Chapter 9

Helioseismic inversion

The expression (8.41) for the splitting caused by spherically symmetric rotation is a particularly simple example of the relation between the observable properties of the oscillation frequencies, and the properties of the solar interior which we wish to determine. The determination of $\Omega(r)$ from the $\delta\omega_{nlm}$ constitutes the simplest example of an *inverse problem*. In particular, there is a linear relation between the observables and the property of the solar interior. In contrast, the oscillation frequencies depend in a non-linear fashion on the structure of the Sun, as specified by for example $\rho(r)$ and $c(r)$ (*cf.* Section 5.1). However, by assuming that the real solar structure can be obtained from the structure of a given reference model by applying small corrections, the differences in frequency between the observations and the reference model can be obtained from a linear perturbation analysis of the oscillation equations, resulting, once more, in a linear relation between the frequency differences and the corrections to the model; this was discussed in some detail in Section 5.5.3 (see eq. 5.84). Thus the linear inverse problem forms the basis for much of the inverse theory for solar oscillations.

Inverse problems have a vast literature, covering their application in, for example, geophysics and radiation theory (*e.g.* Parker 1977; Craig & Brown 1986; Deepak 1977; Tarantola 1987). The application to the solar inverse problem was discussed by Gough (1985), and Christensen-Dalsgaard, Schou & Thompson (1990) made a systematic comparison of different inversion techniques, as applied to the problem of spherically symmetric rotation. The following discussion is to a large extent based on their results.

9.1 Inversion of the rotational splitting

Given the simplicity of the rotational-splitting inverse problem, it serves as a very useful prototype of more general inversions. Furthermore, a determination of the solar internal rotation is of great intrinsic interest. For these reasons, in this section I consider rotational inversion in some detail.

9.1.1 One-dimensional rotational inversion

I first consider the inversion for a spherically symmetric rotation rate $\Omega(r)$. Thus the data are

$$\Delta_i = \int_0^R K_i(r)\Omega(r)\mathrm{d}r + \epsilon_i \ , \qquad i = 1, \ldots, M \ , \tag{9.1}$$

where, for notational simplicity, I represent the pair (n, l) by the single index i; M is the number of modes in the data set considered. Δ_i is the scaled rotational splitting $m^{-1}\beta_{nl}^{-1}\delta\omega_{nlm}$, the kernels K_i having been normalized as in equation (8.44), and ϵ_i is the observational error in Δ_i. The goal of the inversion is to determine an approximation $\bar{\Omega}(r_0)$ to the true angular velocity, as a function of position r_0 in the Sun; obviously this is only possible for those parts of the Sun about which the oscillations provide data. In most cases considered so far, the inversion is carried out through linear operations on the data. Hence $\bar{\Omega}$ is linearly related to the data: for each r_0 there exists a set of *inversion coefficients* $c_i(r_0)$ such that

$$\bar{\Omega}(r_0) = \sum_i c_i(r_0)\Delta_i \ . \tag{9.2}$$

It follows from equation (9.1) that this may be written as

$$\bar{\Omega}(r_0) = \int_0^R \mathcal{K}(r_0, r)\Omega(r)\mathrm{d}r \ , \tag{9.3}$$

where the *averaging kernel* $\mathcal{K}(r_0, r)$ is given by

$$\mathcal{K}(r_0, r) = \sum_i c_i(r_0)K_i(r) \ . \tag{9.4}$$

The inversion coefficients and averaging kernels clearly depend on the choice of inversion method, and of possible parameters that enter into the method; indeed, the inversion may be thought of as a way to determine coefficients and averaging kernels such as to obtain as much information about the angular velocity as possible. As discussed below, the determination of the coefficients must take into account the estimated errors in the data. On the other hand, once the method and parameters have been chosen, the coefficients and averaging kernels are independent of the data values. Hence they can be used to make a data-independent comparison of different inversion methods; this was the approach taken by Christensen-Dalsgaard *et al.* (1990).

The averaging kernels provide an indication of the resolution of the inversion; it is clearly desirable to achieve averaging kernels that are sharply peaked around $r = r_0$, and with small amplitude far away from that point. As a quantitative measure of resolution it is common to use a width of $\mathcal{K}(r_0, r)$ obtained as the distance between the quartile points, defined such that one quarter of the area of $\mathcal{K}(r_0, r)$, regarded as function of r, lies below the lower, and one quarter of the area above the upper, quartile point. The inversion coefficients give information about the propagation of errors from the data to the solution $\bar{\Omega}(r_0)$. If the standard error on Δ_i is $\sigma(\Delta_i)$, the standard error in the result of the inversion is given by

$$\sigma[\bar{\Omega}(r_0)]^2 = \sum_i c_i(r_0)^2\sigma(\Delta_i)^2 \ . \tag{9.5}$$

(The generalization to a non-diagonal covariance matrix is fairly obvious.) In particular, if (somewhat unrealistically) $\sigma(\Delta_i) = \sigma$ is assumed to be the same for all the observed modes

$$\sigma[\bar{\Omega}(r_0)] = \Lambda(r_0)\sigma \ , \tag{9.6}$$

where I introduced the *error magnification*

$$\Lambda(r_0) = \left[\sum_i c_i(r_0)^2\right]^{1/2} . \tag{9.7}$$

The optimization of the inversion techniques is often based on a trade-off between width of the averaging kernels and error or error magnification magnification (*cf.* Figure 9.2 below).

A procedure which is based explicitly on the determination of the inversion coefficients is the technique of *optimally localized averages*, developed by Backus and Gilbert (1970); this has been used extensively for helioseismic inversion. The goal is to choose the coefficients $c_i(r_0)$ such as to make $\mathcal{K}(r_0, r)$ approximate as far as possible a delta function $\delta(r - r_0)$ centred on r_0; then $\bar{\Omega}(r_0)$ provides an approximation to $\Omega(r_0)$. This is achieved by determining the coefficients $c(r_0)$ such as to minimize

$$\int_0^R (r - r_0)^2 \mathcal{K}(r_0, r)^2 \mathrm{d}r + \mu \sum_{ij} E_{ij} c_i c_j , \tag{9.8}$$

subject to the constraint

$$\int_0^R \mathcal{K}(r_0, r)\mathrm{d}r = 1 ; \tag{9.9}$$

here E_{ij} is the covariance matrix of the data. This is equivalent to solving the set of linear equations

$$\sum_j W_{ij} c_j = b , \tag{9.10}$$

and

$$\sum_j c_j = 1 , \tag{9.11}$$

where b is a Lagrange multiplier. Here

$$W_{ij} = S_{ij} + \mu E_{ij} , \tag{9.12}$$

where

$$S_{ij} = \int_0^R (r - r_0)^2 K_i(r) K_j(r) \mathrm{d}r . \tag{9.13}$$

Furthermore, μ is a parameter which, as discussed below, must be adjusted to optimize the result.

The effect of the minimization is most easily understood for $\mu = 0$. Minimizing equation (9.8) subject to equation (9.9) ensures that $\mathcal{K}(r_0, r)$ is large close to r_0, where the weight function $(r - r_0)^2$ is small, and small elsewhere. This is precisely the required "delta-ness" of the combined kernel. However, with no further constraints, the optimization of the combined kernel may result in numerically large coefficients of opposite sign. Hence, the variance in $\bar{\Omega}$, which can be estimated as

$$\sigma^2(\bar{\Omega}) = \sum_{ij} E_{ij} c_i c_j , \tag{9.14}$$

would be large. The effect of the second term in equation (9.8), when $\mu > 0$, is to restrict $\sigma^2(\bar{\Omega})$. The size of μ determines the relative importance of the localization and the size of

the variance in the result. Hence, μ must be determined to ensure a trade-off between the localization and the error, measured by the width of $\mathcal{K}(r_0, r)$ and $\Lambda(r_0)$, respectively; μ is generally known as *the trade-off parameter*.

The principal difficulty of this method is computational expense: at each target radius r_0 it involves the solution of a set of linear equations whose order is the number of data points. Jeffrey (1988) proposed an alternative version where the coefficients were determined by minimizing the difference between $\mathcal{K}(r_0, r)$ and the delta function $\delta(r - r_0)$. This is computationally more efficient, in that only one matrix inversion is required, but results in averaging kernels with somewhat undesirable properties.

