SOLAR INVERSE THEORY

Douglas Gough
Institute of Astronomy and Department of Applied Mathematics
and Theoretical Physics, University of Cambridge, and Joint
Institute for Laboratory Astrophysics, University of Colorado
and National Bureau of Standards, Boulder, Colorado 80309

Helioseismological inversion, as with the inversion of any other
data, is divided into three phases. The first is the solution of the
so-called forward problem: namely, the calculation of the eigenfre-
frequencies of a theoretical equilibrium state. The second is an attempt to
understand the results, either empirically by determining how those fre-
quencies vary as chosen parameters defining the equilibrium model are
varied, or analytically from asymptotic expansions in limiting cases of
high order or degree. A familiarity with at least the qualitative de-
pendence of the eigenfrequencies on various properties of the solar
model is necessary not only for personal enlightenment but also for
arming oneself to interpret the rather more abstract third phase. That
phase is to pose and solve an inverse problem, which seeks to find a
plausible equilibrium model of the Sun whose eigenfrequencies are con-
sistent with observation.

The three phases are briefly discussed in this review, and the
third, which is not yet widely used in helioseismology, is illustrated
with some selected inversions of artificial solar data.

1. INTRODUCTION

There is now a substantial body of seismological data from the
Sun. Therefore we can contemplate graduating from the hit-and-miss
model-fitting investigations that have dominated the subject in the past
to the more systematic approach provided by inverse theory. There are
many methods for attempting to invert seismological data, developed
mainly by geophysicists and applied mathematicians interested in ge-
ophysical problems, so we are in the fortunate situation of being able
to draw on their experience. There appears to be a variety of schools
of opinion about which methods are generally preferable, or even which
approach is likely to be the most fruitful for a particular problem.
What seems to be universally agreed, however, is that when attacking a
totally new problem it is expedient to test a method first, using arti-
ficial data generated from a theoretical model.
In this and the accompanying discussion (Christensen-Dalsgaard and Gough, 1984a), we report briefly on two potentially useful methods of inverting helioseismological data. One is based on localized averages, and the other is an expansion (spectral expansion) in the weighting functions (data kernels) that characterize the contribution of different regions of the star to the data. Both methods have been used in geophysics, and have their origins in the work of Backus and Gilbert (e.g. 1970). A detailed description of the methods, and their relation to other techniques that have been used, will not be given here; for this the reader is referred to the various reviews that already exist (e.g. Wiggins, 1972; Parker, 1977a,b; Sabatier 1977, 1978). The examples discussed here and in the accompanying article of the application of these methods to artificial data suggest that both of the procedures offer hope of being useful for inverting solar data.

The two aspects of the solar structure that are most likely to be determined in the foreseeable future are the internal rotation and the hydrostatic stratification. From a theoretical point of view the former is the simpler problem, because it is essentially linear; moreover the result is likely to be quite reliable, because it is not sensitive to uncertainties in the physics of stellar material. The latter is non-linear, and depends on a knowledge of the equation of state and, to a lesser extent, on the dynamics of convection. Nevertheless there are good grounds to believe that a useful inversion, which sets certain stringent constraints on the stratification, will soon be possible.

2. PHASE ONE: THE FORWARD PROBLEM

The forward problem is discussed in the standard texts on stellar pulsation theory (Ledoux and Walraven, 1958; Unno et al., 1979; Cox, 1980). Since to a first approximation the Sun may be regarded as being spherically symmetrical (the oblateness of the density distribution is nowhere more than about $10^{-5}$), the oscillation eigenfunctions are separable. For example, the displacement eigenvector $\xi(r, \theta, \phi)$ may be written, with respect to spherical polar coordinates $(r, \theta, \phi)$:

$$\xi = (\Xi(r)P_{\ell}^m, r^{-1}H(r)\partial \phi P_{\ell}^m, r^{-1}\csc \theta H(r) P_{\ell}^m \partial \phi \cos(\psi - \omega t))$$

(2.1)

where $P_{\ell}^m(\cos \theta)$ is the associated Legendre function of the first kind and $t$ is time; $\ell$ and $m$ are respectively the degree and the azimuthal order of the mode.

To a good first degree of approximation (no worse than 1 part in $10^3$) the oscillations may be regarded as being adiabatic. This property has a good aspect and a bad aspect. The good aspect is that the physics of adiabatic oscillations is relatively well understood. So an accurate and fairly reliable solution to the forward problem is possible. The dynamics of the oscillations hardly senses the perturbations to the heat flow, and therefore it is not necessary to have an accurate knowledge of the opacity or the nuclear energy generation rates, or how the convective
heat flux is modulated by the oscillations, at least for the lower-frequency modes. The highest frequencies that have been observed do seem to have a somewhat greater sensitivity to the nonadiabatic effects of convection, however, the theory of which is quite uncertain. Nevertheless, since these are significant only in a very thin layer near the top of the convection zone, it is possible, at least in principle, to eliminate their influence by considering appropriate combinations of eigenfrequencies. I shall not discuss that here, but merely assume that they can be neglected, and consider the straightforward linearized perturbation analysis of a truly hydrostatic star.

The bad aspect relates to the good aspect: because heat flow is essentially irrelevant, the oscillation eigenfrequencies provide us with no direct information about temperature. Adiabatic oscillations result solely from Newton's second law of motion: pressure gradients exert a force on material with inertia. Therefore all we can learn directly is a relation between pressure and density. If this is known throughout the Sun it can be coupled with the constraint of hydrostatic balance and Newtonian gravitation to provide us, at least in principle, with the density and pressure stratification. Of course it is necessary to know the equation of state, so that one can calculate the perturbation to the pressure gradient associated with a given compression or rarefaction, and this is perhaps the major uncertainty in the entire theory at present.

The dominant symmetry-breaking agent is the linear advection term coming from the angular velocity of the Sun. Its effect is to split the degeneracy of the eigenfrequencies of like order \((n)\) and degree \((\ell)\). From the splitting one can hope to infer the internal angular velocity \(\Omega(r, \theta, t)\). Because the splitting is small, quadratic effects, such as centrifugal forces acting on both the equilibrium state and the oscillations, can be ignored. Therefore the problem is, to a good approximation, linear. Several reports of observations of rotational splitting have already been made (Deubner et al., 1979; Claverie et al., 1981; Hill et al., 1982; Deubner, 1983; Rhodes et al., 1983a; Scherrer and Delache, 1984), and more are no doubt imminent.

In addition to \(\Omega\), large-scale convection currents (giant cells) should also produce diagnostically useful observable consequences (Gough and Toomre, 1983; Hill et al., 1983,1984). Locally, the horizontal component of the convective flow acts on high-degree oscillations in much the same way as rotation.

3. PHASE TWO: UNDERSTANDING THE FORWARD PROBLEM

Much of the understanding of the forward problem comes from analytical analyses in simplified circumstances. Useful information about high-degree oscillations, for example, can be extracted by approximating the equilibrium state by a polytrope or by assuming it to be isothermal (e.g. Lamb, 1932; Stein and Leibacher, 1974). Low-degree oscillations can be analyzed asymptotically at high order (e.g. Vandakurov, 1967; Zahn, 1970; Ledoux and Perdang, 1980; Tassoul, 1980). It is from analyses such as
these that one can learn what aspects of the observed dispersion relation pertain to different levels in the star. In particular, the small separation between the frequencies of modes \((n,\ell)\) and \((n-1,\ell+2)\) of low-degree five-minute oscillations provides information about the stratification in the energy-generating core (e.g., Gough, 1983a). Another example is presented in these proceedings, where it is pointed out that the latest observations of five-minute oscillations by Harvey and Duvall (1984) indicate an error in a standard solar model at or near the base of the convection zone (Christensen-Dalsgaard and Gough, 1984b).

