INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION
INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
I.C.T.P., P.O. BOX 586, 34100 TRIESTE, ITALY, CABLE: CENTRATOM TRIESTE

H4.SMR/480-5

WORKSHOP ON EARTHQUAKE SOURCES
& REGIONAL LITHOSPHERIC
STRUCTURES FROM SEISMIC WAVE DATA

19 - 30 November 1990

An Easy Introduction Into Seismic Tomography

Guust Nolet

University of Utrecht
Dept. of Theoretical Geophysics
Utrecht, The Netherlands

AN EASY INTRODUCTION INTO SEISMIC TOMOGRAPHY

Guust Nolet

Lecture Notes, Trieste 1990

The underdetermined inversion problem

Suppose you go to the market to buy some food. You decide to buy one apple and one banana, and the vendor asks you one Piaster (or whatever the local currency is). At home your wife asks you: "how much does an apple cost?". Your correct answer should probably be:

$$a + b = 1 .$$

which will drive most housewives into a fit. In that case it won't help you to refer to figure 1, where the line $b=1-a$ delineates all possible solutions:

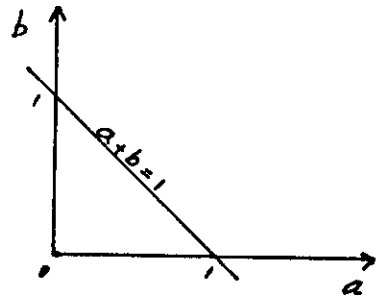


Figure 1: all solutions of $a+b=1$ lie on a straight line

It seems like a stupid example, but it actually illustrates a very basic characteristic of geophysical inversion problems. For instance, you measure the two-way travel time of a seismic reflection - say it is 830 milliseconds (figure 2). Convinced that this particular reflection is strong enough to signal the presence of a lucrative gas field, the oil company manager asks you to: "how deep is the gas field?". Your correct answer should probably be:

$$z = \frac{0.830}{v_{av}}$$

where (you guessed it), v_{av} is the average seismic velocity over depth z. This answer, too, might throw the questioner into a fit.

The similarity between the two examples is that *one* measurement is available to say something about *two* unknowns: a and b, or z and v_{av}. The problem is *underdetermined*, like most, if not all realistic geophysical problems. We seem to be at a loss. Or are we?

In the case of the apple, we can certainly be a bit more specific than just saying $a+b=1$. For example, we know that cost is generally a positive quantity, so we may state that $a>0$ and $b>0$. Thus, instead

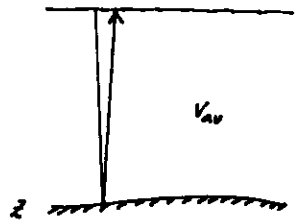


Figure 2

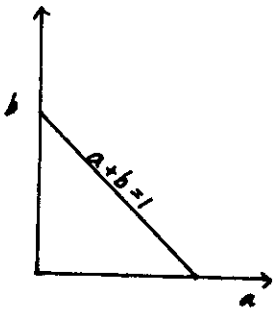


Figure 3a

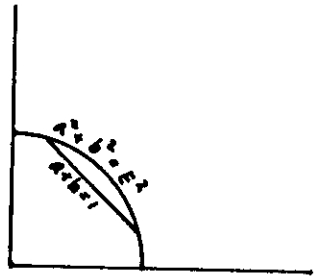


Figure 3b

of figure 1, we have a more limited set of solutions (figure 3a): Such nonlinear constraints are very helpful to restrict the *solution space*. In figure 2, the solution space is one dimensional (a line). In some cases the constraints are a little more complicated. If the unknowns are the harmonic coefficients of the Earth magnetic field, the finite energy of the field puts a quadratic bound on the length of the vector of such coefficients. In our case, the equivalent constraint would be $a^2+b^2 \le E^2$ (see figure 3b). If such a constraint could be applied to our apple problem, we might even end up with an unique solution if $E^2=½$! Backus (1988, 1989) has made an in-depth study of the influence of such bounds on geophysical inverse problems, although the more pedestrian approach of Jackson (1979) is probably easier to understand for starters in the subject.