More recently Pijpers & Thompson (1992, 1994) have developed this method further, by matching $\mathcal{K}(r_0, r)$ instead to a prescribed *target function* $\mathcal{T}(r_0, r)$ which more closely matches the behaviour that can be achieved with the given mode set. They dubbed this the *SOLA* technique (for Subtractive Optimally Localized Averaging), to distinguish it from the *MOLA* technique (for Multiplicative Optimally Localized Averaging) discussed above. Specifically, the coefficients $c_i(r_0)$ are determined by minimizing

$$\int_0^R [\mathcal{K}(r_0, r) - \mathcal{T}(r_0, r)]^2 \, \mathrm{d}r + \mu \sum_{ij} E_{ij} c_i c_j \ , \tag{9.15}$$

where again μ is a tradeoff parameter. In addition, the width of $\mathcal{T}(r_0, r)$ functions as a parameter, in most cases depending on r_0, of the method. As before, the inclusion of the last term in equation (9.15) serves to limit the error in the solution. The minimization leads to the following system of linear equations for the $c_i(r_0)$:

$$\sum_j (K_{ij} + \mu E_{ij}) c_j(r_0) = T_i(r_0) \ ; \tag{9.16}$$

here

$$K_{ij} = \int_0^R K_i(r) K_j(r) \mathrm{d}r \ , \tag{9.17}$$

and

$$T_i(r_0) = \int_0^R \mathcal{T}(r_0, r) K_i(r) \mathrm{d}r \ . \tag{9.18}$$

In equation (9.16) the coefficient matrix on the left-hand side is independent of r_0. Thus it can be inverted or, more efficiently, suitably factored, once and for all; after this the determination of the coefficients at each target point r_0 is virtually free. Compared with the MOLA technique the computational effort is therefore reduced by roughly a factor given by the number of target locations. An additional advantage of the technique is the ability to choose the target function such as to tailor the averaging kernels to have specific properties. In addition to the usual trade-off parameter μ controlling the weight given to the errors, the method obviously depends on parameters controlling the properties of the target functions $\mathcal{T}(r_0, r)$. These are often taken to be of gaussian shape; it was argued by Thompson (1993) that the radial resolution, for inversion of acoustic modes, is proportional to the sound speed c, and hence the width of $\mathcal{T}(r_0, r)$ is generally taken to be proportional to $c(r_0)$, the constant of proportionality serving as a parameter characterizing the targets.

A second commonly used technique is the regularized least-squares, or Tikhonov, method (see, for example, Craig and Brown 1986). Here the solution $\bar{\Omega}(r)$ is parameterized, often as a piecewise constant function on a grid $r_0 < r_1 < \ldots < r_N$, with $\bar{\Omega}(r) = \Omega_j$ on the

interval $[r_{j-1}, r_j]$; the parameters Ω_j are determined through a least-squares fit to the data. In general, this least-squares procedure needs to be regularized to obtain a smooth solution. This is achieved by including in the minimization a term which restricts the square of $\bar{\Omega}$, or the square of its first or second derivative. Thus, for example one may minimize

$$\sum_i \left[\int_0^R K_i(r)\bar{\Omega}(r)\mathrm{d}r - \Delta_i \right]^2 + \mu^2 \int_0^R \left(\frac{\mathrm{d}^2\bar{\Omega}}{\mathrm{d}r^2} \right)^2 \mathrm{d}r \ , \tag{9.19}$$

where in the last term a suitable discretized approximation to $\mathrm{d}^2\bar{\Omega}/\mathrm{d}r^2$, in terms of the Ω_j, is used. The minimization of equation (9.19) clearly leads to a set of linear equations for $\bar{\Omega}_j$, defining the solution; however, it is still the case that the procedure can be formulated as in equation (9.2) and hence leads to the determination of inversion coefficients and averaging kernels. By restricting the second derivative the last term in equation (9.19) suppresses rapid oscillations in the solution, and hence ensures that it is smooth; the weight μ^2 given to this term serves as a trade-off parameter, determining the balance between resolution and error for this method.

The asymptotic expression (8.48) for the frequency splitting provides the basis for a final example of an inversion method in widespread use. The right-hand side of that equation is a function $\mathcal{H}(\omega/L)$ which is in principle determined by the observed splittings after scaling with S. Given $\mathcal{H}(w)$, the angular velocity can be obtained from

$$\Omega(r) = -\frac{2a}{\pi} \frac{\mathrm{d}}{\mathrm{d}\ln r} \int_{a_s}^a (a^2 - w^2)^{-1/2} \mathcal{H}(w)\mathrm{d}w \ , \tag{9.20}$$

where $a = c/r$ and $a_s = a(R)$. This is entirely equivalent to equation (7.153) determining the sound-speed difference from the function $\mathcal{H}_1(\omega/L)$ fitted to the scaled frequency differences. In practice, since the asymptotic expressions are not exact, the scaled splittings are not precisely functions of ω/L. Hence, an approximation to $\mathcal{H}(\omega/L)$ is obtained by making a least-squares fit to $S\Delta\omega_{nl}$ of a function of that form, for example by representing it as a spline over a suitably chosen set of knots. The number of knots determines the resolution achieved in representing $\mathcal{H}(w)$ and hence in the inferred solution $\bar{\Omega}(r)$; therefore, in this case the number of knots serves as trade-off parameter. Again, the processes of carrying out the spline fit to the scaled data and evaluating the integral in equation (9.20) are linear, and hence the method allows the evaluation of inversion coefficients and averaging kernels (see Christensen-Dalsgaard *et al.* 1990 for details).

An illustration of the use of these methods is provided by the results obtained by Christensen-Dalsgaard *et al.* (1990). They considered a set consisting of about 830 modes at selected degrees between 1 and 200, and frequencies between 2000 and 4000 μHz; for simplicity, the standard errors were assumed to be the same for all modes. Figure 9.1 shows examples of averaging kernels $\mathcal{K}(r_0, r)$ for the MOLA, regularized least-squares and asymptotic methods. The trade-off parameters were chosen such that the error magnification at $r_0 = 0.5R$ was close to 1 in all three cases. It should be realized that the kernels entering into the combination are of the form shown in Figure 8.2. Thus, a very large degree of cancellation has been achieved of the dominant contribution from near the surface. Nevertheless, it is obvious that the averaging kernels are only approximate realizations of delta functions; structure on a scale smaller than roughly $0.05R$ is not resolved.

This limitation is inherent in any inversion method. Indeed, it is evident that from a finite set of data one can never completely resolve the function $\Omega(r)$. To obtain a definite