Asymptotic relations usefully provide a coarse diagnostic of the solar interior. The frequencies of high-order modes depend predominantly on aspects of the equilibrium structure that vary on a scale much greater than the characteristic wavelength of the oscillation eigenfunctions, and can be expressed as integrals of the equilibrium structure that do not depend explicitly on the detailed variation of these eigenfunctions. In particular, Duvall's law (1982) describes the \(p\) modes, and appears to be approximately true for high-degree modes even when the order is not large (Christensen-Dalsgaard et al., 1984). The law can be inverted analytically to obtain an integral formula for the sound speed in terms of the observed dispersion relation (Gough, 1984).

Aside from discontinuities, aspects of the equilibrium structure that vary on a scale much shorter than a wavelength are not easily treated by analytical methods. They do influence the eigenfrequencies, however. In such cases, the effect on the frequencies depends on the detailed structure of the spatial oscillations in the eigenfunctions. Though it is often possible to see a posteriori how the frequencies are influenced, by noticing the locations of nodes and antinodes of the eigenfunctions in relation to the small scale variations in the equilibrium structure, it is usually not easy to make quantitative predictions.

It is certainly evident, therefore, that a systematic procedure is required for analyzing the eigenfrequencies to measure the equilibrium structure on a relatively small scale. In the present state of development of the theory, it is probably necessary to use such a procedure for much of the large-scale structure too, since the asymptotic approximations are not always reliable. Inverse theory will no doubt provide that procedure. But before embarking on my discussion of inverse theory it is perhaps appropriate simply to mention some of the conjectures that have arisen from the results of model-fitting. I do this to emphasize the importance of a simple procedure that provides an easy guide to the results that might subsequently be found by more sophisticated analyses.

The first conjecture from model-fitting was deduced from the fact that the frequencies of high-degree five-minute oscillations observed by Deubner (1975) were lower than the predictions (Ulrich, 1970; Ando and Osaki, 1976) of contemporary solar model envelopes. From a polytropic analysis (Gough, 1977) this seemed to imply that the models overestimated the entropy jump across the upper superadiabatic convective boundary layer, implying that the Sun's convection zone was actually deeper than the models predicted. Numerical experiments by Ulrich and
Rhodes (1977) were consistent with this idea, and subsequent more detailed numerical work (Lubow et al., 1980; Berthomieu et al., 1980), in which many of the other uncertain aspects of the theory were varied, suggested that the conjecture was likely to be the only possibility.

A second conjecture arose from a calibration performed by fitting to observations the eigenfrequencies of low-degree five-minute modes of a sequence of evolved solar models with different compositions (Christensen-Dalsgaard and Gough, 1980, 1981; Shibahashi et al., 1983; Ulrich and Rhodes, 1983). This suggested an initial solar helium abundance of 0.25 ± 0.02 (see also Gough, 1983b). Interestingly, it was found that a perfect fit between theory and observation was not possible, and subsequent numerical experiments with the theory failed to remove the discrepancy. As I have already pointed out, we have a strong indication of where the models are incorrect (Christensen-Dalsgaard and Gough, 1984b), but the nature of the error is yet to be found. Perhaps we must await an inverse calculation before we know the answer.

4. WHAT DO OSCILLATION FREQUENCIES TELL US?

Oscillations sample an extended region of the Sun, and so provide an integral measure of the structure over that region. This is most easily seen from the variational formulation of the adiabatic eigenvalue problem for a nonrotating star (e.g. Ledoux and Walraven, 1958):

\[ \omega^2 \int \rho \dot{\xi} \dot{\xi} dv = \int [\gamma p (\text{div} \tilde{\xi})^2 + 2 \xi \cdot \nabla p \text{ div} \tilde{\xi} + \rho^{-1} \xi \cdot \nabla \rho \tilde{\xi} \cdot \nabla \rho] dv \\
- G \int |\xi - \xi'|^{-1} \text{div} (\rho \xi) \text{ div'} (\rho (\xi') \xi (\xi')) dv' dv , \quad (4.1) \]

where \( p \) and \( \rho \) are the pressure and density of the equilibrium state, \( \gamma \) is the adiabatic exponent \( (\partial p / \partial n)_d \), \( G \) is the gravitational constant and the integrals are over the volume of the star. Thus the frequencies are a combination of weighted averages of nonlinear functions of the equilibrium pressure and density and their derivatives, the weighting depending on the oscillation displacement eigenfunctions \( \xi \). Since different eigenfunctions weight the structure differently, the hope is that with a sufficient variety of data one can obtain an estimate of how \( p \) and \( \rho \) vary with \( r \).

Most of the inversion procedures that have been developed apply only to linear integral equations. Therefore to make some progress it is expedient to develop an iterative procedure, like a generalized Newton-Raphson method, that improves (hopefully) upon a trial model of the Sun. One postulates an initial guess: \( p_0, \rho_0, \gamma_0 \), taken, for example, from a standard model of the Sun. Then one carries out the forward problem, calculating the eigenfunctions \( \xi_0 \) and eigenfrequencies \( \omega_0 \) corresponding to the modes for which observational data are available. One presumes that the physics is correctly described by equation (4.1), so that \( \omega, p, \rho, \xi \) refer to the actual Sun. Then one writes down the equation satisfied by \( \delta \omega^2 = \omega^2 - \omega_0^2 \), by subtracting two equations of the type (4.1),
and hopes that the trial model is sufficiently close to reality for linearization in the differences \( \delta \rho = \rho - \rho_0 \), \( \delta \rho = \rho - \rho_0 \), etc. to be valid. There results an equation in which \( \delta \omega^2 \) is expressed as a linear functional of the differences \( \delta \rho \), \( \delta \rho \), etc. It is at this point that one appreciates the variational formulation of the problem. Since equation (4.1) is stationary to variations in the functions \( \xi \) about the true eigenfunctions, it follows that the terms that are linear in \( \xi - \xi_0 \) cancel, and \( \delta \omega^2 \) can be expressed in terms of \( \xi_0 \) alone. Thus it is not necessary to perturb the forward problem.

One can proceed further by imposing the constraint of hydrostatic support

\[
\frac{\delta p}{dr} = -\frac{Gm}{r^2}
\]

where \( \tilde{m} \) is the usual mass variable which satisfies

\[
\frac{d\tilde{m}}{dr} = 4\pi r^2 \rho
\]

By substituting these equations into the equation for \( \delta \omega^2 \) and integrating by parts (and noticing that the surface integrals are negligible), it is possible to write the equation in the form

\[
\frac{\delta \omega^2}{\omega_0^2} = \int S(x_0, \xi_0, r) \frac{\delta p}{\rho_0} r^2 \rho_0 dr
\]

provided the equation of state is known, where \( x_0(r) \) represents the equilibrium structure \( (\rho_0, \rho_0, etc.) \) of the trial model. The formula for the differential kernel \( S \) is quite complicated, so I refrain from presenting it here. Several examples are plotted by Gough (1978a).

If equation (4.4) can be inverted to estimate \( \delta \rho \), an improved estimate of \( \rho \) can be deduced, and the whole procedure could be repeated.

In deriving equation (4.4) I presumed the equation of state was known. This is required for computing the variation in \( \gamma \). What is required is not only a knowledge of the microphysics of the solar material, but also the composition. Of course for the first iteration one has a trial composition taken from the standard solar model. But for subsequent iterations one has no such information, and the procedure must be generalized.

For the purposes of inferring the structure of the core (if it can be done without a detailed knowledge of the structure of the envelope) this issue is not very important. There the material is highly ionized, and \( \gamma = 5/3 \). But in the convection zone where the abundant elements are partially ionized, the problem is important. Formally it is quite straightforward to overcome the difficulty by generalizing equation (4.4) to include additional integrals that are weighted averages of the abundances of the elements that influence \( \gamma \), and generalizing the inversion procedure described in the following section. [The geophysical
inverse problem originally formulated by Backus and Gilbert (1967) was posed for a vector function.] No investigation of the properties of such a procedure applied to the Sun has yet been undertaken. Alternatively, it may be simpler to separate the problem of determining the composition of the convection zone and solving that first (Gough, 1984). This should be possible because in the convection zone the stratification is close to being adiabatic and the chemical composition is presumably homogeneous.

