

Lecture 18: Delay time tomography

David Al-Attar, Michaelmas term 2023

Outline and Motivation

In the previous lecture we discussed the use of ray theory in determining the Earth's spherical velocity structure. Today we consider an extension of these ideas, known as **delay time tomography**, that can be used to produce quantitative images of lateral velocity variations within the Earth. This method falls into a broader class of techniques known collectively as **seismic tomography**, and we will say more about some other variants in later lectures. Such work is done to better understand convection in the mantle, with the tomographic models obtained acting as an indirect observational constraint on this process. Delay time tomography is an example of a linear inverse problem, and we discuss a practical method for its solution known as **regularised least squares**.

Laterally varying structure from seismology

Though the Earth is close to being spherically symmetric, the small deviations from sphericity are still of great interest. This is because they can tell us about processes such as mantle convection deep within the Earth that are inaccessible to direct observation. The area of seismology concerned with “imaging” lateral variations in the Earth’s interior is known as **seismic tomography**, and is one of the main areas of current research within geophysics. The result of such studies are **tomographic models** that show variations in seismic wave speed and other parameters within the Earth. These variations are defined relative to a fixed spherically symmetric reference model, and so include only the laterally varying perturbations. An early example of such a tomographic model is shown in fig.1 which was produced by John Woodhouse and Adam Dziewonski who pioneered seismic tomography at Harvard University during the 1980s.

The basic idea underlying seismic tomography is that if we *knew* the Earth’s internal structure then we could calculate what *would* be observed after an earthquake, and so we can try to *work backwards* to determine this structure from seismic observations. This is an example of an **inverse problem**, and in this lecture we shall discuss the solution of such problems in further detail. In particular, we will focus on travel time measurements and that are modelled using ray theory. This was the earliest tomographic approach developed, with the first papers written in the late 1970s. Many of the basic ideas, however, carry over to more complicated tomographic techniques including **waveform tomography** that we will discuss next time.

Delay time measurements

If we know the location and origin time of an earthquake, then we have seen that it is possible to measure the travel times of various seismic phases recorded on seismograms around the globe. If the Earth were spherically symmetric then the observed travel times would (within the limits of observational errors) depend only on the epicentral angle

