AT721 Section 11:

Introduction to Inverse Problems

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

$$y(z) = \alpha(z)x(\chi, z) + \int_{a(z)}^{b(z)} K(\chi, z, z')x(\chi, z')dz' \tag{11.1}$$

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 (such as the single scatter albedo, asymmetry parameter or other properties that are not retrieved) or it may be a vector quantity representing desired information to be retrieved.

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

•$\alpha(z) = 0$ – (11.1) is an integral equation of the first kind
•$\alpha(z) = 1$ –(11.1) is an integral equation of the second kind
•a, b are constants – (11.1) is of a *Fredholm* type (first or second depending on $\alpha(z)$)
•a =constant, $b = z$ – (11.1) is of a *Volterra* type.

Inverse problems can be further classified depending on the focus of attention. If $x(\chi, z')$ 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, for example, then the problem is non linear. We will be concerned with inversions of (11.1) of both linear and non-linear types of problems.

11.1 The nature of the inverse problem

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

$$y(z) = \int_a^b K(z, z')x(z')dz' \tag{11.2}$$

and by introducing some type of quadrature to discretize this in the form

$$\mathbf{y} \approx \mathbf{K}\mathbf{x} \tag{11.3}$$

where $\mathbf{y}$ is a column vector of N measurements, $\mathbf{x}$ is a column vector of N source functions and $\mathbf{K}$ is an $N \times N$ matrix. Equation (11.3) is now in the form of a linear discrete inverse problem in contrast to its the continuous inverse counterpart (11.2). The solution of this discrete problem then takes the form

$$\mathbf{x} \approx \mathbf{K}^{-1}\mathbf{y} \tag{11.4}$$

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 the issues with the following problem. Suppose we make two measurements such that

$$\mathbf{y} = \begin{pmatrix} 2.0 \\ 4.0001 \end{pmatrix} \tag{11.5}$$

and suppose our problem is characterized by the Kernel of the form

$$\mathbf{K} = \begin{pmatrix} 1.0 & 1.0 \\ 2.0 & 2.0001 \end{pmatrix} \tag{11.6}$$

then we can determine that the solution to the pair of equations implied in (11.2) is

$$\mathbf{x} = \begin{pmatrix} 1.0 \\ 1.0 \end{pmatrix} \tag{11.7}$$

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

$$\mathbf{y} + \varepsilon = \begin{pmatrix} 2.0 \\ 4.0 \end{pmatrix} \tag{11.8}$$

then

$$\mathbf{x} = \begin{pmatrix} 2.0 \\ 0.0 \end{pmatrix} \tag{11.9}$$

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 $\mathbf{x}$ obtained from improvements in data accuracy is small.
(ii) The representation of the observing system in the from of (11.1) or its approximate form (11.3)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.

11.2 Further characteristics of inverse problems

Underdetermined problems: When the equation fails to provide enough information to determine uniquely (although not stably) all the elements of $\mathbf{x}$, the problem is said to be *underdetermined*. These problems arise where there are more unknowns that data, i.e. when $\mathbf{y}$ is an N column vector of data and $\mathbf{x}$ 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, however, is not quite as obvious as 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 values of the Kernel matrix 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).

11.3 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 such large systems. These systems are also not always even determined and thus not necessarily square - i.e. the $\mathbf{K}$ 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 $\mathbf{K}^T$ of a matrix $\mathbf{K}$ is obtained by interchanging rows of matrices with columns of the matrix. The diagonal elements are thus unaffected. A symmetric matrix has the property $\mathbf{K}^T = \mathbf{K}$ and the product $\mathbf{K}^T\mathbf{K}$ is referred to as the symmetric product since it results in a symmetric and square matrix even when $\mathbf{K}$ is asymmetric and non square. This product is encountered frequently in following chapters. The inverse of the square matrix $\mathbf{K}^T\mathbf{K}$

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

so

$$\mathbf{K}^{-1} = (\mathbf{K}^T\mathbf{K})^{-1}(\mathbf{K}^T)^{-1} \tag{11.10}$$

that is a simple post-multiplication of $(\mathbf{K}^T\mathbf{K})^{-1}$ by $\mathbf{K}^T$ yields $\mathbf{K}^{-1}$. Since $\mathbf{K}$ can be non-square, this procedure enables any real matrix to be inverted. The inverse of $\mathbf{K}$ as given by (11.10) will also be frequently encountered below. Other matrix/vector properties encountered include:

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

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

The quadratic form of $\mathbf{K}$: Any square (general) matrix $\mathbf{K}$ 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

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

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

$$\mathbf{x}^T\mathbf{K}\mathbf{x} = \mathbf{x}^T\frac{1}{2}(\mathbf{K}+\mathbf{K}^T)\mathbf{x} \tag{11.110}$$

11.4 Eigenvalues and eigenvectors

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

$$\mathbf{Ku} = \lambda\mathbf{u} \tag{11.12}$$

When the relationship of this sort is satisfied, the vector $\mathbf{u}$ 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 $N \times N$ matrix possesses 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 $\mathbf{K}$ is symmetric and non-singular the eigenvectors $\mathbf{u_1}, \ldots, \mathbf{u}_n$ and the eigenvalues $\lambda_1, \ldots, \lambda_n$ categorize $\mathbf{K}$ completely. For non symmetric $\mathbf{K}$, there is a second eigen-equation

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

which has the same eigenvalues as the first.

