Chapter 5

Solution of the Inverse Problem

Once the travel time differences have been computed as a function of travel distance Δ and latitude λ, it remains to compute the subsurface velocity $u(r, \lambda)$. This process is rather more complicated than it might appear at first glance, and its solution requires some knowledge of inverse problems. The terminology and formalism of this subject are briefly described here.

5.1 The Integral Equation

Equation 2.44 relates the horizontal velocity u of the solar plasma to a particular measured time difference $\delta\tau$, under the assumptions of the ray approximation. The object, then, is to determine the function $u(r, \lambda)$ which best fits the measurements. A naive approach would be to discretize the problem and then find the "best fit" model u in a least-squares sense. In the discussion which follows, I will describe this procedure in slightly more detail, explain why it must fail, and then describe the use of regularization to obtain a stable and reasonable solution.

5.2 The Regularized Least Squares Method

5.2.1 Discretization of the problem

In order to convert the equation 2.44 into a matrix equation which can be solved by a computer, it is simply necessary to model the region of wave propagation with a suitable grid. One approach is to create a grid which is equally spaced in radius and latitude. Assume that there are M_1 latitude gridpoints and M_2 radial gridpoints, where $M = M_1 M_2$. These M model velocities can be thought of as forming a vector u. For a particular measurement $\delta\tau_i$, where the index i denotes the coordinates (λ, Δ) which determine the ray path, the relationship between the model and the data can be expressed as

$$\delta\tau_i = \sum_{j=1}^{M} K_{ij} u_j. \tag{5.1}$$

The *sensitivity kernel* K_{ij} has elements which are given by

$$K_{ij} = 2 \int_i \frac{v_{gh}}{v_{gr} c^2} dr_j, \tag{5.2}$$

where the integral is along the particular element of the ray path determined by i which lies within grid cell j. In other words, the velocity u is assumed to be constant within each cell of the grid.

The choice of the grid spacing is intimately related to the question of the resolution of the solution, which is discussed in section 5.2.4. However, the radial coordinate deserves special mention. The simplest choice is to make a grid which is equally spaced in radius as well as latitude. However, analysis of the inversion results has shown that this offers a somewhat misleading view of the resolution of the solution deep in the convection zone. A better choice is to create a grid which has equal spacing in *acoustic radius*, defined[1] as

$$r_\tau \equiv \int_0^r \frac{dr'}{c}. \tag{5.3}$$

[1]Here I use the symbol r_τ rather than the standard τ, so as not to become confused with travel time.

In this case the grid points are equally spaced in travel time, assuming radially propagating waves and ignoring dispersion.

A typical sensitivity kernel is shown in figure 5.1. Since each measurement $\delta\tau_i$ actually is made from the average of a large number of cross correlations, the kernel is not actually constructed for a single ray path but for the weighted mean of all of the ray paths which have contributed to the average. The weight for each ray path is proportional to the number of pairs of pixels used which had exactly the same distance, latitude, and direction (see also section 4.3.2).

Once the problem has been discretized, the equation 5.1 can be recast in matrix form,

$$Ku = z, \tag{5.4}$$

where the N elements of the vector z are $z_i = \delta\tau_i$, the vector u has the M model velocities for components, and each row of the $N \times M$ matrix K is a sensitivity kernel with components given by equation 5.2. If we define two new quantities A and b,

$$b_i = z_i/\sigma_i, \qquad i = (1, \ldots, N) \tag{5.5}$$

$$A_{ij} = K_{ij}/\sigma_i, \qquad j = (1, \ldots, M) \tag{5.6}$$

then we want to find the solution $\hat{u}$ which minimizes

$$\chi^2 \equiv |Au - b|^2 \tag{5.7}$$

with respect to u. This is equivalent to solving the set of linear equations (see for example Press *et al.* (1992))

$$A^T Au = A^T b. \tag{5.8}$$

In practice, however, this will not give us a useful answer for $\hat{u}$, if it gives an answer at all. The problem here is that since the operator K is an integral operator it will act to smooth the response of the data to the model. This "smoothed out" information cannot be recovered by solving the set of linear equations in equation 5.8. In addition, since the operator K is a smoothing operator, the inverse operator (if it exists), when operating on the noisy measurements, will cause the solution to be unstable. Finally,