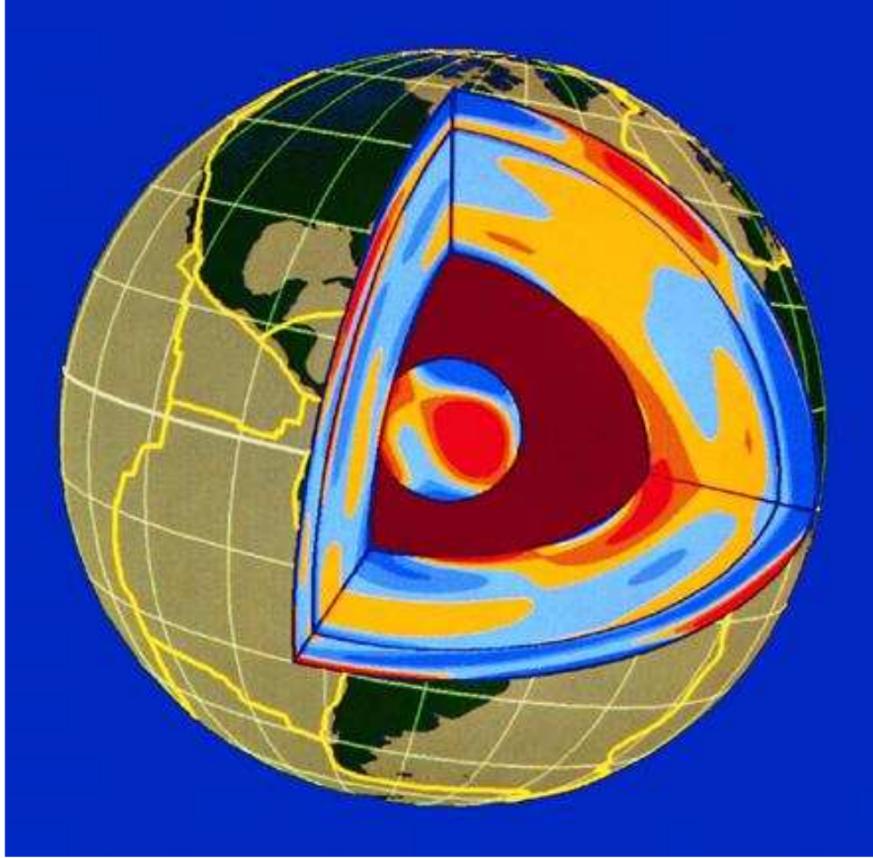


Fig. 1: An image of a tomographic model of S-wave velocity variations within the Earth produced by John Woodhouse and Adam Dziewonski. The colours show the perturbation in S-wave velocity relative to a spherical reference model, with blue areas being faster than normal, and red areas slower. Areas of faster velocity are thought to be colder than normal, while those that are slower are hotter, and hence the chosen colour scale. These temperature anomalies have associated changes in density, and so the models tell us something about the buoyancy forces that drive mantle convection.

and the source depth. Though this is a reasonable first approximation, there do exist variations in observed travel times due to lateral variations within the Earth. For a given travel time measurement, we define its associated **delay time** by

$$\delta T = T - T_0, \quad (1)$$

where T is the measured travel time, and T_0 is the travel time of the seismic phase predicted by a spherically symmetric reference model. We note that this definition is such that the delay time will be positive if the observed phase arrives later than expected. Such delay times can be measured for many source-receiver pairs and for many seismic phases, and constitute the data in **delay time tomography**. To some extent these times can be measured by hand. But more often this is done using signal processing tools such as cross-correlation, or increasingly using forms of machine learning¹.

¹ In my opinion, this is one of the few applications of machine learning of clear value within seismology. If you would enjoy hearing a questionably informed rant about how machine learning is just rebranded statistics mostly done badly, please do ask in the supervisions.

Relating delay times to lateral variations

Suppose we have obtained the delay time of a seismic phase for a given source-receiver pair. Neglecting for the moment observational errors, this delay time would vanish if the Earth's velocity structure was equal to that of the spherical reference model. In practice, measured delay times are small (e.g. a tens of seconds for a phase whose travel time is tens of minutes), and so we can conclude that the Earth's true structure cannot be too far from the spherical reference model. We know that the travel time of a seismic phase can be calculated using ray theory. But this travel time is related to the velocity structure in a complicated non-linear manner through solution of the Hamiltonian ray equations, and we cannot in general write an explicit formula for T in terms of the velocity, v . Because, however, lateral velocity variations within the Earth seem to be small we can simplify the problem by seeking an *approximate linear relation* between the observed delay time and velocity perturbations within the Earth. Here we need the **functional derivative** of T with respect to v as defined through a first-order Taylor expansion

$$T(v + \delta v) = T(v) + \langle DT(v)|\delta v\rangle + O(\delta v^2). \quad (2)$$

As in our discussion of functional derivative in Lecture 12 in the context of Hamilton's principle, we see that this functional derivative, $DT(v)$, is an object that acts on a velocity perturbation, δv , to produce the first-order change in the travel time. We denote the action of the functional derivative on the velocity perturbation using the inner product notation $\langle \cdot | \cdot \rangle$. We will see below that the interpretation of this action as an inner product is problematic in the present case, but we will not dwell on this technical issue².

The desired functional derivative can be obtained very easily using of **Fermat's principle** for elastic ray theory which you have established in Example sheet 3. Having solved the relevant ray tracing equations, we know that the delay time can be expressed as the following line integral

$$T = \int_{\text{ray}} \frac{1}{v} ds, \quad (3)$$

where s denotes the arc-length along the ray. Fermat's principle states that this travel time is stationary with respect to all first-order perturbations to the ray geometry that leave the end-points fixed. If we perturb the velocity of the earth model to be $v + \delta v$, then the travel time is affected by both (i) a change in the ray geometry, and (ii) the change in velocity along the new ray path. To first-order accuracy the contributions of these effects can be added, but by Fermat's principle the first-order term associated with the change in ray geometry is zero. It follows that the delay time can be written

$$\delta T = - \int_{\text{ray}} \frac{\delta v}{v^2} ds, \quad (4)$$

where this integral is taken along the *unperturbed ray*. To identify the functional derivative here, it seems natural to use the conventional inner product

$$\langle u|w\rangle = \int_M uw d^3\mathbf{x}, \quad (5)$$

for any two real-valued scalar fields defined in the earth model. Looking at eq.(4), it follows that the functional derivative can be suitably defined only if we allow it to be a

² It has an entirely rigorous solution, but one that requires use of functional analysis.

generalised function such that is non-zero only along the reference ray path, while along the ray path it is infinite in such a way that

$$\langle DT(v)|\delta v\rangle = - \int_{\text{ray}} \frac{\delta v}{v^2} ds. \quad (6)$$

This definition should be compared with the familiar one for a Dirac delta function. The singular nature of this functional derivative causes difficulties in some approaches to solving the inverse problem, but these issues are avoided (but at a price) within the most common approach discussed below.