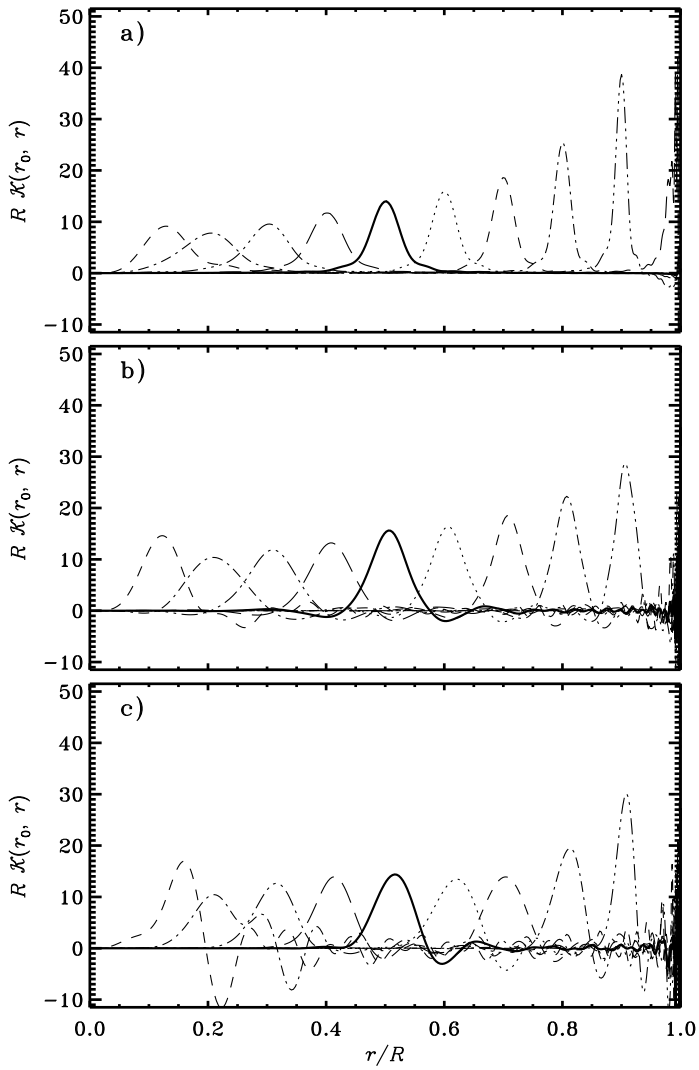

Figure 9.1: Averaging kernels $\mathcal{K}(r_0, r)$ at selected radii ($r_0/R = 0.1, 0.2, \ldots, 1.0$) for inversion by means of the MOLA technique (panel a), the Tikhonov inversion with second-derivative smoothing (panel b) and asymptotic inversion (panel c). The parameters in each inversion method have been chosen to obtain approximately the same error magnification for $r_0 = 0.5R$. In each case, the kernel at $r_0 = 0.5R$ is shown as a bolder curve. From Christensen-Dalsgaard *et al.* (1990).

solution additional constraints must be invoked. The constraints used here essentially demand that the solution be smooth. This is ensured in the method of optimally localized kernels by representing the solution by smooth averaging kernels whose shape is determined by the minimization in equation (9.8). For the Tikhonov method smoothness is explicitly

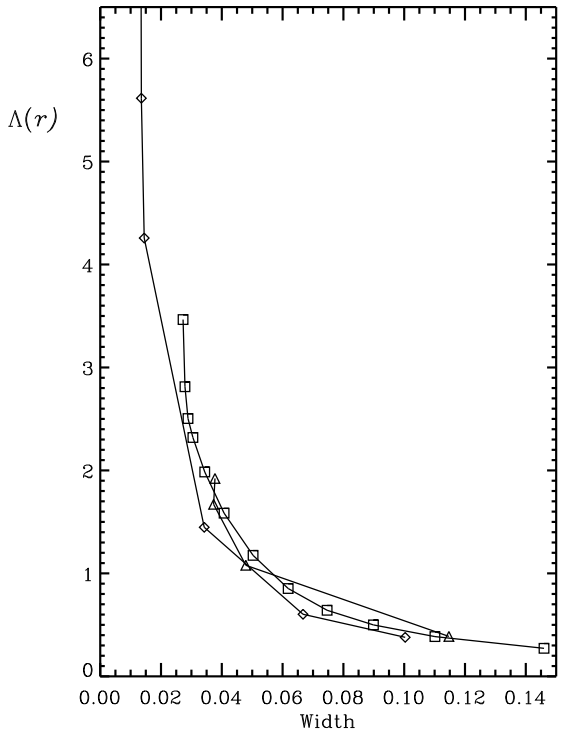

Figure 9.2: The trade-off between error magnification $\Lambda(r_0)$ (*cf.* eq. 9.7) and width (defined as distance between quartile points), for inversion at $r_0 = 0.5R$. Results are shown for the MOLA technique (□), Tikhonov inversion (◇) and asymptotic inversion (△), in each case varying the relevant trade-off parameter over a wide range. From Christensen-Dalsgaard *et al.* (1990).

demanded by constraining the second derivative of the solution, whereas in the case of the asymptotic technique the constraints lie partly in using the asymptotic description, which in itself assumes that the solution varies on a scale larger than the wavelength of the modes, partly in the spline fit to the scaled data.

All methods contain trade-off parameters which determine the relative weight given to the demands of resolution on the one hand, and smoothness or minimizing errors on the other. To illustrate this balance, it is common to consider trade-off diagrams, where a measure of error is plotted against a measure of the width of the averaging kernels. An example is shown in Figure 9.2. The similarity, in terms of such global measures, between the three conceptually rather different methods is quite striking.

It is also of interest to consider in detail the way in which the different methods utilize the data, as expressed in terms of the inversion coefficients. In the case of the asymptotic technique it may be shown that these depend on ω/L alone; hence for the purpose of comparison it is sensible to plot the coefficients as a function of ω/L in all cases. In Figure 9.3 the coefficients for the optimally localized averages and the Tikhonov inversions are compared with those obtained with the asymptotic technique, for $r_0 = 0.5R$. It is evident that

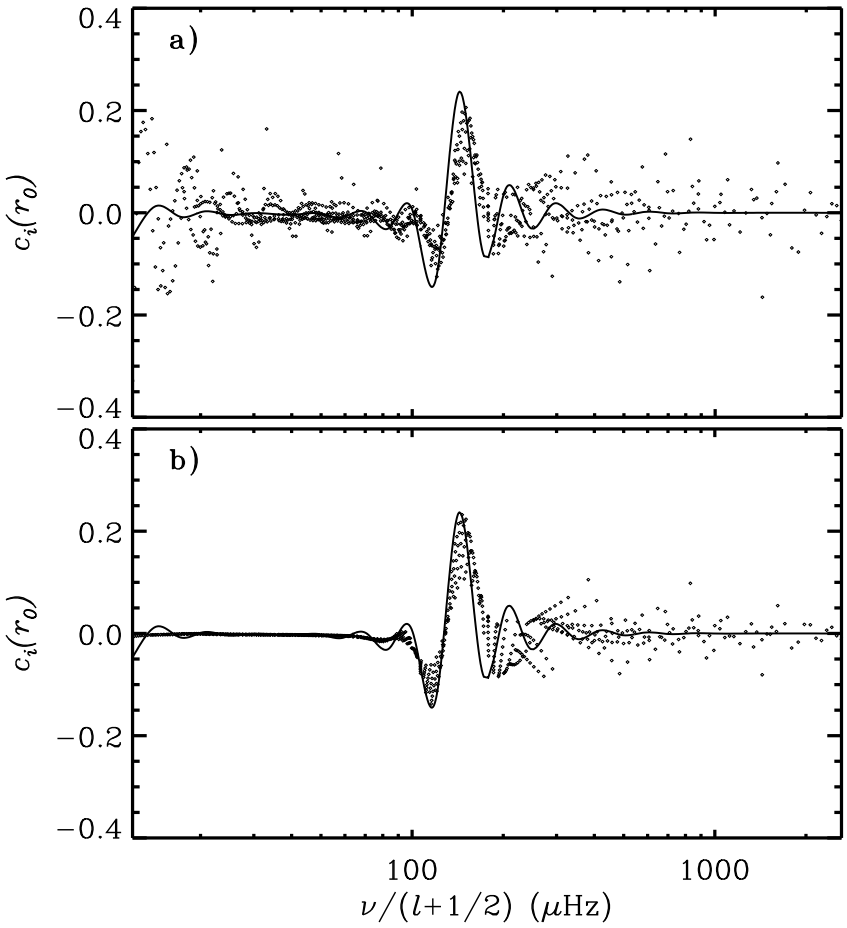

Figure 9.3: Inversion coefficients $c_i(r_0)$ for the inversions illustrated in Figure 9.1, at $r_0 = 0.5R$; they have been plotted against $\nu/(l+1/2)$ which, according to equation (5.28), measures the depth of penetration of the modes. The continuous line in both panels shows coefficients for the asymptotic method which are functions of $\nu/(l+1/2)$. Panel (a) gives the coefficients obtained with the MOLA technique, whereas panel (b) shows coefficients for Tikhonov inversion. From Christensen-Dalsgaard *et al.* (1990).

the overall behaviour of the coefficients is quite similar in all three cases; in particular, the modes dominating the inversion are those whose turning point r_t is in the vicinity of the target radius r_0. At a more subtle level, there are significant differences. In particular, unlike the Tikhonov case, inversion by optimally localized averages makes substantial use of modes of high degree. It may be shown that these are used essentially only to improve the averaging kernels near the surface; in fact, as can be seen in Figure 9.1 the $\mathcal{K}(r_0, r)$

for Tikhonov inversion have quite substantial amplitude at very short wavelength near the surface, whereas such structure is entirely suppressed by the minimization in equation (9.8) for the optimally localized averages. Christensen-Dalsgaard, Hansen & Thompson (1993) found that considerable insight into such details of the behaviour of the inversion methods can be obtained by analyzing the Tikhonov case by means of the so-called Generalized Singular Value Decomposition; this provides a natural basis for expressing the inverse problem which can then be used to study the properties of other inversion methods.