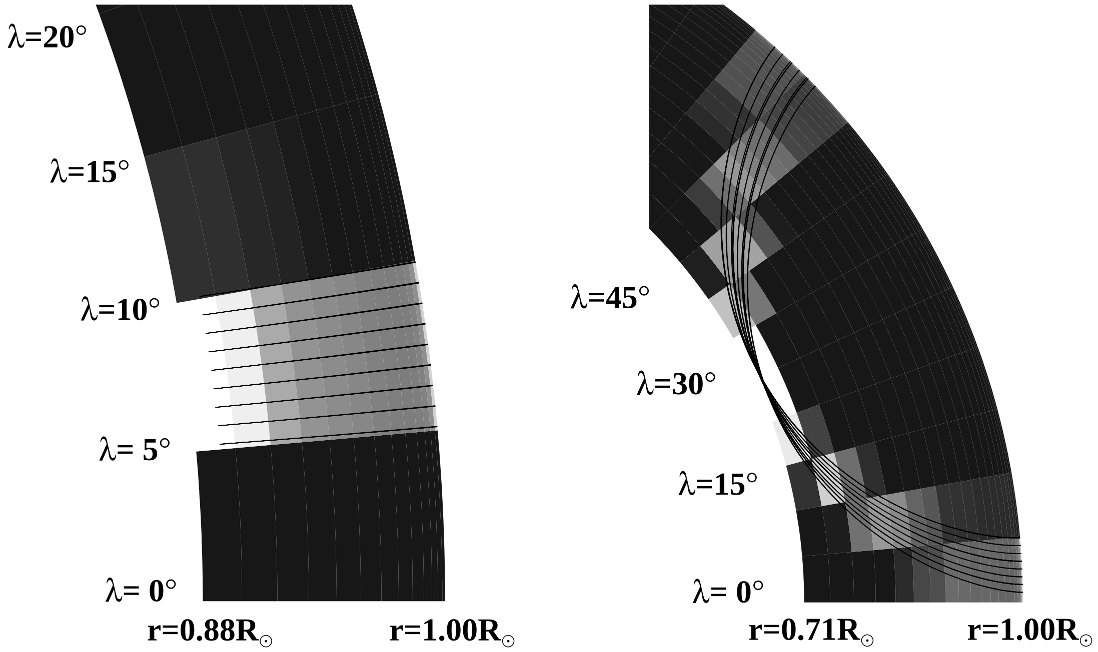

Figure 5.1: The left-hand figure shows the sensitivity kernel for the zonal velocity, for the measurement with $\lambda = 7.5°$, $\Delta = 18.1°$. The right-hand figure shows the sensitivity to the meridional velocity for the measurement with $\lambda = 24.4°$, $\Delta = 44.1°$. The greyscale shows the sensitivity to horizontal flows, with black areas being those with no sensitivity. The black curves depict ray paths for the individual measurements which were averaged together to get the single kernel shown here; that is, each black curve represents a different pair of MDI pixels. There are 52 rays in the left-hand figure, and 39 on the right, although many of the curves overlap. In the left-hand figure, note that the rays do not propagate at constant latitude.

the problem may be *underdetermined* — we may attempt to solve for more model parameters than there are measurements. This is, strictly speaking, mathematically impossible, but can be accomplished if some *a priori* knowledge can be used to select the proper solution from the set of possible solutions given by the measurements.