Setting up the inverse problem

We now consider how the inverse problem can be formulated. Our data in such a study comprises delay times for a variety of seismic phases measured from many source-receiver pairs. Typically the number of delay times used will be of order 10^5 - 10^8 . For definiteness, we focus on P-wave phases, meaning the data are only sensitive to variations in the P-wave velocity, α . Near identical methods can be applied to studies based on S-waves, or those that consider both P- and S-wave measurements simultaneously.

At the outset, we shall assume that (i) velocity variations in the Earth are sufficiently small that they can be related linearly to the delay times, and (ii) that the length scale of heterogeneity in the Earth is such that ray theoretic approximations are appropriate. Given these assumptions, each delay time measurement δT_i for $i = 1, \dots, n$ can then be expressed in terms of the P-wave velocity perturbation $\delta\alpha$ through the equation

$$\delta T_i = - \int_{\text{ray}_i} \frac{\delta\alpha}{\alpha^2} ds + e_i, \quad (7)$$

where the integration is taken along the appropriate unperturbed ray path, α denotes the P-wave velocity in the reference model, and we have written e_i for the associated measurement error. We will assume initially that the e_i represent zero-mean, uncorrelated Gaussian errors with known variances σ_i^2 . More complicated statistical models can, of course, be applied but this is not the place for such a discussion.

From the data set $\{\delta T_i\}_{i=1}^n$ we wish to recover the velocity perturbation $\delta\alpha$ which is a (piece-wise) continuous function within the earth model. Clearly a finite amount of data can never be sufficient to recover such a continuous function uniquely, and so we are dealing with a fundamentally under-determined problem. There are a range of methods for dealing with this problem, but we consider one that is both widely used and comparatively simple (the latter point, of course, contributing to the former).

First we assume that the velocity perturbation can be expanded using a finite number of **basis functions**

$$\delta\alpha(\mathbf{x}) = \sum_{j=1}^m \delta\alpha_j \varphi_j(\mathbf{x}). \quad (8)$$

For example we might expand the angular dependence of the velocity perturbation in spherical harmonics up to some finite degree, and then use a set of polynomial functions to describe the radial dependence. Equally, some studies split the earth model into a series of small blocks, assuming a constant velocity in each one. Having expanded the velocity perturbations, the problem is reduced to one of linear algebra that can be solved

using elementary methods. But clearly we are introducing bias into our solutions because we cannot say *a priori* what an appropriate basis should be.

Accepting this strong assumption on the form of the velocity model, we can then write the i th delay time as

$$\delta T_i = \sum_{j=1}^m \left(- \int_{\text{ray}_i} \frac{\varphi_j}{\alpha^2} ds \right) \delta \alpha_j + e_i, \quad (9)$$

which gives a system of linear equations relating the velocity expansion coefficients $\{\delta \alpha_j\}_{j=1}^m$ to the delay time data $\{\delta T_i\}_{i=1}^n$. It will be useful to write this equation as

$$\mathbf{d} = \mathbf{A}\mathbf{m} + \mathbf{e}, \quad (10)$$

where \mathbf{d} is an n -dimensional vector of the measured delay times, \mathbf{m} is an m -dimensional vector of the velocity expansion coefficients, and \mathbf{e} is a vector of the random errors. The elements of the n -by- m matrix \mathbf{A} are given by

$$A_{ij} = - \int_{\text{ray}_i} \frac{\varphi_j}{\alpha^2} ds, \quad (11)$$

and can be determined from knowledge of the reference model and the basis functions.