Graphs such as Figure 9.2 are very useful for the choice of the trade-off parameter; however, it is evident that this choice depends critically on the particular application, including the errors in the data. The question of how to fix the trade-off parameter, or indeed even how to choose the inversion method, has given rise to a great deal of debate, occasionally of an almost philosophical (or, dare one say, religious) nature. It has been suggested that inversion methods should be chosen which aim at fitting the data; this makes the method of optimally localized averages, whose goal is instead to design the averaging kernels, less attractive. Also, a great deal of emphasis has been placed on techniques for objectively determining the trade-off parameters, based on the errors in the data and possibly the properties of the solution. In contrast, the approach taken in helioseismology has to a large extent been pragmatic: in fact, it can be argued that since no method, or choice of trade-off, can provide the exact solution given the necessarily incomplete data, the most important aspect of the inversion is to be able to interpret the result and its significance. In this respect, the averaging kernels which graphically illustrate the resolution, and the inversion coefficients which allow evaluation of effects of errors in the data, are clearly very useful. By choosing different inversion methods, and different values of the trade-off parameters, considering in each case the properties of the resulting inversion, one can hope to obtain a more complete impression of the underlying solution. In this process prior knowledge, or prejudices, about the solution clearly play a significant role; these should ideally be formulated in a well-defined statistical sense, but probably often are not.

9.1.2 Two-dimensional rotational inversion

So far, I have considered inversion for a function that depends on r alone. It is evidently desirable, however, to carry out inversion for more general properties which are functions both of r and θ. Here I concentrate on the case of determining the angular velocity $\Omega(r,\theta)$; it should be noticed, however, that another interesting inverse problem concerns the departure of the structure from spherical symmetry caused, for example, by a latitude dependence of the energy transport in the convection zone.

As is evident from Chapter 8, the latitude dependence of rotation is reflected in the dependence of the frequencies on azimuthal order m. Thus in general individual frequencies ω_{nlm} have to be analyzed. This greatly increases the amount of data to be considered, compared with the simple case discussed in Section 9.1.1, and computational efficiency becomes a crucial consideration. However, it is still generally the case that the inferred angular velocity $\bar{\Omega}(r_0,\theta_0)$ at some location (r_0,θ_0) is linearly related to the data; as discussed by Schou *et al.* (1992) this allows the introduction of inversion coefficients and generalized averaging kernels $\mathcal{K}(r_0,\theta_0,r,\theta)$ defined such that $\bar{\Omega}(r_0,\theta_0)$ is related to the true angular velocity $\Omega(r,\theta)$ through

$$\bar{\Omega}(r_0,\theta_0) = \int_0^\pi \int_0^R \mathcal{K}(r_0,\theta_0,r,\theta)\Omega(r,\theta)\, r\mathrm{d}r\mathrm{d}\theta \; . \tag{9.21}$$

The form of the inverse problem evidently depends on the representation of the data. The general problem has the form

$$\omega_{nlm} - \omega_{nl0} = \delta\omega_{nlm} = m \int_0^R \int_0^\pi K_{nlm}(r,\theta)\Omega(r,\theta) r \mathrm{d}r\mathrm{d}\theta \tag{9.22}$$

(*cf.* eq. 8.35). However, it is often the case that the data do not allow determination of individual frequencies ω_{nlm}. In that case, it is customary to make fits of the general form shown in equation (2.35), *i.e.*,

$$\omega_{nlm} = \omega_{nl0} + 2\pi \sum_{j=1}^{j_{\max}} a_j(n,l)\, \mathcal{P}_j^{(l)}(m) \ , \tag{9.23}$$

where $\mathcal{P}_j^{(l)}$ is a polynomial of degree j. Since the fitting procedure is in general linear, the *a coefficients* $a_j(n,l)$ are linearly related to the frequency splittings,

$$2\pi a_j(n,l) = \sum_m \gamma_j(l,m)(\omega_{nlm} - \omega_{nl0}) \ , \tag{9.24}$$

for some coefficients $\gamma_j(l,m)$. It immediately follows from equations (8.35) and (8.36) that rotation gives rise to odd a coefficients, related to $\Omega(r,\theta)$ by

$$2\pi a_{2j+1}(n,l) = \int_0^R \int_0^\pi K_{nlj}^{(a)}(r,\theta)\Omega(r,\theta) r \mathrm{d}r\mathrm{d}\theta \ , \tag{9.25}$$

where the kernels $K_{nlj}^{(a)}$ can be determined in a straightforward manner from the kernels $K_{nlm}(r,\theta)$.

As discussed in Section 8.4 it may also be convenient to expand the dependence of $\Omega(r,\theta)$ on θ on the form

$$\Omega(r,\theta) = \sum_{s=0}^{s_{\max}} \Omega_s(r)\psi_{2s}^{(1)}(\cos\theta) \ , \tag{9.26}$$

where $\psi_{2s}^{(1)}(x)$ is a polynomial in x^2 of degree s. [Note that, as discussed at the end of Section 8.2, the rotational splitting is sensitive only to the component of Ω that is symmetrical around the equator; thus only even powers of $\cos\theta$ enter into the expansion (9.26).] Based on the expansions given in equations (9.23) and (9.26), the inverse problem can be formulated as the determination of the expansion functions $\Omega_s(r)$ through a series of one-dimensional inversions of the a coefficients $a_j(n,l)$ (see, for example, Korzennik *et al.* 1988; Brown *et al.* 1989; Thompson 1990). Such procedures are often called 1.5-dimensional (or 1.5D) inversions. By chosing the polynomials $\mathcal{P}_j^{(l)}$ as defined after equation (2.35) (Ritzwoller & Lavely 1991; Schou, Christensen-Dalsgaard & Thompson 1994) and the corresponding expansion of Ω (*cf.* eq. 8.61), the relations between the a coefficients and the expansion functions of Ω simplify to

$$2\pi a_{2s+1}(n,l) = \int_0^R K_{nls}^s(r)\Omega_s(r)\mathrm{d}r \ , \tag{9.27}$$

so that the separate 1-dimensional inversions can be carried out independently.

The 1.5-dimensional techniques very considerably reduce the computational efforts required for the inversion. However, the expansion of Ω evidently imposes a rather special

structure on the solution, unless a large number of terms is included. An alternative is to perform direct two-dimensional inversions, based either on equations (9.22) or equations (9.25). Even given the large amount of data (as many as 200 000 individual frequencies ω_{nlm}, or of order 50 000 a coefficients $a_j(n,l)$), this can be handled by means of a straightforward regularized least-squares technique (*e.g.* Sekii 1991; Schou 1991; Schou, Christensen-Dalsgaard & Thompson 1992, 1994). Here the inferred $\bar{\Omega}(r,\theta)$ is represented on a suitable grid (r_p,θ_q) in r and θ, $p=1,\ldots,n_r$, $q=1,\ldots,n_\theta$, by expansion coefficients Ω_{pq}. These expansion coefficients can be determined through a regularized least-squares fitting technique, analogous to the one described in equation (9.19). Assuming that splittings $\delta\omega_{nlm}$ for individual models are available, related to $\Omega(r,\theta)$ by equation (9.22), the solution is determined by minimizing

$$\sum_{nlm}\left(\frac{\int_{r,\theta}K_{nlm}\bar{\Omega}r\mathrm{d}r\mathrm{d}\theta-\delta\omega_{nlm}}{\sigma_{nlm}}\right)^2 \qquad (9.28)$$
$$+\mu_r\int_{r,\theta}f_r(r,\theta)\left(\frac{\partial^2\bar{\Omega}}{\partial r^2}\right)^2\mathrm{d}\theta\mathrm{d}r+\mu_\theta\int_{r,\theta}f_\theta(r,\theta)\left(\frac{\partial^2\bar{\Omega}}{\partial\theta^2}\right)^2\mathrm{d}\theta\mathrm{d}r\ ;$$

here σ_{nlm} is the standard deviation for the observed splitting $\delta\omega_{nlm}$. The last two terms serve to regularize the solution, as before, and depend on the weight functions f_r and f_θ and the trade-off parameters μ_r and μ_θ. Instead of $\delta\omega_{nlm}$, expansion coefficients a_{2j+1} with the corresponding kernels may evidently also be used. As in the 1-dimensional case, the trade-off parameters must be determined such as to ensure a balance between resolution and error. However, here one must also balance the resolution in the radial and latitude directions (*e.g.* Schou *et al.* 1994).