5.2.2 Regularization

A common and very useful approach to this problem is to add an additional constraint which causes the chosen solution to be "smooth" in some way. For example, rather than minimize χ^2 alone, we can try to find the solution $\hat{u}$ which minimizes the quantity

$$|Au - b|^2 + \gamma u^T H u, \tag{5.9}$$

where the second term represents some sort of smoothing operation. As examples, four forms have been used for the matrix H in this work:

$$\begin{aligned}
u^T H_0 u &\Longrightarrow \int\int |u|^2 r\, dr\, d\theta && (5.10)\\
u^T H_1 u &\Longrightarrow \int\int |\nabla u|^2 r\, dr\, d\theta && (5.11)\\
u^T H_{u/r} u &\Longrightarrow \int\int \left|\nabla \frac{u}{r}\right|^2 r\, dr\, d\theta && (5.12)\\
u^T H_\Omega u &\Longrightarrow \int\int \left|\nabla \frac{u}{r\cos\lambda}\right|^2 r\, dr\, d\theta && (5.13)
\end{aligned}$$

where the matrix operations on the left are discrete approximations of the integrals on the right. The chosen matrix H is called the *regularization operator*, and adding it to the problem constrains the solution to be smooth in some sense.

Minimizing the quantity 5.9 is equivalent to solving the set of linear equations

$$(A^T A + \gamma H)u = A^T b. \tag{5.14}$$

The solution $u = \hat{u}$ obviously depends not only on the data z, the errors σ, and the model K, but also on the regularization scheme chosen (choice of H). The relative influence of the regularization is controlled by the free parameter γ in equation 5.14, which is called the *regularization trade-off parameter*.

5.2.3 Trade-off parameter

When γ is large, the influence of H is great compared to the influence of the model A, and the solution obtained, although very smooth, probably will not be a good fit to the data. On the other hand, when γ is small, the influence of the regularization term is relatively small. The solutions in this case can provide an arbitrarily good fit to the (noisy) data, but will have unreasonably large magnitudes and derivatives.

In general, model solutions $\hat{u}$ are usually computed for a range of different values of γ, and then the "best" model in some sense is chosen from the set of solutions. At the two extremes, it is easy to rule out the solutions obtained, but the determination of the "best" solution is often a matter of art.

One way to look at the effect of the trade-off parameter in a somewhat quantitative way is to plot the magnitudes of the two terms in the quantity 5.9. Such a diagram is shown in figure 6.1 and is known as an *L-curve*. The trade-off parameter is increasing in magnitude from right to left, so the smoothness of the solution is increasing. The bend in the L-curve is an optimal solution for this particular choice of regularization; it is the smoothest solution possible with a "small" value of χ^2. If the errors σ in the measurements can be trusted, then we can hope that the bend in the L-curve will be located near a solution with $\chi^2 = N$, the number of measurements.

5.2.4 Spatial averaging kernels

Another way to think about the effect of γ is to examine the interplay between the stability and the resolution of the inversion solution. For the sake of a convenient notation, define the matrix D as

$$D \equiv A^T A + \gamma H. \tag{5.15}$$

We can therefore express the solution $\hat{u}$ to equation 5.14 as

$$\hat{u} = D^{-1}(A^T b). \tag{5.16}$$

Substituting for b from equation 5.8,

$$\hat{u} = D^{-1} A^T A u. \tag{5.17}$$

Equation 5.17 shows that each parameter of the model solution $\hat{u}$ can be expressed as a linear combination of the parameters of the "real" velocity u. For this reason, each row of the $M \times M$ matrix $D^{-1}A^T A$ is called an *averaging kernel* for the solution. For the particular problem under consideration here, each element of the vector u corresponds to a particular latitude and radius, so the kernel gives a representation of the spatial resolution of the solution.

5.2.5 Error analysis

As in the case of all least squares fitting techniques, the uncertainties in the model parameters are determined by the uncertainty in the measurements and by the characteristics of the kernels K. In both the regularized and standard forms, the error information is contained in the covariance matrix of the solution. In the case of regularized least squares, however, the formulation is slightly more complicated.

