the skew-symmetric part. If we consider to column vectors **x, y** then the product $\mathbf{x}^{\mathbf{T}}\mathbf{K}\mathbf{y}$ is called the bilinear form. For **x=y**, the product $\mathbf{x}^{\mathbf{T}}\mathbf{K}\mathbf{x}$ is the quadratic form and this depends only on the symmetric portion of **K**

$$\mathbf{x}^{\mathbf{T}}\mathbf{K}\mathbf{x} = \mathbf{x}^{\mathbf{T}}\frac{1}{2}(\mathbf{K} + \mathbf{K}^{\mathbf{T}})\mathbf{x}$$

10.4 Eigenvalues and eigenvectors

If **u** is a vector and **K** is a (square) matrix, then the product **Ku** produces another vector **y** which, in general, has no simple relation to **u**. It is natural to ask that for any arbitrary (square) matrix **K** there is a choice or choices of x that make **Ku** a simple scalar multiple of **u**, namely

$$\mathbf{K}\mathbf{u} = \lambda\mathbf{u}$$

When the relationship of this sort is satisfied, the vector **x** is designated as the eigenvector and the scalar quantity λ is the eigenvalue (the expression characteristic vector and characteristic value is also encountered. This expression implies that there is a single eigenvector and a single eigenvalue but a general NxN matrix will possess N eigenvectors and N eigenvalues. This follows from the fact that the matrix-vector equation system is a system of N scalar equations in N unknowns components of **u** and a single unknown λ.

If **K** is symmetric and non-singular the eigenvectors $\mathbf{u}_1,\ldots\mathbf{u}_n$ and the eigenvalues $\lambda_1,\ldots,\lambda_n$ categorize **K** completely. If K is not symmetric, there is a second eigen-equation

$$\mathbf{K}^{\mathbf{T}}\mathbf{v} = \lambda\mathbf{v}$$

which has the same eigenvalues as the first.

We will encounter the notion of the eigenvector and eigenvalue throughout this book, In the context of inverse problems, one of the important aspect of the eigenvalue/eigenvector decomposition of **K** follows be considering a vector defined as a linear combination of eigenvectors $\mathbf{u}_1,\ldots\mathbf{u}_n$

$$\mathbf{y} = \xi_1\mathbf{u}_1 + \xi_2\mathbf{u}_2 \ldots + \xi_n\mathbf{u}_n$$

Then the repeated pre-multiplication of **x** by **K** produces

$$\mathbf{K^m y} = \lambda_1^m\xi_1\mathbf{u}_1 + \lambda_2^m\xi_2\mathbf{u}_2 \ldots + \lambda_n^m\xi_n\mathbf{u}_n$$

From this expression we note that the inverse $\mathbf{K}^{-1}$ involves the division by the eigenvectors

$$\mathbf{K^{-1} y} = \lambda_1^{-1}\xi_1\mathbf{u}_1 + \lambda_2^{-1}\xi_2\mathbf{u}_2 \ldots + \lambda_n^{-1}\xi_n\mathbf{u}_n$$

The smallest eigenvalue thus tends to dominate the inversion. Now we begin to the difficulty with the direct inversion of our matrix **K.** If the smallest eigenvalue is very small its reciprocal is very large. If a small error in **y** creeps into the measurements, as it always does, then the error in the inversion $\mathbf{K^{-1}y}$ gets greatly magnified.

Returning to our simple 2X2 example,

$$\mathbf{K} = \begin{pmatrix} 1.0 & 1.0 \\ 2.0 & 2.0001 \end{pmatrix}$$

then λ_1=0.00033 and λ_2=3.0007. The smallness of the first eigenvalue could have been anticipated give the intrinsic instability of our simple problem.

Two important points drawn from this discussion warrant emphasis

(i) The fundamental nature of the inversion problem, its stability, uniqueness, and as we will see later, the information content characteristic of the problem is governed by the **K** matrix. As we will see below and may have anticipated previously, this **K** matrix is in turn directly related to the radiative transfer equation and the physical processes it represents.

(ii) The characteristics of the matrix **K** are fully expressed in terms of the eigenvalues and eigenvectors of this matrix.

10.5 Least Squares Solution

One reaction to the instability problem we have raised here is to seek more data in hopes that this will alleviate the difficulty. Clearly this will not solve our problem as the instability is an intrinsic property of **K.** Nevertheless, solutions to overdetermined problems are common and the most popular approach to solving over-determined problems is by invoking the least squares solution. We will also see that a number of other inversion methods derive from this most common approach.

Least squares provides a solution whereby we obtain the vector **x** of length N that minimizes the norm of the residual $\mathbf{Kx - y}$. The square norm, which gauges the magnitude of $\mathbf{Kx - y}$ can be written

$$\|\ell\|^2 = \mathbf{(Kx - y)^T (Kx - y)}$$
$$= \sum_{\mathbf{i}}^{\mathbf{N}} \left[\sum_{\mathbf{j}}^{\mathbf{M}} K_{ij} x_j - y_j \right] \left[\sum_{\mathbf{k}}^{\mathbf{M}} K_{ik} x_k - y_k \right]$$

With expansion and some rearrangement

$$\|\ell\|^2 = \sum_j^M \sum_k^M x_j x_k \sum_i^M K_{ij} K_{ik} - 2 \sum_j^M x_j \sum_i^N K_{ij} y_i + \sum_i^N y_i y_i$$

This square norm may be thought of in some sense as an error between a model prediction **Kx** and the data **y.** We can derive **x** to minimize this error from

$$\frac{\partial \|\ell\|^2}{\partial \mathbf{x}} = 0$$

which implies that

$$\mathbf{K^T K x - K^T y} = 0$$

or

$$\mathbf{x = (K^T K)^{-1} K^T y}$$

This is the least squares solution. Its geometric interpretation is illustrated in Fig. 10.2. If **Kx** is the closest point to **y** in the whole column space of **A**, then the line from **y** to **Kx** is perpendicular to that space.