Non-uniqueness and the null space

Having reduced delay time tomography into a set of linear algebraic equations

$$\mathbf{d} = \mathbf{A}\mathbf{m} + \mathbf{e}, \quad (12)$$

we can start to think how the model \mathbf{m} might be recovered. If we assume that the observational errors are negligible, then we might be tempted to solve this equation using the inverse of \mathbf{A} . However, we recall that \mathbf{A} is an n -by- m matrix, where n is the number of delay time measurements, and m the number of model parameters, and as we will typically have $n \neq m$ the inverse of this matrix cannot be defined. We could, of course, select the number of model parameters to equal to the number of data ($n = m$), but there is no *physical* reason why this should be the case. Indeed, it is better to determine the model parametrisation based on some idea of what the data should reasonably be able to resolve within the Earth. For example, if we are looking at waves of a given dominant frequency, then we can estimate their wavelength in different parts of the Earth using the reference velocity structure. We would not expect to resolve features within the Earth whose size is substantially less than the wavelength of the seismic waves, and can build our model parametrisation accordingly. When this is done it is often the case that $m \gg n$, which is to say that there are many more model parameters than data.

Considering first the idealised case in which there are no data errors, suppose we have found a solution, \mathbf{m} , of the equation

$$\mathbf{d} = \mathbf{A}\mathbf{m}. \quad (13)$$

If there exist a model vector \mathbf{m}_0 such that

$$\mathbf{A}\mathbf{m}_0 = \mathbf{0}, \quad (14)$$

then it is clear that for any scalar a we have

$$\mathbf{A}(\mathbf{m} + a\mathbf{m}_0) = \mathbf{A}\mathbf{m} + a\mathbf{A}\mathbf{m}_0 = \mathbf{A}\mathbf{m} = \mathbf{d}, \quad (15)$$

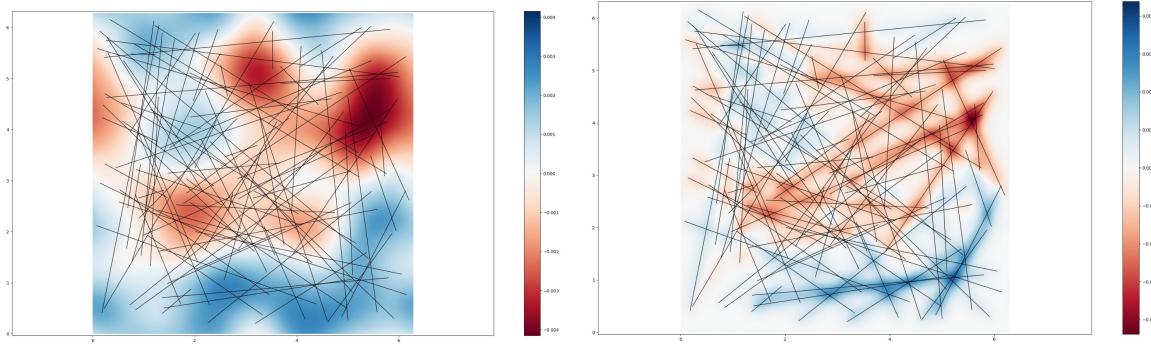


Fig. 2: An example of a simple two-dimensional tomographic problem. There are a number of “earthquakes” and “seismograms”, and the straight lines denote ray paths. The colours plotted denote the velocity perturbation. On the left is the input model, and on the right the inverted model obtained using a certain technique. In regions where there is no ray coverage, the recovered model sets the velocity to zero, this giving a clear example of what is meant by the null space.

and so $\mathbf{m} + a\mathbf{m}_0$ is still an exact solution to the problem. The set of these vectors forms a linear subspace known as the **null space**³ of \mathbf{A} . If the null space is not empty then it is clear that the inverse problem cannot have a unique solution. When there are more model parameters than data then \mathbf{A} will always have a non-empty null space, but even if $n \geq m$ there may still be a non-empty null space due to some of the observations failing to be linearly independent. Almost all inverse problems associated with delay time tomography will have a high-dimensional null space, this largely being due to insufficient ray coverage throughout the Earth (See fig.2).