To illustrate the resolution properties of the inversions we may consider the averaging kernels which, as discussed above, can be defined both for inversions assuming an expansion of Ω in powers of $\cos\theta$ and for the full 2-dimensional techniques. Examples of such kernels are shown in Figure 9.4, both for an inversion using expansions of the splittings and Ω and for a full, two-dimensional inversion. In the former case, only a_1, a_3 and a_5 were included and the latitude information is consequently relatively limited; hence the kernels have a substantial extent in latitude. A particularly striking feature is the fact that the attempt to determine the angular velocity close to the pole results in what contains aspects of extrapolation from lower latitudes: indeed, it is obvious that the rotation of the region very near the pole has little effect on the frequency splittings and hence cannot be determined from the inversion. The inversion based on individual splittings provides substantially better resolution in latitude, as might have been anticipated. Within the convection zone it is possible to determine the rotation over a region extending only a few per cent of the solar radius in both radial and latitude directions, for realistic sets of observed frequency splittings.

The MOLA and SOLA techniques, discussed in Section 9.1.1, may obviously also be generalized to the two-dimensional case, in principle. They offer considerable advantages in terms of the ability to control the resolution, and possibly other properties of the averaging kernels. However, a naive implementation would involve a prohibitive computational effort, since this requires inversion of matrices whose order is given by the number of data values. Fortunately, by utilizing the special properties of the kernels, very substantial improvements of computational efficiency can be achieved. One class of techniques, the so-called $\mathbb{R}^1\otimes\mathbb{R}^1$

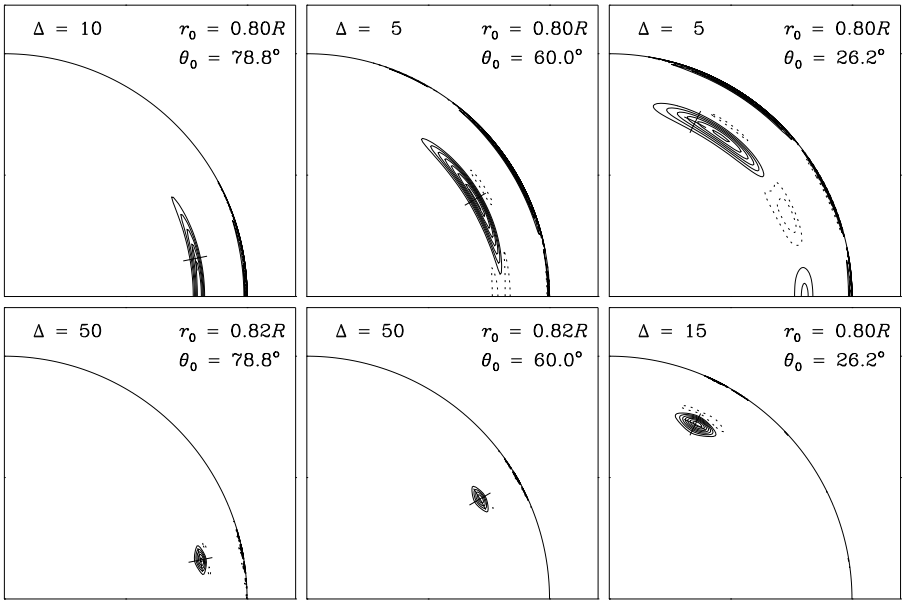

Figure 9.4: Contour plots of two-dimensional averaging kernels $R^{-2}\mathcal{K}(r_0, \theta_0, r, \theta)$ [*cf.* equation (9.21)]. Results are shown at a target radii near $r_0 = 0.8R$ and three different target co-latitudes, as indicated. The plots are in the (r, θ) plane, with the polar axis towards the top of the page. Positive contours are indicated by solid lines, negative contours by dashed lines; Δ is the value of the separation between contour levels. The cross shows the target location (r_0, θ_0). The top row shows kernels for an inversion based on just a_1, a_3 and a_5 and using a corresponding expansion of Ω. The bottom row shows results for a full two-dimensional regularized least-squares inversion, for a mode set aiming at representing the results of 1 year's observations with high spatial resolution. Adapted from Schou *et al.* (1994).

methods, uses explicitly the fact that the kernels can be written as

$$K_{nlm}(r, \theta) = F_{1\,nl}(r) G_{1\,lm}(\theta) + F_{2\,nl}(r) G_{2\,lm}(\theta) \tag{9.29}$$

[*cf.* equations (8.31), (8.34) and (8.35)], where the first term generally dominates; a similar expression applies in the case of kernels for a coefficients (*e.g.* Pijpers 1997). It was pointed out by Sekii (1993) and Pijpers & Thompson (1996) that as a result the inversion can essentially be separated into distinct, and computationally more manageable, inversions in r and θ.

Larsen & Hansen (1997) showed that the linear equations [equations (9.16) – (9.18)] arising in the SOLA technique can be solved efficiently by using explicitly the discretized representation of the kernels. By using also the structure given by equation (9.29), Larsen (1997) has developed an iterative technique which allows two-dimensional SOLA inversion

to be carried out with fairly modest means (see also Larsen *et al.* 1998). These efficient techniques have been applied to the very extensive and accurate data obtained with the GONG and SOHO projects.

Further details on the implementation of rotational inversion, and tests of the various techniques, were provided by Schou *et al.* (1998).

9.2 Inversion for solar structure

In its most general form, the dependence of the oscillation frequencies on solar structure may be expressed as

$$\omega_{nl} = \mathcal{F}_{nl}[\rho(r), c^2(r), \ldots] \; , \tag{9.30}$$

where, as indicated, other properties beyond the "mechanical" structure as characterized by ρ and c^2 may affect the frequencies. This equation is often approximated by the corresponding equation for the adiabatic frequencies, *i.e.*,

$$\omega_{nl}^{\rm ad} = \mathcal{F}_{nl}^{\rm ad}[\rho(r), c(r)] \; , \tag{9.31}$$

where the functional $\mathcal{F}_{nl}^{\rm ad}$ is determined through the solution of the equations of adiabatic oscillation. The inverse problem for solar structure then consists of inferring properties of the structure by "solving" equation (9.30) or (9.31), given a set of observed frequencies $\{\omega_{nl}^{\rm (obs)}\}$.