The case of rotational splitting of nonaxisymmetrical modes can be treated similarly. The perturbations \( \omega_1 \) to the eigenfrequencies imposed by the angular velocity \( \Omega(r, \theta, t) \) can be computed by linearizing the variational principle of Lynden-Bell and Ostriker (1967). Alternatively it can be obtained in the form of a consistency condition if one expands the eigenvalue problem about that for the nonrotating state. The result is a functional of \( \Omega \) which, in the special case when \( \Omega \) is a function of \( r \) alone, can be written in the form (Gough, 1981; cf. Hansen et al., 1977)

\[
\frac{\omega_1}{\omega_0} = m \int K(x_0, z_0, r) \Omega(r)dr.
\]

(4.5)

Quadratic and higher order terms can be computed, if desired, but these are small compared with the linear term represented in equation (4.5) (Dziembowski and Goode, 1984; Gough and Taylor, 1984). Once again, the data, namely the splitting frequencies \( \omega_1 \), are weighted averages of the function to be determined (this time \( \Omega \)), with weight functions that depend on the equilibrium model and its eigenfunctions. If the previous inverse problem has already been solved (using the observed frequencies of the axisymmetrical modes) \( K \) can be regarded as being known. This inverse problem is therefore linear, and requires no iteration.

5. PHASE THREE: INVERSE METHODS

In this section I consider explicitly the inversion of the idealized equation (4.5) to obtain \( \Omega \) from rotational splitting data. Now it is convenient to label the modes with an index \( i \) which identifies \( n, k \) and \( m \). Actually, in the special case considered here where \( \Omega \) depends only on \( r \), \( K \) is independent of \( m \) and therefore \( i \) need represent only \( n \) and \( k \). Thus one writes

\[
\omega_i = \int_0^R K_i(r) \Omega(r)dr
\]

(5.1)

where \( \omega_i = \omega_{1i}/(m \omega_{0i}) \) are observable quantities, \( K_i \) are known kernels and \( R \) is the radius of the Sun. In practice, observations contain errors, so if the \( \omega_i \) are regarded as the observations, equation (5.1) holds only approximately.
As pointed out by Backus and Gilbert (1967), a solution to the problem represented by (5.1), if it exists, is not unique, even if strict equality is assumed. One reason is that since one can have only a finite number of observations, the kernels $K_i$ cannot possibly span the space of functions on the interval $(0,R)$. Therefore there is an infinite number of functions, $f_k$, orthogonal to all the $K_i$, and any linear combination of them can be added to $\Omega$ without modifying any of the integrals. Let $0$ denote the subspace spanned by the $K_i$. The functions $f_k$ lie in its complement, $A$, called the annihilator. Evidently the data contain no information about the projection of $\Omega$ in $A$.

Permitting only approximate equality broadens the possibilities further. Therefore the problem is not simply to find an approximate solution to equation (5.1), but to select which of the infinite number (if there are any at all) is the most likely. Here prejudice reigns, and opinions therefore differ.

Before proceeding I shall assume that the physics embodied in equation (5.1) is correct (and that the data are consistent). In that case a solution must exist.

Spectral expansion

This is essentially an expansion of $\Omega$ in terms of the kernels $K_i$. The idea of using $K_i$ as a basis is, at first sight, natural, since the $K_i$ span the subspace $0$ that is accessible to the observations. However, as I shall soon discuss, it is actually more useful to transform, at least conceptually, to a new basis, $\psi_i$, that takes into account the degree to which the data can measure the projection of $\Omega$ onto each basis function.

In geophysics, the expansion first arose out of a procedure formulated by Backus and Gilbert (1967). Suppose one has a preconceived idea, $W(r)$ say, of the function $\Omega$. Let us assume that $W$ is close to the truth, and seek that function $\Omega$ that minimizes the least-squares deviation

$$\mathcal{E} = \int_0^R (\Omega - \overline{W})^2 \, dr \quad .$$  \hspace{1cm} (5.2)

Forgetting errors for the moment, the minimization must be performed subject to the constraints (5.1), with exact equality. The result is

$$\Omega = W + \sum_i a_i^* K_i \quad ,$$  \hspace{1cm} (5.3)

with $a_i^*$ being the solutions of the Euler equations

$$\sum_j A_{ij} a_j^* = \omega_i - a_i \quad ,$$  \hspace{1cm} (5.4)

where

$$A_{ij} = \int_0^R K_i K_j \, dr \quad , \quad a_i = \int_0^R K_i \, \overline{W} \, dr \quad .$$  \hspace{1cm} (5.5)
Thus the difference between $\Omega$ and $W$ is expressed, in equation (5.3), as a linear combination of the kernels $K_i$.

The foregoing analysis suggests that one might simply express $\Omega$ directly as an expansion in $K_i$:

$$\Omega = \sum_i \alpha_i K_i$$  \hspace{1cm} (5.6)

If once again errors are ignored, one can determine the coefficients in the usual way by projecting the constraints (5.1) onto the basis, which leads to

$$\sum_i A_{ij} a_j = w_i$$  \hspace{1cm} (5.7)

The difference between the solutions (5.6)-(5.7) and (5.3)-(5.4) is in the annihilator $A$, so the data $w_i$ cannot distinguish between the two. Therefore, if $w_i$ constitute the only information that is available about $\Omega$, one has no basis for choosing between the two possibilities. An inherent advantage of the Backus-Gilbert formulation, however, is that one can write into $\Omega$, via $W$, additional constraints (obtained from information other than $w_i$) that are not otherwise easily incorporated into a procedure for solving (5.1).

There are two related problems that one would encounter in trying to carry out the straightforward method outlined above. First, if there are redundant data, the matrix $A_{ij}$ is singular, and some care must be taken in solving equations (5.4) or (5.7). If the data $w_i$ were truly error-free, that would be possible, at least in principle, since, by hypothesis, the equations would be consistent, and one would need only to reject the redundant equations and solve the reduced set that remains. In practice, however, the data are erroneous, and equations (5.4) or (5.7) are formally inconsistent. One could still reject redundant data, but that, of course, would be unwise. Retention of redundancy is always important under these circumstances for reducing the influence of random errors. Therefore some kind of averaging procedure is required.

The second problem concerns error magnification. In addition to formal redundancy it is usually the case that there are different combinations of the data that give nearly but not strictly the same information. This leads to the matrix $A_{ij}$ being ill-conditioned, or nearly singular. Once again, one wishes to average in some way the almost identical information that is contained in the two or more combinations; that is analogous to what must be done for the genuinely redundant data. But in addition one must reject the apparent information contained in the difference between the almost equivalent combinations, for that is dominated by the errors in the data.

The second problem is analogous to trying to measure a vector $y$ by making independent measurements of its components ($a_1, a_2$) in the directions of the unit vectors $y_1$ and $y_2$, which are known to be nearly parallel. Thus if
\[ \mathbf{y} = a_1\mathbf{y}_1 + a_2\mathbf{y}_2 = \frac{1}{2} (a_1+a_2)(\mathbf{y}_1+\mathbf{y}_2) + \frac{1}{2} (a_1-a_2)(\mathbf{y}_1-\mathbf{y}_2) \]  

(5.8)

and the measurements of \( a_1 \) and \( a_2 \) have errors \( \varepsilon_1 \) and \( \varepsilon_2 \), one cannot measure the component of \( \mathbf{y} \) in a direction roughly perpendicular to \( \mathbf{y}_1 \) and \( \mathbf{y}_2 \) (e.g., in the direction of \( \mathbf{y}_1 - \mathbf{y}_2 \)) if either of \( \varepsilon_1 \) or \( \varepsilon_2 \) has a magnitude comparable with the difference \( a_1 - a_2 \) between the two measurements. Even though formally the vectors \( \mathbf{y}_1 \) and \( \mathbf{y}_2 \) can be used as basis vectors for a plane, in practice the erroneous data provide information only along the line parallel to \( \mathbf{u}_1 = (1/2)(\mathbf{y}_1+\mathbf{y}_2) \). For practical purposes the component of \( \mathbf{y} \) parallel to \( \mathbf{u}_2 = (1/2)(\mathbf{y}_1-\mathbf{y}_2) \) is inaccessible. Thus the situation is essentially the same as if \( \mathbf{y}_1 \) and \( \mathbf{y}_2 \) were genuinely parallel, which is analogous to the first of the problems mentioned above.