We will encounter the notion of the eigenvector and eigenvalue throughout the remainder of this book, In the context of inverse problems, one important aspect of the eigenvalue/eigenvector decomposition of $\mathbf{K}$ follows be considering a vector defined as a linear combination of eigenvectors $\mathbf{u_1}, \ldots, \mathbf{u}_n$

$$\mathbf{y} = c_1\mathbf{u_1} + c_2\mathbf{u_2} + \ldots + c_n\mathbf{u_n}$$

Then pre-multiplication of $\mathbf{y}$ by $\mathbf{K}$ produces

$$\mathbf{Ky} = c_1\mathbf{Ku_1} + c_2\mathbf{Ku_2} + \ldots + c_n\mathbf{Ku_n}$$

and it follows from (11.12) that

$$\mathbf{Ky} = c_1\lambda_1\mathbf{u_1} + c_2\lambda_2\mathbf{u_2} + \ldots + c_n\lambda_n\mathbf{u_n}$$

and by inference m premultiplications yield

$$\mathbf{K}^m\mathbf{y} = c_1\lambda_1^m\mathbf{u_1} + c_2\lambda_2^m\mathbf{u_2} + \ldots + c_n\lambda_n^m\mathbf{u_n}$$

Thus the inverse $\mathbf{K}^{-1}$ follows given $m = -1$ and thus involves the division by the eigenvectors

$$\mathbf{K}^{-1}\mathbf{y} = c_1\lambda_1^{-1}\mathbf{u_1} + c_2\lambda_2^{-1}\mathbf{u_2} + \ldots + c_n\lambda_n^{-1}\mathbf{u_n} \tag{11.13}$$

and furthermore reveals how the smallest eigenvalue tends to dominate the inversion. This now puts us in a position to understand the difficulty with the direct inversion of our matrix **K**. If the smallest eigenvalue is very small, approaching zero, then the reciprocal of this eigenvalue is very large. If a small error ε in **y** creeps into the measurements or even the model (i.e. **K** itself, as it always does, then the error in the inversion gets greatly magnified.

Two important points can be drawn from this discussion:

(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 may have been 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.

11.4.1 Hypothetical Illustrations

We considered here two examples that underscore these important points. Consider first our simple 2X2 example,

$$\mathbf{K} = \begin{pmatrix} 1.0 & 1.0 \\ 2.0 & 2.0001 \end{pmatrix} \tag{11.6}$$

for which $\lambda_1 = 0.00033$ and $\lambda_2 = 3.0007$. The smallness of the first eigenvalue could have been anticipated give the intrinsic instability of our simple problem.

Another hypothetical example is taken from Twomey (1977), chapter 6. Consider the retrieval problem defined as follows

$$y = \int_0^1 te^{-kt}x(t)dt$$

for the function $x(t) = 1 + 4(t - 1/2)^2$. By introduce quadrature, we have

$$y_i \approx K_{ij}x_j$$

where $x_j = x(t_j)$ and K_{ij} are the values of the function $t_je^{-k_it_j}$ determined for discrete values of t_j. Table 11.1 lists values of y obtained from this integral derived both analytically (labelled 'exact') and using a trapeziodal quadrature scheme (labelled 'quadrature') for different values of k. Table 11.2 lists results of the inversion

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

where $\mathbf{x}$ is a vector of values of $x(t_j)$ for discrete values of t_j and is obtained both using exact values of y_i and approximate values derived from quadrature. Both results are bad when compared to the original $x(t)$ yet the values obtained for y when using either one is very close to the original y as indicated in Table 11.1.

11.5 Least Squares Solutions

One reaction to the instability problem we have raised here is to seek more data in hopes that this will alleviate the difficulty thereby creating an overdetermined problem. Clearly this will not solve our problem as the instability is an intrinsic property of **K** and thus an intrinsic property of the physical basis of the remote sensing problem itself. Nevertheless, 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 $\mathbf{x}$ of length N that minimizes the norm of the residual $\mathbf{Kx} - \mathbf{y}$. The square norm, which gauges the magnitude of $\mathbf{Kx} - \mathbf{y}$ can be written

$$\|\ell\|^2 = (\mathbf{Kx} - \mathbf{y})^T(\mathbf{Kx} - \mathbf{y})$$

$$= \sum_i^M [\sum_j^M K_{ij}x_j - y_j][\sum_k^M K_{ik}x_k - y_k]$$

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 $\mathbf{Kx}$ and the data $\mathbf{y}$. We can derive $\mathbf{x}$ to minimize this error from

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

which implies that

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

or

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

This is the least squares solution. Its geometric interpretation is illustrated in Fig. 10.2. If $\mathbf{Kx}$ is the closest point to $\mathbf{y}$ in the whole column space of $\mathbf{K}$, then the line from $\mathbf{K}$ to $\mathbf{Kx}$ is perpendicular to that space.