Regularised least squares solutions

To solve the inverse problem, we might be tempted to apply a least-squares method familiar from statistics, and hence consider the following weighted least squares misfit⁴

$$J(\mathbf{m}) = \frac{1}{2} \sum_{i=1}^n \frac{[(\mathbf{Am})_i - d_i]^2}{\sigma_i^2}, \quad (16)$$

where the factor of one half has been included for later convenience. If we introduce the diagonal covariance matrix $\mathbf{C} = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$ then the misfit above can be written more concisely as

$$J(\mathbf{m}) = \frac{1}{2} (\mathbf{Am} - \mathbf{d})^T \mathbf{C}^{-1} (\mathbf{Am} - \mathbf{d}) \quad (17)$$

This second expression is also valid for a general covariance matrix that accounts for correlations between the different measurement errors. The problem with this approach, however, is that there is not a unique model that minimises this quantity. Indeed, suppose there exists a non-zero vector \mathbf{m}_0 lying in the null space of \mathbf{A} . It is then clear that $J(\mathbf{m} + a\mathbf{m}_0) = J(\mathbf{m})$ for all scalars a .

A non-ideal but pragmatic way around this problem is to seek a model that not only fits the data well, but possesses a property that we see as desirable. For example, we might

³ The term **kernel** is also used in the mathematical literature for this object, and it is my own preference. Null space does have the advantage of being more suggestive of its meaning.

⁴ Some of you might note that this is proportional to the χ^2 -statistic.

think that the true velocity structure must be close to that of the spherical reference model, and hence prefer models that are small in some sense. A common way to measure the “size” of the velocity perturbation through the L^2 -norm

$$\int_M \delta\alpha^2 d^3x. \quad (18)$$

To proceed we substitute our finite-dimensional expansion of the velocity perturbation into this expression to obtain

$$\int_M \delta\alpha^2 d^3x = \sum_{j=1}^m \sum_{k=1}^m \delta\alpha_j \delta\alpha_k \int_M \varphi_j \varphi_k d^3x, \quad (19)$$

which in matrix-vector notation can be written

$$\int_M \delta\alpha^2 d^3x = \mathbf{m}^T \mathbf{B} \mathbf{m}, \quad (20)$$

where the matrix \mathbf{B} has components

$$B_{jk} = \int_M \varphi_j \varphi_k d^3x. \quad (21)$$

Given this result we now consider the modified least squares misfit

$$J(\mathbf{m}) = \frac{1}{2} (\mathbf{A}\mathbf{m} - \mathbf{d})^T \mathbf{C}^{-1} (\mathbf{A}\mathbf{m} - \mathbf{d}) + \frac{\lambda}{2} \mathbf{m}^T \mathbf{B} \mathbf{m}, \quad (22)$$

where λ is a positive constant. The first term on the right hand side measures how well the data is fit, while the second is proportional to the quantity we believe will be small for the true model. The parameter λ provides a means of trading-off between the two terms. By making λ smaller we favour models that fits the data more closely, while larger values favours models compatible with our preconceptions.

Let us show that minimisation of eq.(22) leads to a unique model for each $\lambda > 0$. To do this we will differentiate J with respect to \mathbf{m} , and show that the equation

$$\nabla J(\mathbf{m}) = \mathbf{0}, \quad (23)$$

has a unique solution corresponding to a minimum of J . Using index notation we have

$$J(\mathbf{m}) = \frac{1}{2} (A_{ij} m_j - d_i) C_{ik}^{-1} (A_{kl} m_l - d_k) + \frac{\lambda}{2} m_i B_{ij} m_j, \quad (24)$$

and so

$$\begin{aligned} \frac{\partial J}{\partial m_p} &= \frac{1}{2} (A_{ij} \delta_{jp}) C_{ik}^{-1} (A_{kl} m_l - d_k) \\ &\quad + \frac{1}{2} (A_{ij} m_j - d_i) C_{ik}^{-1} (A_{kl} \delta_{lp}) \\ &\quad + \frac{\lambda}{2} \delta_{ip} B_{ij} m_j + \frac{\lambda}{2} m_i B_{ij} \delta_{jp} \\ &= A_{ip} C_{ik}^{-1} (A_{kj} m_j - d_k) + \lambda B_{pj} m_j, \end{aligned} \quad (25)$$

where we have used the fact that both \mathbf{C} and \mathbf{B} are symmetric. Reverting back to matrix vector notation this result becomes

$$\nabla J(\mathbf{m}) = (\mathbf{A}^T \mathbf{C}^{-1} \mathbf{A} + \lambda \mathbf{B}) \mathbf{m} - \mathbf{A}^T \mathbf{C}^{-1} \mathbf{d}. \quad (26)$$