You may wish to know at this stage what all this has to do with seismic tomography. The answer is that the simple apple example is able to show most of the features of the large tomographic problem, but at a much smaller scale of 2 or 3 unknows and data. We could use a tomographic analogon, using an Earth model that is represented by only two velocities - taking constant velocity in the crust and mantle, for example. The apple problem just seems a bit more real at small dimensions.

The minimum norm solution

Even though we have reduced the dimensionality of the model space, a solution in the form of a subspace such as shown in figure 3 is not practical - if only because more realistic inverse problems have more than 2 dimensions. Some tomographic problems may have a dimension of order 10^6! We would often like to present a preferred model (or *point estimate*). We obtain a preferred model by specifying a criterion that distinguishes the "good" models from the others. This always involves some degree of subjectivity, and we shall in general be obliged to state our subjective preference clearly.

As an example, we might state that we would prefer the price of an apple to be as close as possible to that of a banana. In figure 4, we now draw a line for equal prices $a=b$:

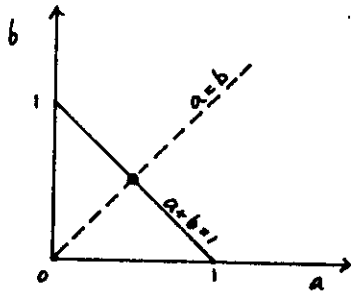


Figure 4: the preferred solution not only has $a+b=1$ but also $a=b$ (other choices are possible as well, depending on our a priori preferences).

This clearly yields a point estimate $a=b=½$. the example is less arbitrary as may seem at first sight. Note that the line $a=b$ is also the solution with minimum norm a^2+b^2. Often our unknowns are *deviations* from a standard model. If the standard model is preferred for some reason, the minimum norm model deviates least from this model. For example, in seismology our preferred model is often homogeneous, and we try to find the model with the least possible heterogeneity that still satisfies the data. Other criteria are possible, depending on the problem at hand and the a priori information available.

In the case of the seismic reflection, it would make little sense to minimize the length of the solution vector (z, v_{av}), since both components have different physical dimension. But perhaps we have some a priori idea. For instance, from other measurements in the area we know that there is a reflective horizon at depth $z_0 \pm \sigma_z$, where σ_z is the statistical standard deviation. And, similarly, we expect the velocity to be near $v_0 \pm \sigma_v$. We can make the problem linear by introducing the slowness $s_{av}=1/v_{av}$, with $s=s_0 \pm \sigma_s$, $\sigma_s=\sigma_v/v_{av}^2$:

$$z - 0.830 s_{av} = 0$$

The solution space is therefore the line $z=0.830s_{av}$. We stay closest to our a priori model if $z=z_0$ and $s=s_0$, but this is not necessarily a solution. What deviation of z from z_0 is to be considered equally bad as a deviation of s from s_0? A logical answer is to weigh the deviations by their standard deviations:

$$\left[\frac{z-z_0}{\sigma_z}\right]^2+\left[\frac{s-s_0}{\sigma_s}\right]^2=MIN$$

Our preferred model is therefore given by that model on $z=0.830s_{av}$ for which this expression is a minimum. If we use scaled parameters $z'=z/\sigma_z$ and $s'=s/\sigma_s$, we can easily find this solution by projection, as in figure 5.

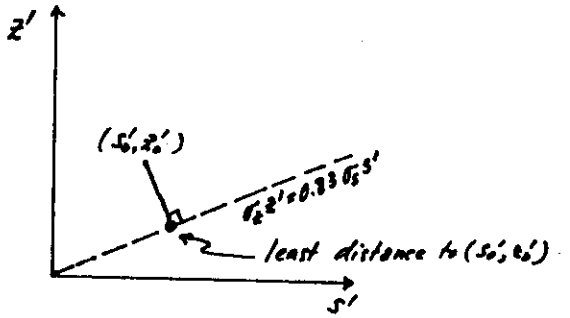


Figure 5: the preferred solution is found by projection, which is orthogonal if we use scaled parameters.