The resolution of problems of this kind had already been discussed by Lanczos (1961). The procedure is to find that part of the subspace 0 that is inaccessible to the data by virtue of the errors, and to relegate it to the annihilator A. The solution \( \Omega \) is the component of the actual function that is in the appropriately diminished subspace \( 0' \) that remains. In terms of the simple vector analogy, one creates a new orthogonal basis \( (\mathbf{u}_1,\mathbf{u}_2) \) of the plane, and recognizes that one can measure only the component \( a_1 + a_2 \) in the direction of \( \mathbf{u}_1 \). Notice that in this analogy the result contains both measurements, so statistical errors are lower than had the second of the measurements, say, been rejected as being redundant and the resulting component estimated by \( 2a_1 \). The Lanczos inverse also has this property, providing a natural way of averaging data in the more complicated case when several different combinations provide similar information.

For the details and the justification of the Lanczos method the reader is referred to the book by Lanczos (1961), and the discussions in the geophysical context by Jackson (1972), Wiggins (1972) and Parker (1977a). The following summary follows Parker (1977a).

Let \( \sigma_1 \) be the standard errors of the data \( w_1 \). Then one weights the constraints (5.1) with \( \sigma_1^{-1} \), yielding

\[ w'_1 = \int_0^R K'_1(r)\Omega(r)dr, \]  

(5.9)

where \( w'_1 = w_1/\sigma_1 \) and \( K'_1 = K_1/\sigma_1 \), so that \( w'_1 \) has unit standard error. The matrix \( A'_{1j} \), defined in the same way as \( A_{1j} \) in equations (5.5) but with \( K_k \) replacing \( K_k \), is positive-definite and symmetric, and can be diagonalized with an orthonormal matrix \( U'_{1j} \) to give

\[ \sum_{k,l} U'_{k1} A'_{kl} U'_{lj} = \lambda'_1 \delta'_{lj} \]  

(5.10)

where \( \lambda'_1 \) are the positive eigenvalues of \( A'_{1j} \) and \( \delta'_{lj} \) is the Kronecker delta. Now consider the new basis \( \psi'_1(r) \) of \( 0 \) defined by

58
\[ \psi_i = \lambda_i^{-1/2} \sum_j u_{ij} k_j' \]  \hspace{1cm} (5.11)

which has the property
\[ \int_0^R \psi_i \psi_j \, dr = \delta_{ij} \]  \hspace{1cm} (5.12)

Regard the functions \( \psi_i \) to be ordered such that \( \lambda_i \) decreases with increasing \( i \), and expand the solution \( \Omega \) in terms of them:
\[ \Omega = \sum_i a_i \psi_i \]  \hspace{1cm} (5.13)

Of course the expansion excludes that part of \( \Omega \) in the annihilator A. In view of equations (5.12) and (5.9), the expansion coefficients are determined by
\[ \lambda_i^{1/2} a_i = \lambda_i^{1/2} \int_0^R \psi_i \omega \, dr = \sum_j u_{ij} w_j' \]  \hspace{1cm} (5.14)

Recalling that the \( w_j' \) have unit standard errors, it follows immediately from equation (5.14) and the orthonormality of \( U_{ij} \) that if the errors in \( w_j' \) are statistically independent the errors in \( a_i \) are also statistically independent, with standard deviation \( \lambda_i^{1/2} \). Thus the uncertainty of the coefficients \( a_i \) increases with decreasing \( \lambda_i \); and it is immediately evident from equation (5.14) that the uncertainty is total when \( \lambda_i = 0 \).

Notice that these statements incorporate the statistics of the errors in the data, as must be the case, because each \( \lambda_i \) depends on the \( \sigma_j \).

The final step in the procedure is to truncate the expansion (5.13), thereby effectively relegating to A the subspace spanned by the eigenfunctions \( \psi_i \) that correspond to small eigenvalues \( \lambda_i \). It can easily be shown that these functions contribute little to the integrals in the constraints (5.9) compared with their contribution to the sum (5.13). It is necessary to decide where (and how) the expansion is to be truncated, and here several options are available (e.g. Wiggins, 1972; Jackson, 1973). In the following section I illustrate a pragmatic approach, which appears to be reliable when systematic errors in the data can be ignored. Moreover, it is particularly useful when the magnitude of the random errors in the data are poorly estimated. It rests on the fact that as \( \lambda_i \) decreases the functions \( \psi_i \) tend to develop more and more small-scale structure and the coefficients \( a_i \) increase (as \( \lambda_i^{1/2} \)) in a random way once they are dominated by errors. If one increases from unity the number of terms retained in the expansion (5.13), the result should first approach the correct solution. Then, once errors dominate, successive approximations diverge, exhibiting structure on smaller and smaller scales with larger and larger amplitude. The best estimate of \( \Omega \) one can obtain by the method is the function to which the expansion appears to be converging before the divergence takes over.
Extracting localized averages

The idea of finding localized averages was introduced by Backus and Gilbert (1968), partly as a means of assessing what they call the resolving power of the data: namely, the degree to which information about \( \Omega \) that is contained in the data can be localized in space.

To begin, the constraints (5.1) are rescaled by dividing each by \( \int_0^R K_i(r)dr \), yielding

\[
w_i^* = \int_0^R K_i^*(r)\Omega(r)dr , \tag{5.15}
\]

where \( K_i^* \) is unimodular: i.e.

\[
\int_0^R K_i^*(r)dr = 1 . \tag{5.16}
\]

Thus \( w_i^* \), which is an observable quantity, is an average of \( \Omega \). The idea now is to seek, for a certain value \( r_0 \) of \( r \), a set of coefficients \( \beta_i(r_0) \) such that in the linear combination

\[
\sum_i \beta_i(r_0)w_i^* = \int_0^R D(r_0,r)\Omega(r)dr \tag{5.17}
\]

the averaging kernel

\[
D(r_0,r) = \sum_i \beta_i(r_0)K_i^*(r) \tag{5.18}
\]

resembles a Dirac delta function centred at \( r = r_0 \). One can assess the degree of success by inspecting \( D \) to see how localized it is. A quantitative measure of the localization can be obtained from the spread, defined by Backus and Gilbert (1970) as

\[
s(r_0;D) = 12 \int_0^R (r-r_0)^2D^2(r_0,r)dr , \tag{5.19}
\]

which, as Backus and Gilbert show, can be large either if \( D \) is not well localized or if \( r_0 \) is far from center \( \overline{r} \) of a well-localized \( D \), where

\[
\overline{r} = \frac{\int_0^R rD^2(r_0,r)dr}{\int_0^R D^2(r_0,r)dr} . \tag{5.20}
\]

The factor 12 is chosen in (5.19) because if \( D = \delta^{-1} \) for \( |r-r_0| < \delta/2 \) and \( D = 0 \) otherwise, then \( s(r_0;D) = \delta \). For any given \( D \), \( s(r_0;D) \) is minimized when \( r_0 = \overline{r} \). That minimum, \( \delta \equiv s(\overline{r};D(r_0,r)) \), is the width of \( D \), and measures the degree to which \( D \) resembles a delta function. Alternative measures have also been used.
The average (5.17) of \( \Omega \), which I denote by \( \bar{\Omega}(r) \), is a useful diagnostic of the function \( \Omega \). Indeed, as Backus and Gilbert (1968) point out, it is only this that is in any sense determined by the data. Thus one can regard \( \bar{\Omega} \) as an estimate of \( \Omega \), bearing in mind that it is really a smoothed version obtained by averaging over a characteristic distance \( \delta \).