Setting this derivative equal to zero we arrive at the linear equations

$$(\mathbf{A}^T \mathbf{C}^{-1} \mathbf{A} + \lambda \mathbf{B}) \mathbf{m} = \mathbf{A}^T \mathbf{C}^{-1} \mathbf{d}, \quad (27)$$

which will have a unique solution so long as the $m \times m$ symmetric matrix

$$\mathbf{A}^T \mathbf{C}^{-1} \mathbf{A} + \lambda \mathbf{B} \quad (28)$$

is invertible. That this is the case can be seen by examining the sign of its eigenvalues. To do this we consider the eigenvalue problem

$$(\mathbf{A}^T \mathbf{C}^{-1} \mathbf{A} + \lambda \mathbf{B}) \mathbf{m} = \mu \mathbf{m}, \quad (29)$$

with \mathbf{m} the eigenvector and μ the corresponding eigenvalue. Taking the inner product of this equation with \mathbf{m} we obtain

$$(\mathbf{A}\mathbf{m})^T \mathbf{C}^{-1} (\mathbf{A}\mathbf{m}) + \lambda \mathbf{m}^T \mathbf{B}\mathbf{m} = \mu, \quad (30)$$

where we have assumed that the eigenvector is normalised. The first term on the left hand side is clearly non-negative, while the second is positive by definition. Thus all eigenvalues are positive, and the matrix is invertible. In this manner we have shown that the unique solution of $\nabla J(\mathbf{m}) = \mathbf{0}$ is given by the simple formula

$$\mathbf{m} = (\mathbf{A}^T \mathbf{C}^{-1} \mathbf{A} + \lambda \mathbf{B})^{-1} \mathbf{A}^T \mathbf{C}^{-1} \mathbf{d}. \quad (31)$$

Finally, we note that this model must be a minimum of J . This is because it is a quadratic function that tends to infinity as \mathbf{m} grows in magnitude, and hence its only stationary point must be a minimum.

What we have obtained in eq.(31) is known as a **regularised least squares solution** of the inverse problem. Such a solution is obtained by choosing a regularisation matrix \mathbf{B} corresponding to a quantity we wish to make small, along with the **trade-off parameter** λ which determines how data misfit and model size are balanced. There are a range of different schemes for regularisation including, for example, those that seek as smooth a model as possible. Indeed, a more general least squares function to minimise is

$$J(\mathbf{m}) = \frac{1}{2} (\mathbf{A}\mathbf{m} - \mathbf{d})^T \mathbf{C}^{-1} (\mathbf{A}\mathbf{m} - \mathbf{d}) + \frac{\lambda}{2} (\mathbf{m} - \mathbf{m}_0)^T \mathbf{B}(\mathbf{m} - \mathbf{m}_0), \quad (32)$$

where we can see the regularisation term as asking that the model be close to some preferred value \mathbf{m}_0 , with “close” being measured by the matrix \mathbf{B} . Following the same method as above, the optimal solution in this case can be shown to be

$$\mathbf{m} = (\mathbf{A}^T \mathbf{C}^{-1} \mathbf{A} + \lambda \mathbf{B})^{-1} (\mathbf{A}^T \mathbf{C}^{-1} \mathbf{d} + \lambda \mathbf{B}\mathbf{m}_0). \quad (33)$$

Though such approaches are all rather ad hoc, they are often the only computationally feasible thing that can be done. And we can at least vary some of the arbitrary parameters in the hope of determining what features are reliable in the recovered model.

A Bayesian perspective

A seemingly very different way of thinking about inverse problems is through **Bayes' theorem**. This result states that

$$p(\mathbf{m}|\mathbf{d}) \propto p(\mathbf{d}|\mathbf{m})p(\mathbf{m}). \quad (34)$$

Here $p(\mathbf{m}|\mathbf{d})$ is the probability density function (PDF) of \mathbf{m} given that \mathbf{d} was observed; $p(\mathbf{d}|\mathbf{m})$ is the PDF for observing \mathbf{d} from the model \mathbf{m} , which is also known as the **likelihood**; and finally $p(\mathbf{m})$ is the PDF of \mathbf{m} prior to observing any data. In words, this theorem tells us how to update uncertain knowledge of the model in light of our observations. This result is fully general, and can be applied to non-linear inverse problems with complicated priors. To do so in practice, however, requires random sampling techniques such as Markov Chain Monte Carlo (MCMC) algorithm. Currently this approach possible only when the model space has quite low dimensions (order hundreds or thousands) and the forward calculations are cheap (as is the case with ray theory).