However, this is not generally how we solve such a system of equations. In the case of the apple problem, we can easily see how to find the minimum norm solution if we view the equation as an inner product of the vector (1,1) with the vector (a,b):

$$(1,1)\cdot(a,b)=1$$

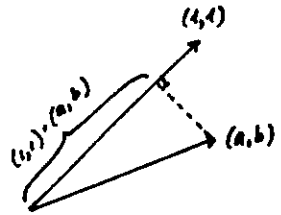


Figure 6: a linear equation is an inner product between the vector of coefficients and the vector of unknowns.

For all vectors (a,b) of a given length, the resulting product is maximized if $(a,b)//(1,1)$. Conversely, the vector (a,b) with the smallest length for which the inner product has the value 1 is a vector parallel to (1,1), or with $a=b$, from which it follows that $a=b=½$.

How does this work with more complicated problems? Suppose we have m linear equations, with $n>m$ unknowns. In matrix notation:

$$\mathbf{A}\,\mathbf{x}=\mathbf{d}$$

where $\mathbf{x}$ is the model, and $\mathbf{d}$ the datum vector. The matrix $\mathbf{A}$ has m rows, which we designate $\mathbf{a}_1,\mathbf{a}_2,\cdots,\mathbf{a}_m$. The matrix multiplication is essentially a series of inner products like in the apple problem. For example, the third datum satisfies: $\mathbf{a}_3\cdot\mathbf{x}=d_3$.

Now, analogous to the one dimensional case, let us suppose that the preferred solution $\mathbf{x}$ is one that can be constructed from vectors parallel to the rows of A, and *no other components*. Thus, we write this particular solution $\mathbf{x}_0$ as a linear combination $w_1\mathbf{a}_1+w_2\mathbf{a}_2+,\cdots,w_n\mathbf{a}_m$:

$$\mathbf{x}_0=\mathbf{A}^T\mathbf{w}$$

$$\mathbf{A}\mathbf{A}^T\mathbf{w}=\mathbf{d}$$

which is a $m\times m$ system of linear equations. We may show that this $\mathbf{x}_0$ is the minimum norm solution. For, suppose the $\mathbf{a}_1$ to $\mathbf{a}_m$ are independent, they span a m-dimensional subspace of the n-dimensional model space. We can supplement this with $n-m$ extra basis vectors $\mathbf{b}_i$ that are mutually orthonormal and also orthogonal to every one of the basis vectors $\mathbf{a}_i$. The most general expression for a solution is given by:

$$\mathbf{x}_{sol}=\mathbf{x}_0+\sum_i v_i\mathbf{b}_i$$

where the v_i can take arbitrary values without affecting the fit to $\mathbf{d}$, since the orthogonality of the vectors $\mathbf{a}_i$ and $\mathbf{b}_i$ implies $\mathbf{A}\mathbf{b}_i=\mathbf{0}$. The norm of $\mathbf{x}_{sol}$ is:

$$|\mathbf{x}_{sol}|^2=|\mathbf{x}_0|^2+\sum v_i^2$$

Since the w_i are determined by the linear system, there is only one $\mathbf{x}_0$, and $\mathbf{x}_{sol}$ is minimized for all $v_i=0$ so that it equals $\mathbf{x}_0$. In case some of the $\mathbf{a}_i$ are dependent, so that $\mathbf{A}\mathbf{A}^T$ is a singular matrix with determinant 0, we can adapt the proof (we just need more than $n-m$ vectors $\mathbf{b}_i$ to span the complete model space).