The determination of the coefficients \( \beta_i(r_0) \) is discussed by Backus and Gilbert (1968, 1970). One chooses a function \( J(r_0,r) \) which vanishes when \( r = r_0 \) and increases monotonically away from \( r_0 \), and one minimizes the functional

\[
\Delta = \int_0^R JD^2dr,
\]

subject at least to the constraint

\[
\int_0^R D(r_0,r)dr = 1.
\]

When \( \Delta \) is small the constraint (5.22) forces \( D \) to be small where \( J \) is large, and permits \( D \) to be large near \( r = r_0 \) where \( J \) vanishes.

In an application of this procedure to artificial high-degree solar p-mode data (Gough, 1978b) it was found that the final result was not particularly sensitive to the form chosen for \( J \), provided it rose steeply enough far from \( r_0 \). I shall not discuss that issue further here. All the examples illustrated in the following sections where computed with \( J(r_0,r) = 12(r-r_0)^2 \), which is the case that Backus and Gilbert (1970) discuss in detail when they consider erroneous data. Then

\[
\Delta = s = \sum_{ij} S_{ij}^* \beta_i \beta_j
\]

where

\[
S_{ij}^* = 12 \int_0^R (r-r_0)^2 K_i^*(r)K_j^*(r)dr.
\]

In practice, errors in the data cause the minimization of \( \Delta \) subject to the constraint (5.22) not to provide a good measure \( \bar{\Omega} \) of the average of \( \Omega \). The reason is that to obtain the most concentrated kernel \( D \) requires coefficients \( \beta_i \) with large magnitudes. The constraint (5.22) requires that \( \beta_i \beta_i = 1 \). Severe cancellations are therefore required in that sum, and in the sum on the left-hand side of equation (5.17). Cancellation in the latter does not actually take place when the data \( w_i \) contain errors.

Backus and Gilbert (1970) assume that an estimate \( \hat{E}_{ij} \) of the covariance matrix of the errors in the data \( w_i \) is known, so that one can estimate the error \( \epsilon \) in \( \bar{\Omega} \):
\[ \varepsilon^2 = \sum_{ij} E_{ij} \beta_i \beta_j \]  

(5.25)

The idea then is to set a limit \( \varepsilon_0 \) on the error \( \varepsilon \), and minimize \( s \) subject to the constraint \( \varepsilon \leq \varepsilon_0 \) and equation (5.22). The result depends on the choice of \( \varepsilon_0 \). Alternatively one can set a limit \( s_0 \) on \( s \), and minimize \( \varepsilon \) subject to \( s \leq s_0 \) and the constraint (5.22).

As Backus and Gilbert prove, the two formulations are essentially equivalent. They point out that \( S^*_{ij} \) and \( E_{ij} \) are both positive definite symmetric matrices, so that \( s = s_0 \) and \( \varepsilon = \varepsilon_0 \) each define hyperellipsoids in the parameter space spanned by \( \beta_i \). The intersections of these hyperellipsoids with the hyperplane \( \xi \beta_i = 1 \) are also hyperellipsoids, which I denote by \( s^\dagger = \text{constant} \) and \( \varepsilon^\dagger = \text{constant} \). In general, \( E_{ij} \) is not a scalar multiple of \( S^*_{ij} \), so the centres of the ellipsoids are not coincident. Consequently, for any given value \( \varepsilon_1 \leq \varepsilon_0 \) of \( \varepsilon \), \( s \) is minimized when the ellipsoid \( s^\dagger = \text{constant} \) (say \( s_2 \)) is tangent to the ellipsoid \( \varepsilon^\dagger = \varepsilon_1 \), and the solution \( \beta_i \) is the point of contact. The smallest value, \( s_0 \), of \( s \) is obtained when that point is as close to the centre of the ellipsoid \( s^\dagger = s_2 \) as is permissible, which implies that \( \varepsilon_1 \) is as large as is permissible: namely, \( \varepsilon_1 = \varepsilon_0 \). The argument is clearly symmetric; \( \varepsilon \) is minimized for \( s = s_1 \leq s_0 \) when the ellipsoid \( \varepsilon^\dagger = \varepsilon_2 \) is tangent to the ellipsoid \( s^\dagger = s_1 \), and the smallest permissible value \( \varepsilon_0 \) of \( \varepsilon \) is achieved when \( s_1 = s_0 \). The conditions for tangency are:

\[
\sum_j (A^*_{ij} + \mu E_{ij}) \beta_j = \nu \quad , \quad (5.26)
\]

\[
\sum_j S^*_{ij} \beta_i \beta_j = s_0 \quad , \quad (5.27)
\]

\[
\sum_i \beta_i = 1 \quad , \quad (5.28)
\]

where \( A^*_{ij} \) is defined as in equation (5.5) with \( K_k \) replaced by \( K^*_k \). Equations (5.26)-(5.28) are to be solved simultaneously for the coefficients \( \beta_i \) and the unknown parameters \( \mu \) and \( \nu \). The solution is determined uniquely in terms of \( s_0 \) with \( \mu > 0 \), provided \( s_0 \) is in its allowable range. Evidently \( s_0 \) cannot be less than its minimum value obtained by ignoring errors; it is also bounded above, as can be appreciated from the definition (5.19) when it is recognized that the magnitude of \( D \) is bounded for finite values of \( \varepsilon \).

In practice it is simpler to solve the problem implicitly, by choosing \( \mu \) and solving for \( \beta_i \), \( \nu \) and \( s_0 \). Backus and Gilbert set \( \mu = w \tan \theta \) and \( \nu = b \sec \theta \) with \( 0 < \theta < \pi/2 \), and \( w > 0 \) chosen, for convenience, to make \( S^*_{ij} \) and \( w E_{ij} \) of comparable numerical size. Then they introduce

\[
W_{ij}(\theta) = S^*_{ij} \cos \theta + w E_{ij} \sin \theta \quad . \quad (5.29)
\]

Equation (5.26) now takes the form

62
\[
\sum_{j} W_{ij} \delta_{j} = b
\]  
(5.30)

The matrix \( W_{ij} \) is symmetric and positive definite, and therefore has a positive definite inverse \( W_{ij}^{-1} \). Therefore any chosen \( \Theta \) equations (5.30) may be solved, whence \( s_0 \) and \( \epsilon^2 \) may be evaluated from equations (5.27) and (5.25).

The parameter \( \Theta \) determines the extent to which errors are to be restricted, at the expense of permitting \( D \) to be less well confined. When \( \Theta = 0 \) the errors are ignored, and when \( \Theta = \pi/2 \) no attempt is made to localize \( D \). The aim is to find a tradeoff somewhere between.

The choice of \( \Theta \) is discussed by Backus and Gilbert (1970). As with the spectral method, the pragmatic approach is useful, especially when the estimate \( E_{ij} \) of the covariance matrix is uncertain. This is illustrated in the accompanying paper (Christensen-Dalsgaard and Gough, 1984a).

Examples of the application of the localized averaging procedure to real solar data are presented by Gough (1982) and Hill et al. (1984).

6. PHASE THREE: SOME NUMERICAL EXAMPLES

The principal object of this section is to illustrate an application of the spectral method to artificial solar data. A similar application of the localized averaging procedure is discussed in the accompanying paper (Christensen-Dalsgaard and Gough, 1984a).

The example I present in some detail concerns the determination of the horizontal component of the large-scale subphotospheric flow using high-degree \( f \) and \( p \) modes. Thus the data mimic those discussed by Deubner et al. (1979), Deubner (1983), Rhodes et al. (1983a) and Hill et al. (1983).

A model of the solar envelope was constructed in the usual way by integrating the equations of stellar structure inwards from the photosphere. The procedure was identical to that used by Hill et al. (1984). Upon this was imposed an artificial horizontal equatorial velocity \( \Omega \) in excess of the surface rotation, which is plotted in Figures 2-4. The form of the function \( \Omega \) was chosen to have several different features to test what the method could resolve: a gentle growing oscillation at low values of \( x = \log p \), followed by three straight portions with corners, chosen such that \( \Omega \) increases at great depth. Straight portions were chosen because the \( K_\ell \) are not straight, and corners test the spatial resolution of the procedure. Otherwise the choice was quite arbitrary.