A difficulty in this process is that the frequencies depend on solar structure in a complicated, nonlinear way. As is common for nonlinear equations, the analysis proceeds through linearisation around an initial reference model. Let $(\rho_0(r), c_0(r))$ correspond to the reference model, which has adiabatic oscillation frequencies $\omega_{nl}^{(0)}$. We seek to determine corrections $\delta_r\rho(r) = \rho(r) - \rho_0(r)$ and $\delta_r c^2(r) = c^2(r) - c_0^2(r)$ to match the differences $\omega_{nl}^{\rm (obs)} - \omega_{nl}^{(0)}$ between the observed frequencies and those of the reference model. As discussed in Sections 5.5.3 and 5.5.4, linearization of equation (9.30), assuming $\delta_r\rho$ and $\delta_r c^2$ to be small, leads to

$$\begin{aligned} \frac{\delta\omega_{nl}}{\omega_{nl}} \; &= \; \int_0^R \left[K_{c^2,\rho}^{nl}(r) \frac{\delta_r c^2}{c^2}(r) + K_{\rho,c^2}^{nl}(r) \frac{\delta_r\rho}{\rho}(r) \right] {\rm d}r \\ &\quad + Q_{nl}^{-1} \mathcal{G}(\omega_{nl}) + \epsilon_{nl} \; , \end{aligned} \tag{9.32}$$

where the kernels $K_{c^2,\rho}^{nl}$ and K_{ρ,c^2}^{nl} are determined from the eigenfunctions in the reference model (see also Section 5.5.3). In equation (9.32) I included a contribution from the uncertainties in the near-surface region, expressed by the term in $\mathcal{G}$; this may be assumed to contain the difference between the "true" function $\mathcal{F}_{nl}$ in equation (9.30) and the adiabatic approximation $\mathcal{F}_{nl}^{\rm ad}$ in equation (9.31). Furthermore, I explicitly included the observational errors ϵ_{nl}. An additional constraint on $\delta_r\rho$ is that the mass of the Sun and the reference model be the same, *i.e.*,

$$4\pi \int_0^R \frac{\delta_r\rho(r)}{\rho(r)} \rho(r) r^2 {\rm d}r = 0 \; . \tag{9.33}$$

In this way the original nonlinear inverse problem is reduced to a linear problem, which may be analyzed by means of techniques similar to those discussed in Section 9.1.1.

Unlike the rotational case, the linearized inverse problem given by equation (9.32) involves three unknown functions: $\delta_r\rho(r)$, $\delta_r c^2(r)$ and $\mathcal{G}(\omega)$. These may, after suitable

parametrization, be determined through least-squares fitting with appropriate regularization (*e.g.* Dziembowski, Pamyatnykh, & Sienkiewicz 1990). Alternatively, some form of optimally localized averages may be used, by forming a linear combination of equations (9.32),

$$\begin{aligned}\sum_i c_i(r_0)\frac{\delta\omega_i}{\omega_i} &= \sum_i c_i(r_0)\int_0^R K^i_{c^2,\rho}(r)\frac{\delta_r c^2}{c^2}(r)\mathrm{d}r \\ &+ \sum_i c_i(r_0)\int_0^R K^i_{\rho,c^2}(r)\frac{\delta_r\rho}{\rho}(r)\mathrm{d}r \\ &+ \sum_i c_i(r_0)Q_i^{-1}\mathcal{G}(\omega_i) + \sum_i c_i(r_0)\epsilon_i \ ,\end{aligned} \tag{9.34}$$

where, as in Section 9.1.1, i labels the modes. If the goal is to determine the correction to $c^2(r_0)$, the coefficients $c_i(r_0)$ must be chosen such that the first term on the right-hand side of equation (9.34) provides an average of $\delta_r c^2/c^2$ localized near $r = r_0$, while minimizing the effect of the remaining terms.

A natural generalization of the SOLA technique is to obtain the coefficients $c_i(r_0)$ through minimization of

$$\int_0^R \left[\mathcal{K}_{c^2,\rho}(r_0,r) - \mathcal{T}(r_0,r)\right]^2 \mathrm{d}r + \beta\int_0^R \mathcal{C}_{\rho,c^2}(r_0,r)^2\mathrm{d}r + \mu\sum_{ij} E_{ij}c_i(r_0)c_j(r_0) \ , \tag{9.35}$$

where

$$\mathcal{K}_{c^2,\rho}(r_0,r) = \sum_i c_i(r_0)K^i_{c^2,\rho}(r) \tag{9.36}$$

is the averaging kernel; the *cross term*

$$\mathcal{C}_{\rho,c^2}(r_0,r) = \sum_i c_i(r_0)K^i_{\rho,c^2}(r) \tag{9.37}$$

measures the influence of the contribution from $\delta_r\rho$ on the inferred $\delta_r c^2$, and E_{ij} is the covariance matrix, as before. The constraint (9.33) is incorporated by adding a fictitious data point, with zero data and zero error, and with zero sound-speed kernel and a density kernel given by ρr^2. The term in $\mathcal{G}(\omega)$, where $\mathcal{G}$ is assumed to be a slowly varying function of frequency, may be suppressed by restricting the combinations of the data to those that are insensitive to a contribution of this form (Däppen *et al.* 1991; Kosovichev *et al.* 1992). Specifically, the coefficients may be constrained to satisfy

$$\sum_i c_i(r_0)Q_i^{-1}\psi_\lambda(\omega_i) = 0 \ , \lambda = 0,\ldots,\Lambda \ , \tag{9.38}$$

for a suitably chosen set of functions ψ_λ. It was shown by Basu *et al.* (1996a) that an equivalent, but potentially more flexible, suppression of the near-surface terms may be based on the filtering technique considered by Pérez Hernández & Christensen-Dalsgaard (1994a).

The SOLA inversion is characterized by the trade-off parameters β and μ controlling the influence of the cross term and the errors, respectively, by the parameters determining the target function $\mathcal{T}(r_0,r)$ and by the number Λ of terms included in the suppression

of the surface effects. The considerations involved in the choice of these parameters were discussed by Rabello-Soares, Basu & Christensen-Dalsgaard (1999a).

The form of the surface term in equation (9.32) assumed that the local properties of the eigenfunctions in the near-surface region are independent of degree; this is what led to the function $\mathcal{G}$ being just dependent on frequency. From an asymptotic point of view this corresponds to assuming that the rays characterizing the modes are nearly vertical in this region. For modes of high degree this approximation no longer holds. Brodsky & Vorontsov (1993) showed how the asymptotic relation (7.1) should be modified in this case, by introducing l-dependent terms in the phase function α. The introduction of the corresponding terms in structure inversion by means of the SOLA or MOLA techniques, generalizing the constraints in equations (9.38), was discussed by Di Mauro *et al.* (2002); they also applied the techniques to preliminary observed frequencies of high-degree modes, obtained by Rhodes *et al.* (1998) from analysis of observations from the SOI/MDI instrument on the SOHO spacecraft.

Although the inversion has been discussed in terms of the pair (c^2, ρ), other sets of variables characterizing the equilibrium structure of the Sun may be used (see also Section 5.1.1). In particular, as discussed in Section 5.5.3, the frequency changes can be expressed in terms of changes $\delta_r\rho$ and $\delta_r Y$ in density and helium abundance, if the equation of state (and the heavy-element abundance) are assumed to be known. Another equivalent, and commonly used, pair is (u, Y), where $u = p/\rho$ is the squared isothermal sound speed. From the point of view of inversion, these pairs have the substantial advantage that the kernels corresponding to $\delta_r Y$ are relatively small and essentially confined to the ionization zones of hydrogen and helium. Thus in the minimization corresponding to equation (9.35), it is comparatively easy to suppress the cross term $\mathcal{C}_{\rho,Y}(r_0, r)$. Furthermore, inversion can be carried out to determine the difference $\delta_r Y$ between the solar and model helium abundance (*e.g.* Kosovichev *et al.* 1992). It should be noted, however, that differences between the solar and model equations of state may introduce systematic errors in the results of such inversions. Basu & Christensen-Dalsgaard (1997) showed how the differences in equation of state might be taken explicitly into account in the inversion, albeit at the expense of an increase in the error in the solution; they also pointed out that the inversion might be carried out to determine the intrinsic difference in Γ_1 between the solar and model equations of state, *i.e.*, the difference at fixed p, ρ and composition.

I finally note that asymptotic inversion techniques to determine the solar internal sound speed were discussed in Section 7.7.2; in particular, it is possible to estimate the sound speed directly from the data, without the use of linearization. Such techniques were originally developed in geophysics (see Brodsky & Levshin 1979); their application to the helioseismic problem was first considered by Gough (1984). In these techniques, the uncertainties associated with the near-surface region are contained in the function $\alpha(\omega)$ or, for the differential technique, in the function $\mathcal{H}_2(\omega)$ (*cf.* Section 7.7.3).