A very important application of the above principle of a minimum norm solution is given by an iterative solution technique used in medical tomography - and briefly applied in seismic tomography as well although there are now better algorithms available. It is now known as the Algebraic Reconstruction Technique (ART), but actually a re-invention of an algorithm proposed by the Polish mathematician Kaczmarz in 1937. In seismic tomography, we face the problem that n and m are very large: often we try to interprete $m=10^6$ data in terms of a model, given by slownesses x_i in $n=10^4$ or more cells - note that in this case $m>>n$ rather than $m<n$, yet the severe dependencies among the matrix rows generally results in a solution space of dimension $k<n$, so that, again, the solution is not unique and we search for a minimum norm solution. In ART we do this as follows. First we fit the first datum, d_1, by choosing a first iterate $\mathbf{x}^{(1)}$ parallel to the first row, given by vector elements A_{1j}.

$$x_j^{(1)}=\frac{A_{1j}d_1}{\sum_k A_{1k}^2}$$

That leaves a residual $\mathbf{r}^{(1)}=\mathbf{d}-\mathbf{A}\mathbf{x}^{(1)}$, where of course $r_1=0$. We now wish to add a correction to $\mathbf{x}^{(1)}$, such that the new vector $\mathbf{x}^{(2)}$ reduces the second element of the residual vector to 0. Again, we find the minimum norm correction by choosing the correction parallel to the row of A that belongs to the

datum we are fitting. A little algebra results in:

$$x_j^{(2)} = x_j^{(1)} + \frac{A_{2j} r_2^{(1)}}{\sum_k A_{2k}^2}$$

Repeating this many times:

$$x_j^{(q+1)} = x_j^{(q)} + \frac{A_{ij} r_i^{(q)}}{\sum_k A_{ik}^2}$$

with $l = q \bmod m$. Since the solution is again built as a linear combination of the rows of A, we shall obtain the minimum norm solution (but see next section for the common case that some of the equations are inconsistent).

You may have noticed that the only matrix multiplication in the problem is a forward multiplication of the form Ax. To compute one element of r we need only one row of A in computer memory. Thus, since A can be stored on disk, iterative methods like ART are very suitable for use in large dimensional problems such as in seismic tomography. Moreover, as we shall see in the second lecture of this series (or see Nolet, 1987), the matrix element A_{ij} contains the length of ray i in cell j if the Earth is parameterized in cells. Since one ray visits only few cells, the number of matrix elements that are nonzero is very much less than n in every row. We can save considerable disk space by storing not every A_{ij}, but a series of pairs j, A_{ij} for nonzero A_{ij} only.

ART is simple, but it is also extremely slow in converging. Let us apply it to the apple problem. You may wish to do this yourself using the following simple BASIC program:

```
  10 DIM A(20,20),R(20),X(20),D(20),ROW(20)
  20 INPUT "how many data";M: INPUT "how many unknowns";N
  30 FOR I=1 TO M: PRINT "input row",I: FOR J=1 TO N: INPUT A(I,J): NEXT J:
NEXT I
  40 PRINT "input data": FOR I=1 TO M: INPUT D(I): NEXT I
  50 FOR I=1 TO M: ROW(I)=0!: FOR J=1 TO N: ROW(I)=ROW(I)+A(I,J)^2: NEXT J:
NEXT I
  60 FOR I=1 TO N: X(I)=0!: NEXT I
  70 DNORM=0!: FOR I=1 TO M: DNORM=DNORM+D(I)^2: NEXT I
  80 EPS=1E-08*DNORM
  85 PRINT "iter      residual     x(1)"
  90 FOR Q=1 TO 1000
  100 GOSUB 160: GOSUB 200
  110 IF Q>10 AND Q MOD 10 <> 0 GOTO 120
  115 PRINT USING "####";Q;: PRINT "  ";
  116 PRINT USING "   #.###^^^^";SQR(RNORM),X(1)
  120 IF RNORM < EPS THEN GOTO 140
  130 NEXT Q
  140 PRINT "solution": FOR I=1 TO M: PRINT I,X(I): NEXT I
  150 STOP
  160 REM compute residual vector
  170 FOR I=1 TO M: R(I)=D(I): FOR J=1 TO N: R(I)=R(I)-A(I,J)*X(J): NEXT
J:NEXT I
  180 RNORM=0: FOR I=1 TO M: RNORM=RNORM+R(I)^2: NEXT I
```



```
  190 RETURN
  200 REM compute updated x
  210 I=(Q+1) MOD M + 1
  220 RR=R(I)/ROW(I)
  230 FOR J=1 TO N: X(J)=X(J)+A(I,J)*RR: NEXT J
  240 RETURN
```