For the purpose of the inversion 45 five-minute sectoral modes were chosen, distributed approximately uniformly along the 9 lowest ridges in the \( k-\omega \) diagram. The modes are indicated by crosses in Figure 1. The adiabatic eigenfrequencies \( \omega_0 \) were computed in the usual way, and the
Fig. 1.  \( k-\omega \) diagram of prograde sectoral modes computed from the observations of Hill et al. (1983). The crosses indicate the modes that were used in the inversions displayed in Figures 2-4.

Perturbations \( \omega_1 \) produced by the artificial function \( \Omega \) were computed from the eigenfunctions using equation (4.5). The modes considered are all concentrated in the outer layers of the Sun; therefore it was adequate to compute only in the region \( r > 0.5 \, R \).

Three inversions were performed. The first used the raw frequencies; the other two used data constructed by adding independent Gaussian-distributed noise to the frequency perturbations \( \omega_1 \). The same set of modes was used in all three cases.

In Figure 2 are displayed the expansions (5.13) truncated at four different points. They include \( I = 9, 15, 20 \) and 33 basis functions, with eigenvalues satisfying \( \lambda_4/\lambda_1 > 3 \times 10^{-2}, 10^{-2}, 10^{-3} \) and \( 10^{-8} \). As expected, convergence generally improves as the number of modes is increased. If even more modes are included the accuracy of the expansion deteriorates slightly, presumably as a result of rounding errors. At depths greater than \( x = 12.5 \), the most deeply penetrating oscillation eigenfunctions and their associated kernels \( K_4 \) drop to zero, and the modes cannot sense \( \Omega \). The inferred velocity falls to zero too.
Fig. 2. Inversions of error-free frequency splittings $\omega_{ij}$ of the 45 modes indicated in Figure 1. Both the results of the inversions and the imposed horizontal velocity $\Omega$ from which $\omega_{ij}$ were computed are shown; the latter is easily identifiable because it is common to all the panels. The eigenvalues $\lambda_i$ retained in the expansions represented in panels (a)-(d) satisfy $\lambda_i > \lambda_m$, where $\lambda_m/\lambda_1 = 3 \times 10^{-2}$, $10^{-2}$, $10^{-3}$ and $10^{-8}$ respectively; corresponding numbers $I$ of basis functions $\psi_i$ are 9, 15, 20 and 33. Since the problem is linear, the ordinate scale is arbitrary. The independent variable is $x = \log_{10} p$. 

65
In Figure 3 are inversions of data with errors having standard deviations of 5 per cent of the raw values. The first two inversions are hardly distinguishable from the corresponding inversions with error-free data. The errors have been averaged to some degree by the inversion procedure, as a result of combining redundant information, leading to an inferred velocity with an error generally rather less than 5 per cent of the mean magnitude of $\Omega$. Evidently the error that is present is a result of imperfect resolution resulting from retaining no more than 15 basis functions. When 20 basis functions are included (Figure 3c), however, the errors are beginning to become important, and erroneous small-scale structure is introduced into the inferred velocity. Presumably this would not have happened at this level of truncation had substantially more than 45 data been included in the inversion. Figure 3d includes 26 modes, which have $\lambda_4/\lambda_1 > 10^{-5}$; now erroneous rapid oscillations dominate the expansion, though the general behaviour of $\Omega$ is still discernable.

The final example, which is illustrated in Figure 4, was computed from data with 50 per cent errors. Even though the error in the inferred velocity is as great as 100 per cent in places, the inverted data still retain the basic features of $\Omega(r)$. When 20 basis functions are used the characteristic wiggles similar to those in Figure 3d are introduced.

The conclusion to be drawn is that the spectral method is likely to be useful for solar inversions to infer the large-scale subphotospheric velocity (including rotation).

As a secondary example I present the results of a single iteration of a density inversion. This work was carried out some years ago in collaboration with A. J. Cooper (unpublished), to ascertain whether it was likely to be possible to infer the distribution of the density of the Sun from whole-disk data taken from a spacecraft. Two simple models of the Sun were constructed, each consisting of an inner polytrope of index 3 to represent the radiative interior and a polytropic exterior of index 1.5 to represent the convection zone. The transition in one case was at $x \equiv r/R = 0.75$, and in the other it was at $x = 0.80$. Eigenfrequencies of the two models were computed for low-degree modes of relatively low order, assuming $\gamma = 5/3$ throughout. One set was considered to be the observations; the other set was the theoretical eigenfrequencies of a trial solar model. No errors were added. Inversions of equation (4.4) were carried out using both the spectral expansion and the localized averaging procedure to try to deduce the density distribution of the model that provided the artificial data. The fact that it had been constructed from polytropes it not important since that information was not available for the inversion: only the mass, the radius and the eigenfrequencies were used. The constraint of preserving the mass is simply of the form (4.4), with the left-hand side replaced by zero and with $S = 1$; it was handled in the same way as the frequency constraints.

The results of the two pairs of inversions are shown in Figures 5 and 6, where the inferred density differences $\delta \rho$ between the trial and
Fig. 3. Inversions of frequency splittings to which independent Gaussian distributed errors with 5% standard deviation have been added. The lower bounds $\lambda_m$ to the eigenvalues retained in panels (a)-(d) are given by $\lambda_m/\lambda_1 = 3 \times 10^{-2}, 10^{-2}, 10^{-3}$ and $10^{-5}$ respectively. Note the difference in the ordinate scale in panel (d).
Fig. 4. Same as Figure 3, except that 50% errors have been added to the frequency splittings and that $\lambda_m/\lambda_1 = 10^{-1}, 3 \times 10^{-2}, 10^{-2}$ and $10^{-3}$. The expansion with $\lambda_m/\lambda_1 = 10^{-1}$, which includes only four terms, is similar in form to similar truncations with 10% or no errors, but its amplitude is roughly twice as great.
Fig. 5. The continuous line is the relative density difference $\delta \rho/\rho$ between the 'true' artificial model of the Sun and a trial. The dashed line represents an estimate of $\delta \rho/\rho$ from the spectral expansion obtained from a single iteration from the trial by the procedure outlined in the text. The filled circles are corresponding localized averages: no well-localized kernels could be constructed with centres at radii smaller than 0.3 R. Dipole ($\ell = 1$) and quadrupole ($\ell = 2$) $p$ modes of orders 1-9 were used in both inversions.

Fig. 6. Same as Figure 5, but using modes $g_4-p_5$ ($\ell = 1$) and $g_3-p_5$ ($\ell = 2$).
the artificial Sun are compared with the actual difference. Both used just 18 modes. The two inversions in Figure 5 were performed with the 9 lowest-order \( p \) modes with \( \ell = 1 \) and \( \ell = 2 \). The results are not very good, though the general trend of the exact curve is exhibited by the inversions. The poor quality of the inversion would have been evident without prior knowledge of the correct \( \delta \rho \), partly because in most places it was not possible to produce nicely localized optimal kernels \( D \), particularly near the centre of the star, and partly because on the whole the two inversion techniques gave different results.

When some of the \( p \) modes are replaced by the quadrupole \( f \) mode and some low-order \( g \) modes, the situation is quite different. Now sharply peaked kernels \( D \) can be produced, and the two methods both yield very similar and quite accurate results. These are shown in Figure 6.

We have also performed some experiments with other groups of modes, and with data containing errors. The results of that investigation will be reported elsewhere.

7. OTHER INVERSION METHODS

Other criteria can be considered for selecting from the infinity of functions \( \Omega \) that satisfy the constraints (5.1). For example, one can choose the smoothest or the flattest function by minimizing

\[
\int_{0}^{R} \left( \frac{d \Omega}{dr} \right)^2 dr \quad \text{or} \quad \int_{0}^{R} \frac{d \Omega}{dr} \frac{d \Omega}{dr} \quad (7.1)
\]

subject to the constraints (5.1). The latter has been used by Gough (1982, 1984) with the solar data reported by Hill et al. (1982) and Scherrer and Delache (1984). In view of the fact that functions in the annihilator \( A \) of (5.1) tend to be rapidly varying, one would expect these procedures to give results similar to the spectral expansion.