Fig. 10.2 A geometric interpretation of the least squares solution

11.5.1 Constrained Least Squares Solutions

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 $(\mathbf{K}^T\mathbf{K})^{-1}$ tend to be even larger that those of $\mathbf{K}^{-1}$ and in some sense creating an even worse solution. 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 $\mathbf{K}$ 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 $\mathbf{x}$ 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 $\mathbf{x}$ might take.

One constraint frequently used is the constraint that seeks to obtain a smooth distribution of $\mathbf{x}$. Suppose that q(x) is a non-negative scalar measure of the deviations of smoothness in x, then $\mathbf{x}$ can be varied such that $q(\mathbf{x})$ becomes a minimum (and zero if $\mathbf{x}$ is completely smooth). We incorporate this into the least squares procedure such that $(\mathbf{Kx} - \mathbf{y})^T(\mathbf{Kx} - \mathbf{y})$ is not minimized but rather $(\mathbf{Kx} - \mathbf{y})^T(\mathbf{Kx} -$

$\mathbf{y}) + \gamma q(\mathbf{x})$ where γ is a parameter that can be somewhat arbitrarily varied from zero to infinity. Obviously with $\gamma \to \infty$, minimization leads to $q(\mathbf{x}) = 0$ and a perfectly smooth solution as judged by the measure q. With $\gamma = 0$, we obtain the least-squares solution. Since the solution $\mathbf{x}$ that minimizes $(\mathbf{Kx} - \mathbf{y})^T(\mathbf{Kx} - \mathbf{y})$ 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, 1977), but one measure is given by the relationship

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

or alternatively as $\mathbf{x}^T\mathbf{Ex}$ where $\mathbf{E}$ is the identity matrix. The minimization of $(\mathbf{Kx} - \mathbf{y})^T(\mathbf{Kx} - \mathbf{y}) + \gamma\mathbf{x}^T\mathbf{Ex}$ leads to

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

which is a solution for a constrained linear inversion. Since is γ arbitrary, the usual approach is to choose several values and *post facto* decide the most appropriate value. As an example, consider our simple 2x2 case as defined by (11.6) and (11.8), noting that the measurement error is included in the specification of $\mathbf{y}$. Suppose we assert a smoothness constraint such that $\gamma = 1$, 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.

11.5.2 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 $\mathbf{x}$ not from some smooth function but from certain statistical properties about $\mathbf{x}$ 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

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

and proceed as above to obtain the solution for $\mathbf{x}$ from the extremum of this relationship. It follows that this solution has the form

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

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 $\gamma = 1$, 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.

11.5.3 Weighted least squares solutions

TABLE 6.1

Values of $\int_0^1 x\, e^{-yx} f(x)\, dx$ computed by quadrature and exactly

y	Exact value	Quadrature	Error
0.10	0.6218	0.6226	0.0008
0.20	0.5803	0.5810	0.0007
0.30	0.5419	0.5426	0.0007
0.40	0.5063	0.5070	0.0007
0.50	0.4734	0.4740	0.0006
0.60	0.4430	0.4435	0.0005
0.70	0.4147	0.4153	0.0006
0.80	0.3886	0.3891	0.0005
0.90	0.3643	0.3648	0.0005
1.0	0.3418	0.3422	0.0004
1.2	0.3015	0.3019	0.0004
1.4	0.2667	0.2670	0.0003
1.6	0.2366	0.2369	0.0003
1.8	0.2106	0.2109	0.0003
2.0	0.1880	0.1882	0.0002
2.5	0.1434	0.1436	0.0002
3.0	0.1116	0.1117	0.0001
3.5	0.0885	0.0886	0.0001
4	0.0714	0.0715	0.0001
5	0.0490	0.0490	0.0000
6	0.0356	0.0356	0.0000
7	0.0271	0.0271	0.0000
8	0.0214	0.0214	0.0000
9	0.0173	0.0173	0.0000
10	0.0144	0.0144	0.0000

TABLE 6.2

Result of direct inversion $f' = A^{-1}g$ on data of Table 6.1 compared with the $f(x)$ from which the data was computed

i	x_i	$f(x_i)$	f'_i from "exact" g	f'_i from quadrature g
1	0.05	1.81	7979	2280
2	0.1	1.64	−26,222	−7287
3	0.15	1.49	48,215	12,860
4	0.2	1.36	−41,271	−9589
5	0.25	1.25	−11,197	6380
6	0.30	1.16	41,763	15,276
7	0.35	1.09	5160	−1504
8	0.4	1.04	35,417	−10,588
9	0.45	1.01	−20,377	2487
10	0.5	1.00	39,681	1265
11	0.55	1.01	24,757	6342
12	0.6	1.04	−39,393	−4331
13	0.65	1.09	−11,035	−2797
14	0.7	1.16	41,353	10,929
15	0.75	1.25	−61,988	−25,922
16	0.8	1.36	67,292	23,292
17	0.85	1.49	−21,011	4742.1
18	0.9	1.64	−23,273	−20,714
19	0.95	1.81	19,687	11,774
20	1.00	2.00	−1158	−318.2