The variance in the model velocity $\hat{u}_j$ is defined as

$$\sigma^2(\hat{u}_j) = \sum_{i=1}^{N} \sigma_i^2 \left(\frac{\partial \hat{u}_j}{\partial z_i} \right)^2 \tag{5.18}$$

where σ_i^2 is the variance in measurement z_i. Writing equation 5.16 in component form,

$$\hat{u}_j = \sum_{k=1}^{M} [d_{jk}]^{-1} \sum_{i=1}^{N} \frac{z_i K_{ki}}{\sigma_i^2} \tag{5.19}$$

so that

$$\frac{\partial \hat{u}_j}{\partial z_l} = \sum_{k=1}^{M} [d_{jk}]^{-1} \frac{K_{kl}}{\sigma_l^2}. \tag{5.20}$$

Here I have represented the elements of D^{-1} symbolically as $[d_{ij}]^{-1}$. Inserting this result into definition 5.18 gives

$$\sigma^2(u_j) = \sum_{k=1}^{M}\sum_{l=1}^{M}[d_{jk}]^{-1}[d_{jl}]^{-1}[A^TA]_{kl}, \tag{5.21}$$

or in matrix form,

$$\sigma^2(u_j) = [D^{-1}A^TA(D^{-1})^T]_{jj}. \tag{5.22}$$

Note that in the case with no regularization ($\gamma = 0$) the variance reduces to $[D^{-1}]_{jj}$ as we expect.

Equation 5.22 shows that the *error propagation* — the way that uncertainty in the measurements is transformed into uncertainty in the velocities — depends not only on the model, but also on the choice of regularization parameter γ and the regularization operator H in equation 5.15.

In fact, another way to visualize the effect of the tradeoff parameter γ is to imagine the balance between the error propagation and the resolution. With little regularization, the averaging kernels are well localized, but the solution is very sensitive to errors in the measurements; when γ is large, the solution is stable with respect to errors but the resolution is very poor.

5.3 Additional Constraints

The previous section describes how *regularization* is used to select a “good” solution from the infinity of models which might be compatible with a set of noisy measurements. However, the use of such a selection mechanism is not really based on physics, but on some *a priori* ideas about what the velocity field should look like.

In addition to some type of smoothness, it would also be desirable to select a model solution where the velocity field $\mathbf{u}$ satisfies the continuity equation 2.40. In the case of the large-scale, steady flows measured here, this reduces to

$$\nabla \cdot (\rho(r)\,\mathbf{u}) = 0. \tag{5.23}$$

Unfortunately, as noted in section 2.3.3, the radial component of the meridional circulation is essentially undetectable in the time-distance technique, so that it is not possible to constrain the flow field by applying equation 5.23 at each grid point. However, there is one important case where the conservation of mass can be applied to constrain the solution.

If the model is allowed to extend all the way to the bottom of the convection zone, and if the meridional circulation is assumed not to penetrate into the radiative zone, then the total amount of mass flowing northward must be equal to the total amount flowing southward. In fact, it is possible to state the constraint in stronger terms: for every latitude, the net amount of mass crossing that latitude must be zero.

The method used in this work to impose this condition is a so-called "barrier method." The condition of mass conservation is cast in matrix form by creating a matrix C which satisfies

$$u^T C u = \sum_{i=1}^{M_1} \left(\sum_{j=1}^{M_2} \rho_j U_{ji} \, dr_j \right)^2 = 0. \tag{5.24}$$

Here the velocity vector u is treated symbolically as a $(M_2 \times M_1)$ matrix U inside the sum; the indices j and i indicate radius and latitude, respectively.

Once the matrix C has been formed, the condition $u^T C u = 0$ is satisfied in an approximate sense by solving a modified version of equation 5.14:

$$(A^T A + \gamma H + \beta C) u = A^T b. \tag{5.25}$$

Thus, the condition that mass be conserved is treated as an extra regularization condition; the deviation from perfect continuity can be made arbitrarily small by making β arbitrarily large.

Solving inverse problems is an art and a science in itself; the methods described here have been used as a starting point. No doubt the methods of inversion, and the approximations used in the modeling, will be greatly improved as time-distance helioseismology becomes more mature. The next chapter will describe the numerous results which can be obtained using this relatively simple approach.