Another procedure that I must mention is that described by Ulrich et al. (1979). It was the first to be used on real solar data (Deubner et al. 1979). Despite the fact that it was demonstrated to be internally inconsistent, yielding unfaithful results in a particular artificial case (Gough, 1978b), it behooves me to discuss it further because Rhodes et al. (1983b) plan to use it again on future observations to measure the solar rotation.

The procedure appears to be based on the incorrect assumption that the kernels \( K^* \) are highly localized. The essence was to replace equation (5.1) by

\[
w_{ii} = \Omega(r_{ii}) \quad , \quad (7.2)
\]

where \( r_{ii} \) measured the "effective depth" of \( K^* \). The definition adopted for \( r_{ii} \) can easily be shown to be equivalent to the first moment of \( K^* \), though Ulrich et al. (1979) did not describe it in those terms. As was
pointed out by Gough (1978b), the procedure would be adequate if \( \Omega \) were a linear function of \( r \), for then equation (7.2) is correct even though \( K^*_1 \) is not localized. Nevertheless, Deubner et al. found empirically that the \( \Omega \) is not a linear function. Subsequent observations by Rhodes et al. (1983a) did not support the finding that \( \Omega \) varies with depth, which led Rhodes et al. (1983b) to suggest that inversion procedures of the type described here are not necessary. However, the more extensive observations of Hill et al. (1984) suggest that \( \Omega \) does vary, both in space and in time.

In support of the continued use of equation (7.2) Rhodes et al. (1983b) argue that it can be regarded as containing just the first two terms of a Taylor expansion of \( \Omega \). If that were so, Deubner et al. (1979) would have represented \( \Omega \) by a linear function of depth. Rhodes et al. (1983b) went on to say that, should future observations merit, higher terms could be included, after which a least-squares analysis could be carried out. Thus one would set

\[
\Omega = \sum_{k} \gamma_k z^k \quad \text{(7.3)}
\]

where \( z = R-r \), and determine \( \gamma_k \) by minimizing

\[
\chi^2 = \sum_{i} \left[ \sum_{k} \gamma_k (z_{ik})^k - \omega_{i} \right]^2 \quad \text{(7.4)}
\]

where

\[
z_{ik} = \left( \int_{0}^{R} z^k K^*_1 dz \right)^{1/k}
\]

which are essentially the higher-order effective depths of Ulrich et al. (1979); the coefficients \( \mu_i \) are constants that weight the data according to the errors (e.g. \( \mu_i = \sigma_i^{-2} \)).

This procedure has already been carried out (Gough, 1982) on the rotational splitting data of Hill et al. (1982) and Claverie et al. (1981) using polynomials of up to eighth degree, and I have shown (unpublished) it to work quite well on artificial data (when it can be reliably tested) provided the assumed variation of \( \Omega \) is on a scale large enough to be resolved by a modest number of terms. However, it does not work when \( \Omega \) varies on a smaller scale. This is illustrated in Figure 7, where polynomials of up to sixth degree are fitted to the data that were inverted by the spectral method to produce Figures 2 and 3. None of the polynomial approximations fits the original curve particularly well. Moreover, without a knowledge of the original function \( \Omega \) from which the \( \omega_i \) were computed, there is no way of knowing where the polynomial approximations are least good. The failure to converge, however, does indicate that the results cannot be trusted. If the degree of the polynomial is increased further, or if substantially larger errors are added to the data, the mismatch is aggravated severely. Of course, one can always find a smooth function whose values are arbitrarily close to any angular velocity \( \Omega \) for which the radius of convergence of a power series
Fig. 7. Polynomial approximations to the same horizontal velocity $\Omega$ as was used in the inversions in Figures 2-4, computed by $\chi^2$ minimization as described in the text. Panels (a) and (b) use the same data as were used for Figures 2 and 3 respectively. The degrees of the polynomials are indicated in the figure.

representation is greater than the range of $r$ over which $\Omega$ can be measured, and so one might think that this method could be used with many more terms. However, the procedure is then unstable to small errors in the data, and in any case the method is impractical because the matrix that must be inverted to compute the coefficients of the power series is ill conditioned.

My conclusion is that there is little point in using power series approximations in cases when the constraints can be written as linear integral equations of the type (5.1). However, in more complicated circumstances it may not be unwise to try it, because the expansion is usually simple to carry out. If one is fortunate enough to obtain a stable apparently convergent representation, then the problem is probably solved. But if the problem is expressible in the form (5.1), it seems to be expedient to expand in the more natural basis functions $\psi_i$.

8. DISCUSSION

The conclusion to be drawn from this and the accompanying article (Christensen-Dalsgaard and Gough, 1984a) is that both the Gilbert-Backus optimal localized averaging procedure and the spectral expansion are likely to be useful for inverting solar oscillation frequencies. It has been shown that the method should certainly work given accurate knowledge of the rotational splitting of modes that have already been
detected observationally, and evidence has been presented to suggest that it should work for inferring the Sun's pressure and density distribution too.

I should stress however, that the more primitive assessments of Phase Two of the inversion will continue to have some value. I have already described how model-fitting calculations have given us estimates of the Sun's initial helium abundance and the present depth of the convection zone, and how an appreciation of the asymptotic structure of high-frequency p modes has permitted us to infer that at least part of the cause of the unresolved discrepancy between theory and observation lies in the convection zone, or near its base. It is in the convection zone that there are substantial uncertainties in solving the forward problem, for there nonadiabatic effects and Reynolds stresses may play a significant role. It is also in the convection zone that the equation of state appears to be most uncertain. Shibahashi et al. (1983, 1984) and Ulrich (1982) have studied the influence of electrostatic interactions on the equation of state, using the Planck-Larkin cutoff, and have assessed how they effect the eigenfrequencies (see Ulrich and Rhodes, 1983). Christensen-Dalsgaard and Gough (unpublished) have made similar studies, using variants of the theory used by Fontaine et al. (1977); internal partition functions computed both from the static screened Coulomb potential, as was first used in oscillation calculations by Berthomieu et al. (1980), and from the confined atom model, using the programme described by Däppen (1980), together with estimates of the additional degree of ionization induced by time-dependent perturbations to the bound-state potential and the influence of neighbouring neutral species, have all been considered. In no case has adequate agreement between observation and theory been achieved. Whether this implies that current procedures for calculating the equation of state are woefully inadequate or whether it implies that there is something else wrong with our models of the Sun is yet undetermined. It may be that an application of the inversion procedures of Phase Three will give us the next clue.

It may also require Phase Three methods to resolve the apparent discrepancy raised by Scherrer and Delache (1984) and van der Raay et al. (1984), who claim respectively that the mean normalized period interval

\[ \left( \frac{P_{n+1, \ell} - P_{n, \ell}}{\ell(\ell+1)} \right)^{1/2} \]

(where \( P_{n, \ell} \) is the period of the mode of order \( n \) and degree \( \ell \)) between dipole and quadrupole g modes in the period range 3-5 hr is 38.6 min and 41.2 min, whereas the mean theoretical separation of the corresponding modes of Christensen-Dalsgaard's standard Model 1 of the Sun is only 34.5 min. This discrepancy may not be real, however, because the oscillations observed may have been misidentified.