Running this on the original problem, we immediately obtain the preferred solution $a=b=½$. So let us now buy one apple and two bananas. the price is 1.40 Piaster, so we now have the double system:

$$a + b = 1 \quad \text{and} \quad a + 2b = 1.4$$

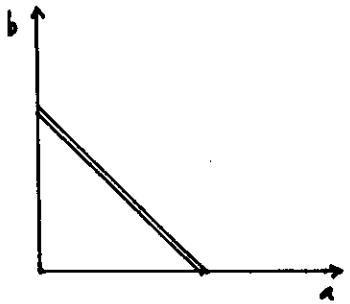
Running the little BASIC program, we see that the solution starts again with $a=b=½$, as expected since it only treats the first equation in step one. In the second step, the result is $a=0.48$, $b=0.46$, in agreement with the second equation and already closer to the exact solution $a=0.6$, $b=0.4$. But in the third step we actually do worse again: $a=0.51$, $b=0.49$. Such oscillatory behaviour is characteristic for ART. After 10 steps we are no further than $a=0.521$, $b=0.439$, and we need 100 steps to get to within 10^{-3} of the true solution.

Inconsistencies

It is interesting to see what happens to the ART solution when two rows are dependent, but the data are inconsistent. For example, if the vendor rounds off, we might by 10 apples and 10 bananas not for 10 Piaster, but for 9.9:

$$a + b = 1 \quad \text{and} \quad 10a + 10b = 9.9$$

A little thinking, or a quick check with the BASIC program, shows you what happens: whenever r_1 is made 0, $a=b=0.500$, but when $r_2=0$ the solution is $a=b=0.495$. Since there is no end to such oscillatory behaviour, ART does not converge. It reflects the obvious dillemma that our problem is not only underdetermined, but that there is no solution that satisfies the equations.



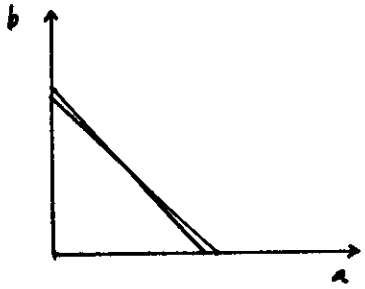


Figure 7: two inconsistent equations because of data errors. Figure 8

The inconsistency is possible because the two equations are dependent. We might be tempted to think that we can get rid of the problem by buying only 9 bananas, so that the equations become

independent. Of course, we shall have no problems if there are no round-offs in the data. For example, if the true prices are a=0.4 and b=0.6, and the vendor asks 1 Piaster for one of each, and 9.4 for a bag with $10a+9b$, we indeed find a=0.4, b=0.6 either by hand or with the BASIC program. Instead of figure 7, the lines are not parallel (figure 8), and there is a solution. But the solution point will shift drastically if there is a small error in the datum vector $\mathbf{d}$ =(1,9.4). For example, suppose the vendor rounds off and asks 9.5 Piaster for the bag - an error of merely 1% in d_2. Computing (a,b) we find: a=b=0.5, an error of 20% in the solution.

Before taking a second look at such problems of numerical instability, let us first see how we should handle incompatible, or overdetermined equations. Consider a more general system of equations $\mathbf{Ax}=\mathbf{x}$, such as:

$$\begin{bmatrix}1 & 1\\ 2 & -3\\ 2 & 5\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix}=\begin{bmatrix}3\\ 2\\ 10\end{bmatrix}$$

A graphical representation (figure 9) immediately shows us that there is no exact solution to this. Every pair of lines crosses in a different point.