As a special case, suppose we consider a linear inverse problem with Gaussian data errors, and assume a Gaussian prior on the model with mean \mathbf{m}_0 and covariance \mathbf{C}_m . The prior PDF then takes the form

$$p(\mathbf{m}) = \frac{1}{\sqrt{(2\pi)^m \det \mathbf{C}_m}} \exp\left[-\frac{1}{2}(\mathbf{m} - \mathbf{m}_0)^T \mathbf{C}_m^{-1} (\mathbf{m} - \mathbf{m}_0)\right], \quad (35)$$

where we recall that m is the number of basis vectors for the model. Similarly with Gaussian errors, the likelihood is can be written

$$p(\mathbf{d}|\mathbf{m}) = \frac{1}{\sqrt{(2\pi)^n \det \mathbf{C}}} \exp\left[-\frac{1}{2}(\mathbf{A}\mathbf{m} - \mathbf{d})^T \mathbf{C}^{-1} (\mathbf{A}\mathbf{m} - \mathbf{d})\right]. \quad (36)$$

From Bayes theorem we then find

$$p(\mathbf{m}|\mathbf{d}) \propto \exp\left[-\frac{1}{2}(\mathbf{m} - \mathbf{m}_0)^T \mathbf{C}_m^{-1} (\mathbf{m} - \mathbf{m}_0) - \frac{1}{2}(\mathbf{A}\mathbf{m} - \mathbf{d})^T \mathbf{C}^{-1} (\mathbf{A}\mathbf{m} - \mathbf{d})\right], \quad (37)$$

and a little effort shows that the exponentiated term can be factored into the form

$$p(\mathbf{m}|\mathbf{d}) \propto \exp\left[-\frac{1}{2}(\mathbf{m} - \mathbf{m}_p)^T \mathbf{C}_p^{-1} (\mathbf{m} - \mathbf{m}_p)\right], \quad (38)$$

where we have defined the posteriori mean

$$\mathbf{m}_p = (\mathbf{A}^T \mathbf{C}^{-1} \mathbf{A} + \mathbf{C}_m^{-1})^{-1} (\mathbf{A}^T \mathbf{C}^{-1} \mathbf{d} + \mathbf{C}_m^{-1} \mathbf{m}_0), \quad (39)$$

and the posteriori model covariance

$$\mathbf{C}_p = (\mathbf{A}^T \mathbf{C}^{-1} \mathbf{A} + \mathbf{C}_m^{-1})^{-1}. \quad (40)$$

We conclude firstly in this special case the posteriori distribution is Gaussian, and can be determined in closed-form. Moreover, the posteriori mean is precisely that obtained via the regularised least squares method subject to the choice $\lambda \mathbf{B} = \mathbf{C}_m^{-1}$.

In this manner we see that solving a regularised least squares problem returns the mean model corresponding to a certain prior distribution. It is sometimes argued that

this result justifies use of the regularised least squares, while knowledge of the posterior covariance \mathbf{C}_p also allows uncertainties on the model to be determined. Perhaps this is true, but it does depend on the rather strong assumption that the prior distribution on the model is Gaussian. Moreover, the application of Bayes theorem in this manner is not nearly as philosophically straightforward as is sometimes thought⁵.

What you need to know and be able to do

- (i) Why the construction of tomographic models is of interest – it gives the main quantitative constraint on mantle convection and hence on the dynamics of the Earth’s deep interior. Here you should be familiar with the idea that seismic slow regions are likely to be hotter than average, while fast regions will be cooler.
- (ii) How Fermat’s principle is used to obtain a simple linearised relation between delay times and velocity perturbations. From this, you should be able to outline how a tomographic problem is reduced into the standard form $\mathbf{A}\mathbf{m} = \mathbf{d} + \mathbf{e}$.
- (iii) What the null space of a linear(ised) inverse problem is and why it matters.
- (iv) How regularised least squares is used to obtain solutions of linearised tomographic problems. Here you should be able to derive the least squares equations from first-principles.
- (v) How Bayes theorem can be applied in the solution of inverse problems, and that the results obtained coincide with regularised least squares for linear(ised) problems with Gaussian errors and priors.

⁵ For those interested I would strongly recommend the article “Some issues in the foundations of statistics” by David Freeman, this being a clear and largely non-technical summary of the main ideas.