9.3 Some results of helioseismic inversion

To illustrate the power of current helioseismic analysis, it is instructive to present a few selected results of the application of helioseismic inversion; results of the use of the differential asymptotic technique were already shown in Figure 7.13. I first consider inversion to determine solar structure. As discussed in Section 5.1, this is carried out to determine

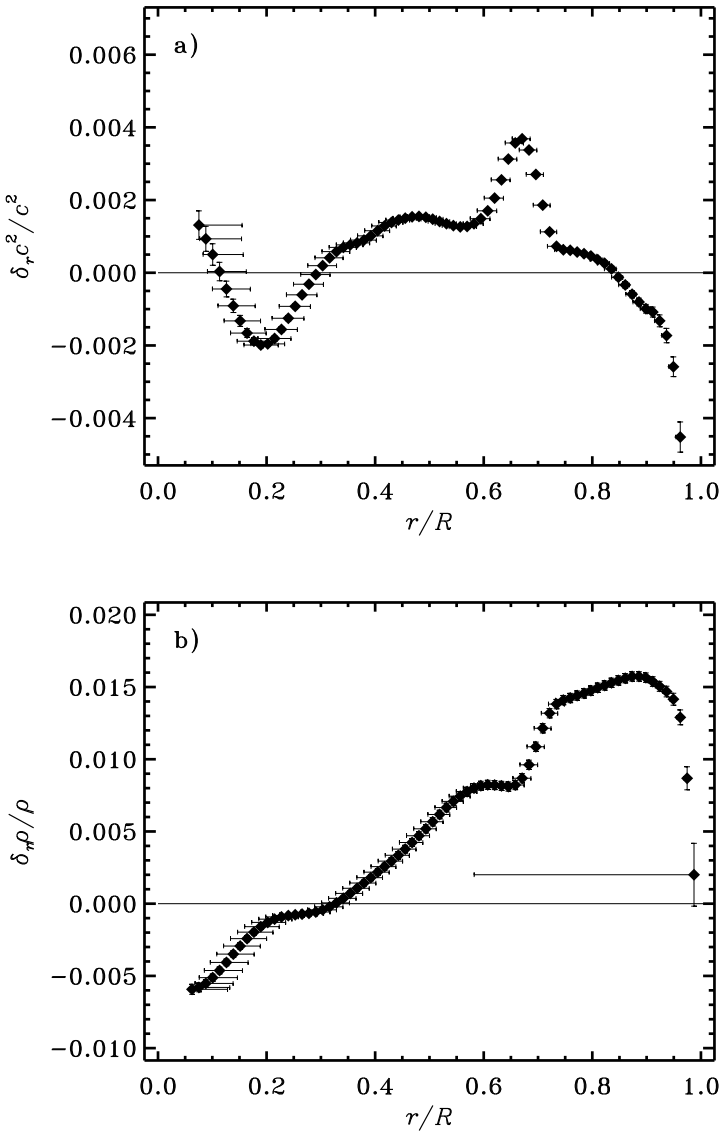

Figure 9.5: Corrections to a reference solar model, obtained by inverting differences between the observed frequencies and the frequencies of the model. Panel a) shows corrections to the squared sound speed c^2, and panel b) shows corrections to ρ. The vertical bars indicate the errors in the results, based on the errors in the observed frequencies, whereas the horizontal bars provide a measure of the resolution in the inversion (from Basu *et al.* 1996b).

corrections to an assumed initial reference model. An example is provided by the results of Basu *et al.* (1996b), based on observations by the LOWL instrument, developed at the High Altitude Observatory (Tomczyk *et al.* 1995). The inversions were carried out in terms of corrections $\delta_r c^2$ and $\delta_r \rho$. In the former case the variable pair (c^2, ρ) was used, while the

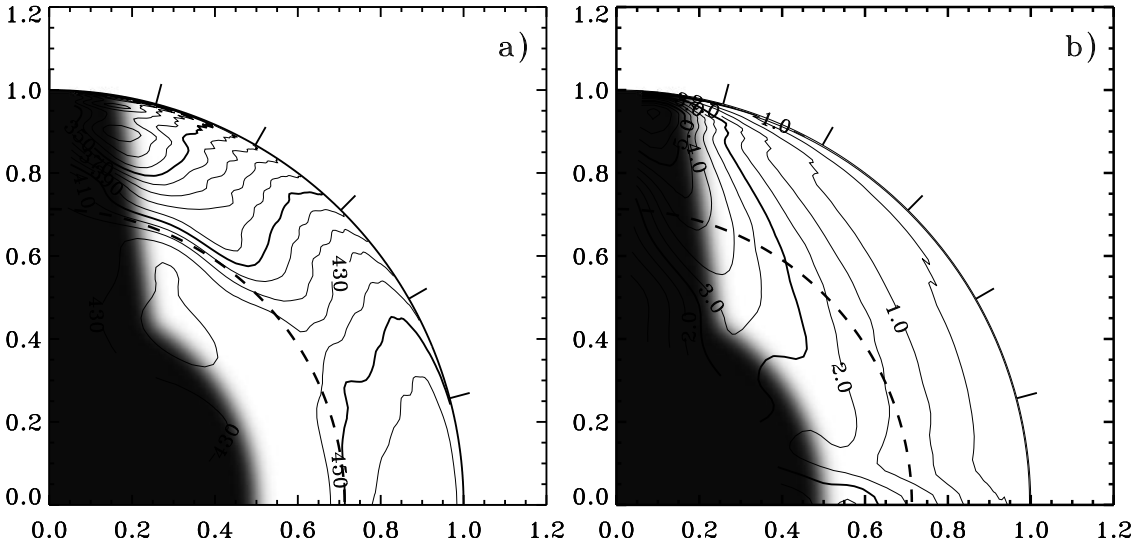

Figure 9.6: Inferred rotation rate $\Omega/2\pi$ (panel a) and the associated error (panel b) in a quadrant of the Sun, obtained by means of SOLA inversion of 144 days of MDI data. The equator is at the horizontal axis and the pole is at the vertical axis, both axes being labelled by fractional radius. Some contours are labelled in nHz, and, for clarity, selected contours are shown as bold. The dashed circle is at the base of the convection zone and the tick marks at the edge of the outer circle are at latitudes 15°, 30°, 45°, 60°, and 75°. The shaded area indicates the region in the Sun where no reliable inference can be made with the present data. (Adapted from Schou *et al.*, 1998.)

density inversion used the pair (ρ, Y); as discussed in Section 5.5.3 it is therefore sensitive to possible errors in the equation of state. As reference was used a model typical of recent normal solar models, computed with little *a priori* attempt to match the observations. The model included settling of helium and heavy elements, and used equation of state and opacity tables from Livermore. The inversion for the corrections was performed by means of the SOLA method discussed in Section 9.2. The results are shown in Figure 9.5; there are evidently systematic differences between the Sun and the model but these are generally fairly small. It is probable that the sharp bump in $\delta_r c^2/c^2$ at $r/R \simeq 0.65$ is a result of weak mixing just beneath the convection zone, which partly offsets the sharp gradient in the hydrogen abundance established by helium settling in this region. Similarly, the negative $\delta_r c^2/c^2$ at the edge of the core could result from partial mixing of the central region. One cannot exclude, however, that errors in the opacity may play a significant role. Nonetheless, it is remarkable that models computed without direct reference to the observations, but based solely on our knowledge of the physics of the solar interior, are so successful in reproducing the structure as inferred from the helioseismic analysis.