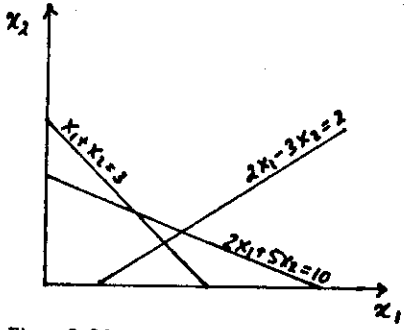


Figure 9: 3 incompatible equations

There is another way to view this. The coefficients x_1 and x_2 of $\mathbf{x}$ are really weights with which the column vectors (1,2,2) and (1,-3,5) are added in the product $\mathbf{Ax}$. From figure 9 it is clear that there is no set of weights that produces the vector (3,2,10). Let us denote the columns of A by a set of n vectors $\mathbf{a}^1,\mathbf{a}^2,\cdots\mathbf{a}^n$ (not to be mixed up with the *row* vectors denoted by low indices such as $\mathbf{a}_k$). Vectors $\mathbf{a}^1$=(1,2,2) and $\mathbf{a}^2$=(1,-3,5) span a two dimensional plane in three dimensional space (figure 10). This plane is known as the *range* of A or $R(\mathbf{A})$ - it is the set of all vectors $\mathbf{Ay}$. The fit to the equations is given by the vector $\mathbf{r}$=$\mathbf{d}$-$\mathbf{Ax}$. Since $\mathbf{r}$ cannot be $\mathbf{0}$, the best we can do is to make its length as small as possible. Thus we wish that $\mathbf{r}$ is perpendicular to $\mathbf{Ay}$ for arbitrary $\mathbf{y}$. In terms of an inner product:

$$(\mathbf{r},\mathbf{Ay})=(\mathbf{d}-\mathbf{Ax},\mathbf{Ay})=0$$

using the fact that $(\mathbf{a},\mathbf{Ab})$=$(\mathbf{A}^T\mathbf{a},\mathbf{b})$:

$$(\mathbf{A}^T\mathbf{d}-\mathbf{A}^T\mathbf{Ax},\mathbf{y})=0$$

Since this has to be true for arbitrary $\mathbf{y}$, the other vector must be identical to $\mathbf{0}$, which implies:

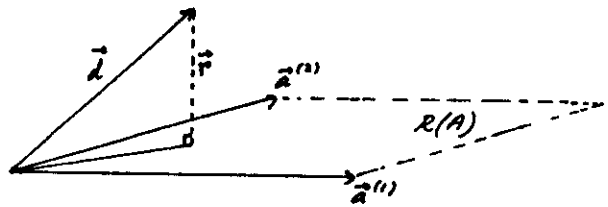


Figure 10: the datum vector $\mathbf{d}$ does not lie in the range of A.

$$\mathbf{A}^T\mathbf{Ax}=\mathbf{A}^T\mathbf{d}$$

This is an $n\times n$ system of equations, which has a guaranteed exact solution since both sides are in the range of $\mathbf{A}^T$. The solution is known as the *least squares solution* to the problem.

There may be a problem if $\mathbf{A}^T\mathbf{A}$ is a singular matrix, because some of the columns $\mathbf{a}^k$ are dependent. In that case we can simply treat it as before, and select the minimum norm solution. This will, however, not free us from the problem of numerical instabilities, caused by columns that are *nearly* parallel.

Numerical instabilities: regularization

We have seen earlier how a 1% error in $\mathbf{d}$ could lead to a 20% error in $\mathbf{x}$. An even more drastc example is given by:

$$\begin{bmatrix}1 & 1\\ 2 & 2.0001\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix}=\begin{bmatrix}3\\ 3.0002\end{bmatrix}$$

which has the exact solution x_1=x_2=1. Now suppose we make a very minor error in the data, so that the right hand side vector is (3,3) instead of (3,3.0002). Solving again, we find x_1=3, x_2=0, quite a difference! Obviously, although the equations are technically independent, they are so nearly the same that small errors may have a large influence.