Formal inversions of the limited rotational splitting data from low-degree modes that have appeared in the literature (Claverie, et al. 1981; Hill et al. 1982; Scherrer and Delache, 1984) have already been
carried out (Gough, 1982, 1984; Hill et al. 1982; Campbell et al., 1983), but, even if the data have been correctly interpreted, they are too sparse for a well resolved angular velocity to be inferred. This has led to some controversy over the value of the gravitational quadrupole moment \( J_2 \) of the Sun that is induced by the centrifugal forces, and its implications regarding tests of theories of gravitation, such as general relativity, from the precession of the perihelion of Mercury. It is undeniable that the splitting data, coupled with the planetary orbit data, do not contradict general relativity (Gough, 1982; Campbell et al. 1983). However, both H. Hill et al. (1982) and Campbell et al. (1983), using a particular kind of coarsely resolved analysis, have concluded that \( J_2 \) is substantially greater than the minimum value permitted by the data, thereby challenging the analysis of the orbit data or supporting alternative theories of gravitation. The estimate \( \tilde{\Omega} \) of the localized averaging procedure, however, is not in conflict; moreover the resolving width is not everywhere smaller than the scale of variation of the \( J_2 \) kernel \( J \) [see Figure 13 of Christensen-Dalsgaard and Gough (1984a)] and \( J_2(\tilde{\Omega}) \) does not necessarily underestimate the true \( J_2 \) (Gough, 1982). It is important to realize that the conclusion that there is a conflict rests crucially on an additional assumption, which is not suggested by the data and which results in both the restriction of the space \( \Omega \) of acceptable functions \( \Omega \) by the imposition of unsubstantiated (and incompletely defined) constraints and the inclusion of components from the annihilator \( \Lambda \). This would be a perfectly acceptable procedure if the restrictions on \( \Omega \) were stated clearly, for then the plausibility of the conclusion could more easily be judged by the casual reader. An advantage of an inversion procedure such as the spectral expansion is that the truncation of (5.13) (or the inclusion of functions from \( \Lambda \)) makes one acutely aware of the assumptions one has made. What is really necessary, however, is a much larger data set, with substantially greater resolving power.

An inversion of frequency splitting of high-degree modes observed by F. Hill et al. (1983) has been undertaken by F. Hill et al. (1984), using the localized averaging technique of Backus and Gilbert (1970). The results suggest variations in the large-scale subphotospheric velocity that may be associated with giant convective cells. Aside from using asymptotic approximations to the frequencies (Gough, 1984), no (Phase Three) inversions of real solar data have yet been performed to obtain the density stratification of the Sun.

What are the differences between the two principal inversion methods I have discussed? Were the data to be free from error, the minimum value of \( \lambda \) that is retained in spectral expansions can, at least in principle, be adjusted to make the solution satisfy the constraints (5.1) as accurately as one pleases. This is not so of the localized averaging technique, for even if errors are ignored, there is always a minimum resolution width of the optimal kernels. Therefore adopting \( \tilde{\Omega}(\tilde{r}) \) as a representation of \( \Omega(\tau) \) can never be exact. For this reason some geophysicists (e.g. Parker, 1977a) prefer not to use the localized averages for inversions, favouring simply a particular example, such as is provided by the spectral expansion, of the infinity of solutions that
satisfy the constraints (5.1) within the limits set by the precision of the data. As is evident from the examples illustrated in Figures 2-4, those solutions have unreal small-scale structure. In interpreting the spectral representations one's eye should smooth out the wiggles, by imagining an average over the resolving width of the optimal localized kernels. That average is provided by \( \tilde{\Omega}(\tau) \), however, which approximates the average of all acceptable solutions to the constraints (5.1), including those that have components in the annihilator \( A \). It is important to realize that \( A \) is determined by the limitations of the observations, and has no other physical significance. Therefore, as Backus and Gilbert (1967) point out, it is only \( \tilde{\Omega} \) that is a valid representation of the true function \( \tilde{\Omega} \), though it is necessary to recognize that \( \tilde{\Omega} \) may have small-scale structure not present in \( \tilde{\Omega} \) that cannot be resolved by the data. What the representation \( \tilde{\Omega} \) means can be assessed by inspecting the optimized kernels \( D \).

One should appreciate also that it can be dangerous to use \( \tilde{\Omega} \) for estimating quantities that are nonlinear functions of \( \Omega \). An example is the gravitational quadrupole moment \( J_2 \), which is an integral of \( \Omega^2 \) weighted with the nonnegative function \( \Psi \) displayed in Figure 13 of Christensen-Dalsgaard and Gough (1984a). The rectification of any small-scale oscillatory component of \( \Omega \) that is not present in \( \tilde{\Omega} \) adds a positive contribution to the estimate \( J_2(\tilde{\Omega}) \) of the moment, which tends to make \( J_2(\tilde{\Omega}) \) an underestimate. In addition the smoothing produced by the averaging causes \( \tilde{\Omega} \) to be systematically above or below the true functions \( \Omega \) in regions where the curvature of \( \tilde{\Omega} \) is positive or negative.

An issue that to some is of considerable importance concerns the computing resources required to carry out the two techniques. The computation of the spectral expansion itself is dominated by calculating the large matrix \( A_{ij} \) and, to a lesser extent, finding its eigenvectors and eigenvalues for constructing the matrix \( U_{ij} \). For each inversion, this is done once. In contrast, the construction of the localized kernels \( D(r_0,r) \) involves a similar amount of work to that required for the entire spectral expansion at each point \( r_0 \). Strictly speaking one should compute \( D \) to determine the resolving power of the data, in order to gauge the reliability of the spectral expansion. But since that takes much longer than the expansion itself, one might be tempted not to.

Another difference between the two techniques is that the spectral expansion always provides a solution, whereas the localized averaging technique may not. The solutions by spectral expansion in Figures 2-4 extend over the entire domain, including the region \( x > 12.5 \) about which the data contain no useful information and where the spectral expansion is incorrect. Of course one could have easily inferred that the expansion cannot be trusted at great depths by inspecting the kernels \( k_j \) and noticing that there they are all zero. However, it is less obvious that the representation is poor when there is a region where all the kernels have a similar functional form but are not zero. It is for these regions particularly that the resolving power must be established. Thus one might consider an apparent advantage of the optimal localized averaging procedure to be that it does not always work, either by
failing to provide a sharply localized kernel centred about a desired value of \( r \), or even by failing to produce any suitable kernel at all. Of course a meticulous worker, not hampered by limitations of computer resources, would certainly compute the resolving power, in which case the advantage would not be real. However, it is then a trivial matter to compute \( H \), so it seems unwise not to do so. It is evident that the spectral method and the averaging technique complement one another, to the extent that it would often be expedient to use both.

Finally I must mention that it is important to be able to identify in the observations which modes correspond to the oscillation frequencies that have been measured. In the past the identification has often been made by indirect methods, and indeed that must always be so to some extent because the order \( n \) can never be measured directly. We know that in some cases, however, it is not necessary to have a precise identification. A systematic study of when that is so has never been undertaken.

ACKNOWLEDGMENTS

The splitting kernels \( K_1 \) used in the spectral expansion reported in this article were taken from the set used by Hill, Gough and Toomre (1984) to invert real solar data. The density inversions were carried out in collaboration with A. J. Cooper. I am very grateful to A. J. Cooper, F. Hill and J. Toomre, and to J. Christensen-Dalsgaard, A. Dziewonski and K. Whaler, for useful discussions.

REFERENCES

Christensen-Dalsgaard, J. and Gough, D. O. 1984a: 'Rotational inversion from global solar oscillations,' these proceedings
Christensen-Dalsgaard, J. and Gough, D. O. 1984b: 'Implications of observed frequencies of solar \( p \) modes,' these proceedings

76
Claverie, A., Isaak, G. R., McLeod, C. F., van der Raay, H. B. and Roca
Astrophys., 72, 177-185
J. Suppl., 35, 293-358
chromosphere and corona (eds. R. M. Bonnet and P. Delache, G. de
Bussch, Clairmont-Ferrand), pp. 3-36
Rösch, CNRS, Paris), pp. 81-103
Gough, D. O. 1983a: in Pulsations in classical and cataclysmic vari-
ables (eds. J. P. Cox and C. J. Hansen, JILA, Boulder), pp. 117-137.
Shaver, D. Kunth and K. Kjär, European Southern Observatory,
Garching), pp. 117-136
151-159
Harvey, J. W. and Duvall Jr., T. L. 1984: these proceedings
in press
425
1794-1797
Lamb, H. 1932: Hydrodynamics (Cambridge Univ. Press)
and nonlinear stellar pulsations (eds. H. A. Hill and W. A.
Dziembowski, Springer, Heidelberg), pp. 300-306
293-310

77
Sabatier, P. C. 1977: J. Geophys., 43, 115-137
Scherrer, P. and De la Che, P. 1984: Mem. Soc. astr. Italiana, in press