To illustrate the determination of the solar internal angular velocity, I present results obtained by Schou *et al.* (1998) from analysis of early data from the MDI instrument on

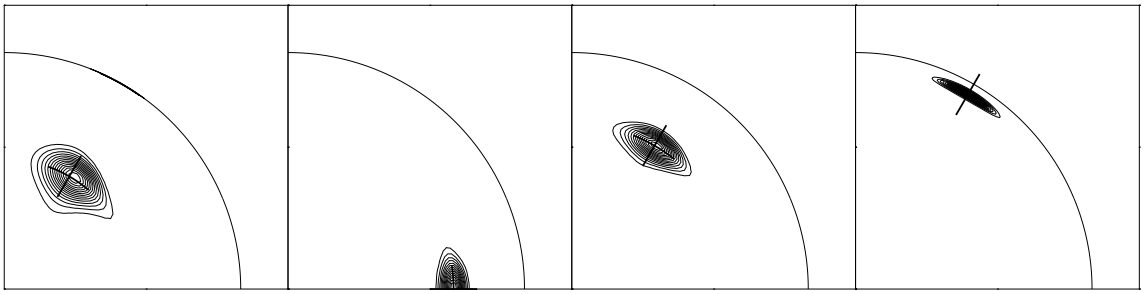

Figure 9.7: Averaging kernels for the SOLA inversion shown in Figure 9.6, targeted at the following radii and latitudes in the Sun: $0.55R$, 60°; 0.7, 0°; 0.7, 60°; $0.95R$, 60°. The corresponding locations are indicated with crosses. (Adapted from Schou et al. 1998.)

the SOHO spacecraft. The inversion was carried out as a fully two-dimensional SOLA inversion of a coefficients, extending to a_{35}, to determine $\bar{\Omega}(r_0, \theta_0)$. The resulting $\bar{\Omega}$ and the estimated errors are presented in Figure 9.6, as contour plots. Strikingly, the error in a substantial part of the Sun is less than 2 nHz. To illustrate the resolution, Figure 9.7 shows selected averaging kernels. Further details of the solution are visible in Figure 9.8, which shows cuts at fixed latitudes, as functions of distance to the centre; here, in addition to the SOLA results, solutions obtained from a two-dimensional regularized least-squares inversion have been included.

There is a striking change in the behaviour of rotation near the base of the convection zone, at a depth of about 28 per cent of the solar radius (as inferred helioseismically; *e.g.* Christensen-Dalsgaard, Gough & Thompson 1991); this is marked by the heavy dashed circle in Figure 9.6 and the heavy dashed line in Figure 9.8. Within the convection zone the variation with latitude in the rotation rate is quite similar to the behaviour observed directly on the surface; in particular, the values at the outermost points in the solution are essentially in agreement with the surface values. (It should be noted that the inversion does not impose continuity with the surface angular velocity.) Near the base of the convection zone there is a transition such that the angular velocity in the radiative interior is roughly independent of latitude, at a value intermediate between the surface equatorial and polar values, but substantially closer to the former. This transition region is known as the *tachocline* (Spiegel & Zahn 1992) and likely plays an important role in the generation of the solar magnetic field and the origin of the solar magnetic cycle. The apparent width of the tachocline in Figure 9.8 in part reflects the finite resolution of the inversion, as determined by the radial extent of the averaging kernels. This must be taken into account in estimating the true width of the tachocline. Charbonneau *et al.* (1999) applied several analysis techniques to LOWL data; they obtained a tachocline width, defined in terms of a representation of the transition by an error function, of $(0.039 \pm 0.013)\,R$ and an equatorial central radius $r_c = (0.693 \pm 0.002)\,R$, essentially placing the transition beneath the convection zone.

Although the overall features of rotation, as presented above, have been found using several different data sets and analysis methods, it should be mentioned that there are problems at the level of finer details, particularly at higher latitudes. These have become

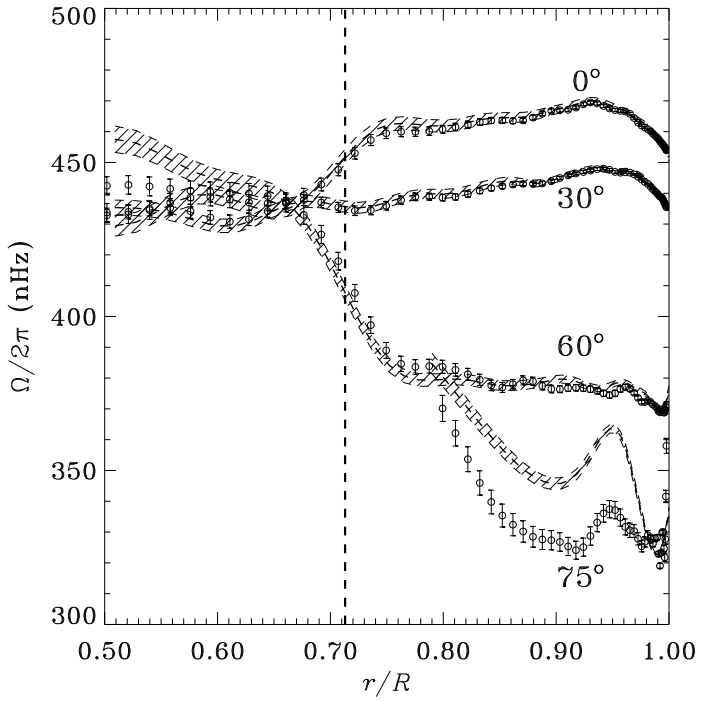

Figure 9.8: Inferred rotation rate $\Omega/2\pi$ as a function of radius at the latitudes indicated, obtained from inversion of 144 days of MDI data. The circles with 1-σ error bars show results of a SOLA inversion, while the dashed lines with 1-σ error band were obtained with regularized least-squares inversion. The heavy vertical dashed line marks the base of the convection zone. (Adapted from Schou *et al.*, 1998.)

apparent in comparisons between results based on data from the GONG and SOI/MDI projects, in both cases analyzed with the procedures used by both projects (*e.g.* Schou *et al.*, 2002). Also, as illustrated by the comparison of the SOLA and least-squares results in Figure 9.8, different inversion methods may give different results at high latitude. Clearly, the underlying causes for these various differences, and how to correct for them, need to be identified.

The data used in the inversions presented in Figures 9.6 and 9.8 did not permit inference of the rotation rate very near the centre. However, analysis of low-degree splittings from the BiSON network provided a tantalizing hint that the core rotation might be *below* the general rotation rate of the radiative interior (Elsworth *et al.* 1995). Chaplin *et al.* (1999) carried out a more detailed analysis of a combination of LOWL and BiSON frequencies, using a version of the MOLA technique especially designed to localize the averaging kernels to the solar core. The results are shown in Figure 9.9. They are consistent with constant rotation of the radiative interior, although with a possible suggestion of a down-turn in the core; analysis of the averaging kernels showed that constraining the measure of rotation to the inner 20 % of the solar radius was only possible at the expense of very substantial

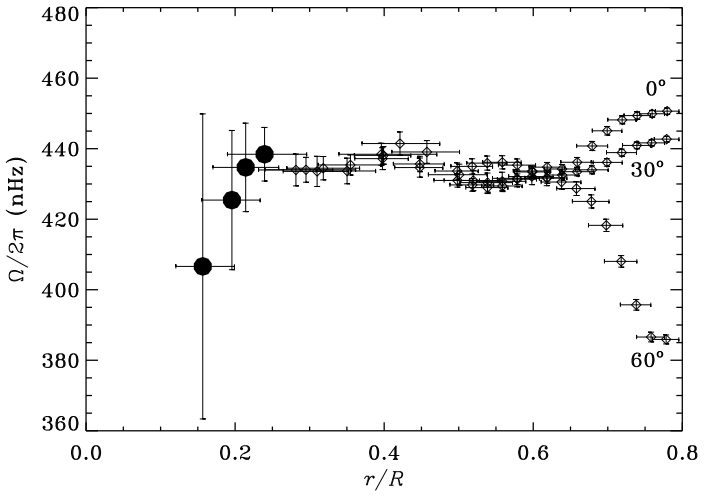

Figure 9.9: The inferred rotation as a function of fractional radius inside the Sun at three solar latitudes: the equator, 30° and 60°; the vertical axis shows the rotation frequency in nHz. The vertical error bars indicate the statistical uncertainty on the rotation rate (±1 standard deviation), whereas the horizontal bars provide a measure of the radial resolution of the inversion. Note that the result becomes much more uncertain in the deep interior, where furthermore the different latitudes cannot be separated. The observations used to infer the rotation were from the LOWL instrument and the BiSON network. (From Chaplin *et al.* 1999.)

errors in the inferred rotation rate.

A more detailed discussion of recent results of helioseismology, with extensive references, was presented by Christensen-Dalsgaard (2002).