If such a minor error blows x_1 up by a factor of 3, how much do we suffer from major errors? Delay times in seismic tomography are of the order of 1-2 seconds for teleseismic P waves, but generally suffer from errors larger than 0.5 seconds. So, what if our data vector is (3,3.5) instead of (3,3.0002)? The solution is now x_1=-9997, x_2=5000! Our intuition is usually able to spot such solutions as unreasonable. This intuition then can be used to 'regularize' the problem.

Usually, we are able to formulate an inverse problem in a form that allows for $\mathbf{x}$=0 as a 'reasonable' solution, and in which solutions with large norm $|\mathbf{x}|$ are suspect. We might add this information to our system of equations. The fact that x_1 and x_2 should be small is expressed by two new linear equations:

$$\epsilon x_1 = 0 \quad , \quad \epsilon x_2 = 0$$

It may seem that we now have a solution x=0, but that is not so, since we have also the two original equations, which are definitely not solved by the null vector. In fact, by adding the two new constraints, we end up with an inconsistent system again. ϵ serves to weight the two parts(the original part and the *damping* part) of the system of equations with respect to each other. The full system becomes:

$$\begin{bmatrix} 1 & 1 \\ 2 & 2.0001 \\ \epsilon & 0 \\ 0 & \epsilon \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 3.0002 \\ 0 \\ 0 \end{bmatrix}$$

And the least squares system is:

$$\begin{bmatrix} 2+\epsilon^2 & 4.0002 \\ 4.0002 & 8.0004+\epsilon^2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 6.0002 \\ 12.0007 \end{bmatrix}$$

in which we assumed 5 figure computing accuracy. If ϵ^2 becomes very large, we see that the diagonal dominates, so that both x_1 and $x_2 \rightarrow 3/epsilon^2$. In that case the damping completely determines the solution. As we relax ϵ^2, the solution gets closer to the original, but on the other hand the adat errors get more influence.

Note that for no $\epsilon > 0$ do we ever get the exact solution, even if the data are without errors. Damping biases the solution towards small solutions. Alternatives are possible. For example, if we wish to minimize the difference between x_1 and x_2, we can add an equation $x_1 - x_2 = 0$.

Seismic tomography

All of the above described features: underdetermined and inconsistent equations, slow convergence of the iterative solutions, amplitude bias because of damping, are important in seismic tomography, as we shall see in the next lecture.

Here we only remark that many better iterative algorithms exist than ART. In particular, projection methods such as LSQR converge much more rapidly, and its solutions possess damped properties if the iterations are stopped prematurely. This we are usually forced to do because computer time is limited. For a very good introduction into the numerical analysis of iterative algorithms, see van der Sluis and van der Vorst (1987).

References

Backus, G., Comparing hard and soft prior bounds in geophysical inverse problems, geophys. J. Roy. astr. Soc., 94, 249-261, 1988.

Backus, G., Confidence set inference with a prior quadratic bound, Geophys. J. Roy. astr. Soc., 97, 119-150, 1989.

Jackson, D.D., Interpretation of inaccurate, insufficient and inconsistent data, Geophys. J. Roy. astr. Soc. 28, 79-109,1972.

Jackson, D.D., The use of a priori data to resolve non-uniqueness in linear inversion, Geophys. J. Roy. astr. Soc., 57, 137-158, 1979.

Nolet, G., Seismic wave propagation and seismic tomography, in: Nolet, G. (ed), *Seismic Tomography*, Reidel, Dordrecht, 1987.

Tarantola, A., *Inverse Problem Theory*, Elsevier, Amsterdam 1987.

van der Sluis, A., and van der Vorst, H.A., Numerical solution of large, sparse linear algebraic systems arising from tomographic problems, in: Nolet, G. (ed)., *Seismic Tomography*, Reidel, Dordrecht, 1987.