Fig. 10.2 A geometric interpretation of the least squares solution

10.5.1 Constrained Least Squares

The least squares solution does not overcome the inherent instability we frequently encounter in inverse problems. The solution is no better than that governed by direct inverse – in fact the elements of $(\mathbf{K^T K})^{-1}$ tend to be even larger that those of $\mathbf{K}^{-1}$. Clearly this does nothing to improve the situation –rather it tends to exacerbate the problem since the root cause for the existence of small eigenvalues of **K** is not addressed.

The ambiguity can be removed by imposing an additional condition or citerion that may be evaluated with the measurements but one that is not derivable from the measurements. The purpose of this additional condition is to enable the selection of one **x** from a set of possible values. In many applications, this new condition is somewhat arbitrary whereas in other applications this new condition might represent our state of knowledge about the acceptable range of values **x** might take.

One constraint frequently used is the constraint that seeks to obtain a smooth distribution of **x**. Suppose that $q(\mathbf{x})$ is a non-negative scalar measure of the deviations of smoothness in x, then x can be varied such that $q(\mathbf{x})$ becomes a minimum (and zero if **x** is completely smooth). We incorporate this into the least squares procedure such that $\left[\mathbf{Kx}-\mathbf{y}\right]^T\left[\mathbf{Kx}-\mathbf{y}\right]$ is not minimized but rather $\left[\mathbf{Kx}-\mathbf{y}\right]^T\left[\mathbf{Kx}-\mathbf{y}\right]+\gamma q(\mathbf{x})$ where γ is a parameter that can be somewhat arbitrarily varied from zero to infinity. Obviously with $\gamma \to \infty$, minimization leads to q(**x**) =0 and a perfectly smooth solution as judged by the measure q. With $\gamma = 0$, we obtain the least-squares solution. Since the solution x that minimizes $\left[\mathbf{Kx}-\mathbf{y}\right]^T\left[\mathbf{Kx}-\mathbf{y}\right]$ does not in general minimize $q(\mathbf{x})$, the solutions obtained with non zero values of γ will produce different kinds of solutions that occur at larger values of the square norm than for the least squares solutions.

There are a number of different measures of smoothness we might adopt (see Twomey, 1997), but one measure is given by the relationship

$$q = \sum_i^N x_i^2$$

or alternatively as

$$\mathbf{x^T I x}$$

where **I** is the identity matrix. The minimization of

$$\left[\mathbf{Kx}-\mathbf{y}\right]^T\left[\mathbf{Kx}-\mathbf{y}\right]+\gamma\mathbf{x^T I x}$$

leads to

$$\mathbf{x} = (\mathbf{K^T K} + \gamma\mathbf{I})^{-1}\mathbf{K^T y}$$

This is one equation for the constrained linear inversion. Since γ is arbitrary, the usual approach is to choose several values of γ and *post facto* decide the most appropriate value for γ. As an example, consider our simple 2x2 case where we have

$$\boldsymbol{y} = \begin{pmatrix} 2.0 \\ 4.0 \end{pmatrix} \text{ and } \mathbf{K} = \begin{pmatrix} 1.0 & 1.0 \\ 2.0 & 2.0001 \end{pmatrix}$$

noting that the measurement error is included in ***y***. Suppose we assert a smoothness constraint such that γ=1, then

$$\mathbf{x} = \begin{pmatrix} 0.9 \\ 0.9 \end{pmatrix}$$

which is a perfectly smooth solution and close to the actual solution which too is perfectly smooth.

10.5.2 Inversion with a priori constraints

For many real inversion problems we have some general expectation of what the solutions should be drawn from accumulated knowledge of the physical problem being inverted. For example, many times the physical parameters represented by **x** need to be non-negative. We would thus like to accommodate this knowledge in some way so we

can discriminate between those solutions that give mathematically acceptable results but physically implausible results from these that are mathematically and physically acceptable. As we have seen, some problems give too broad a range of plausible results and a priori constraints can also be used to restrict this range to a smaller set of reasonable solutions.

It only requires the incorporation of these expectations into the constraints to push the solutions toward the constraint. There are a number of ways the expectation could be included – for example we could use the departure of x not from some smooth function but from certain statistical properties about x. This could be an average value derived from climatological data base. It is relatively straightforward to account for the tendencies that exist in past data and constrain the solution in some way to these tendencies. A simple way to do so is to derive a mean value for $\mathbf{x}$, say $\mathbf{x_a}$, and use the quadratic form

$$\left[\mathbf{Kx - y}\right]^T \left[\mathbf{Kx - y}\right] + \gamma(\mathbf{x - x_a})^{\mathbf{T}}(\mathbf{x - x_a})$$

and obtain $\mathbf{x}$ from the extremum of this relationship. It follows that

$$\mathbf{x} = (\mathbf{K^T K} + \gamma \mathbf{I})^{-1}(\mathbf{K^T y} + \gamma \mathbf{x_a})$$

When there is a reasonable basis for selecting $\mathbf{x_a}$ and γ, then this approach gives reasonable results. Consider again our simple 2x2 example as above with γ=1, and with

$$\mathbf{x_a} = \begin{pmatrix} 1.2 \\ 1.1 \end{pmatrix}$$

then the solution becomes

$$\mathbf{x} = \begin{pmatrix} 1.06 \\ 0.96 \end{pmatrix}$$

which (by design) more closely resembles the actual solutions.

10.5.3 Weighted least squares