1994A&A...281..231P

Astron. Astrophys. 281, 231–240 (1994)

ASTRONOMY AND ASTROPHYSICS

The SOLA method for helioseismic inversion

F.P. Pijpers and M.J. Thompson

Astronomy Unit, Queen Mary and Westfield College, University of London, Mile End Road, London E1 4NS, U.K.

Received April 2, accepted May 12, 1993

Abstract. The Subtractive Optimally Localized Averages (SOLA) method is a versatile and efficient technique for inverting helioseismic data. The SOLA method is based on explicit construction of Backus-Gilbert averaging kernels, but whereas the more usual formulations of the optimally localized averages (OLA) method use a multiplicative penalty function to localize the kernels, the distinctive idea of SOLA is that one specifies a desired target form for the kernels and then minimizes the integrated squared difference between the kernels and the target form. This allows great versatility in the choice of target form, and furthermore SOLA has the significant advantage of being computationally more efficient than the usual OLA formulations. A Gaussian target function is a useful choice, and we use the example of determining the Sun's internal rotation to explore how the parameter values (such as the Gaussian's width) should best be chosen. Some alternatives to using a Gaussian function as target function are discussed and applied to artificial data in a blind experiment. In particular we show that it is possible to invert directly for the gradient of the rotation. This may be of interest if there are localized large gradients in the rotation rate.

Key words: Sun: oscillations of – Sun: rotation of – Sun: structure of – numerical methods

1. Introduction

In recent years observations of normal modes of oscillation of the Sun have increased enormously in both quantity and quality. Various large observational projects – notably GONG and MDI-SOI on board the SOHO satellite – are now in the final stages of planning. To make optimal use of the large sets of data that will be available, the algorithms for helioseismic inversion need to become efficient in handling several thousands of such modes simultaneously. Recently (Pijpers & Thompson 1992; hereafter PT) we presented the Subtractive Optimally Localized Averages (SOLA) method as an alternative to the usual formulations of the Optimally Localized Averages (OLA) inversion method of Backus & Gilbert (1968, 1970); see also e.g. Gough 1985. All OLA techniques are based on explicitly constructing averaging kernels by taking linear combinations of mode kernels. The usual formulations, which we call Multiplicative OLA (MOLA) formulations, involve minimizing the integral of the square of the averaging kernel multiplied by some function that penalizes the kernel for having large amplitude away from the intended location. The distinctive idea of the SOLA method is instead to minimize the integrated squared difference between the kernel and some chosen target form. In addition to the method's flexibility in terms of the target form that one chooses, a significant advantage of the SOLA method over the MOLA formulations is that it is computationally more efficient since it requires just one matrix inversion. The method is not generally as fast with large datasets as the regularized least squares method (e.g. Christensen-Dalsgaard et al. 1990; hereafter CDST) but it can be shown that the quality of the SOLA averaging kernels is better and that the systematic errors introduced by nonlocal effects in the inversion are therefore smaller. A recent comparison of the various methods currently in use is presented by CDST.

Although the SOLA method is described in PT, it is useful to lay out the method briefly in this paper, before proceeding to discuss choices of the free parameters of the method and alternatives for the target forms. The basic equations solved in the SOLA method are presented in Sect. 2. In Sect. 3 the possible choices for the free parameters with Gaussian target forms are discussed and a strategy for choosing parameter values is presented. In Sect. 4 a blind test is carried out and alternative forms of a target averaging kernel are explored. Our conclusions and ideas for future developments are presented in Sect. 5.

2. The SOLA method

In a spherically symmetric, non-rotating star, the frequency ω_{0nl} of a spheroidal mode of oscillation of radial order n and degree l is independent of the mode's azimuthal degree m. For a slowly rotating star the frequency is given by

$$\omega = \omega_{0nl} + m\omega_{1nl} , \quad (1)$$

where ω_{1nl} depends on the interior rotation rate Ω through

Send offprint requests to: M.J. Thompson

© European Southern Observatory • Provided by the NASA Astrophysics Data System

1994A&A...281..231P

$$\omega_{1nl} = \int_0^1 K_{nl}(x)\Omega(x)\mathrm{d}x + \epsilon_{nl} . \quad (2)$$

Here K_{nl} is the rotational kernel for the mode, presumed to be a known function of $x \equiv r/R$ (R being the surface radius), and ϵ_{nl} is the error in datum ω_{1nl}. To obtain the solar rotation rate at some radius r_0 the integral (2) must be inverted.

In OLA methods a set of coefficients $c_i(r_0)$ is constructed such that the *averaging kernel*

$$\mathscr{K}(x_0, x) \equiv \sum_{i\in\mathscr{M}} c_i(x_0)K_i(x) \quad (3)$$

is peaked around $x = x_0 \equiv r_0/R$ and is small everywhere else. Here $\mathscr{M}$ is the set of observed solar oscillation modes being used. For convenience we have introduced a single subscript i to represent each mode instead of the subscript pair (nl). An extra constraint on the averaging kernel $\mathscr{K}$ is that

$$\int_0^1 \mathscr{K}(x_0, x)\mathrm{d}x = 1 . \quad (4)$$

An estimate $\langle\Omega(x_0)\rangle$ of the rotation rate at $x = x_0$ is then obtained with this set of coefficients by defining

$$\langle\Omega(x_0)\rangle \equiv \sum_{i\in\mathscr{M}} c_i(x_0)\omega_{1i}$$

$$= \int_0^1 \mathscr{K}(x_0, x)\Omega(x)\mathrm{d}x + \sum_i c_i\epsilon_i . \quad (5)$$

The first term on the right of eq. (5) is a weighted average of Ω over the solar interior in which $\mathscr{K}$ is the weighting function. The other term is the error propagated from the data errors. The distinctive idea of the SOLA method is that one aims to make the averaging kernel resemble some chosen target form $\mathscr{T}(x_0, x)$, while at the same time moderating the effect of data errors, by minimizing

$$\int_0^1 \left[\mathscr{K}(x_0, x) - \mathscr{T}(x_0, x)\right]^2 \mathrm{d}x + \mu \sum_{ij} \mathbf{E}_{ij}c_ic_j \quad (6)$$

subject to the constraint (4). Here $\mathbf{E}$ is the error variance-covariance matrix of the observed frequencies. The MOLA formulation would be to minimize

$$\int_0^1 \left[\mathscr{K}(x_0, x)\right]^2 \mathscr{J}(x_0, x)\mathrm{d}x + \mu \sum_{ij} \mathbf{E}_{ij}c_ic_j . \quad (7)$$

Here $\mathscr{J}$ is the weighting function which is often chosen to be $\mathscr{J} = 12(x - x_0)^2$. Note that the definition of μ differs from previous usage. It is given by

$$\mu \equiv \mu_0 \left(\sum_{i\in\mathscr{M}} \mathbf{E}_{ii}/M\right)^{-1} , \quad (8)$$

where M is the total number of modes in mode set $\mathscr{M}$, and μ_0 is a trade-off parameter whose value may be chosen to determine the relative desirability of making the first and second terms in (6) small. Thus account is taken of the fact that the observed frequencies will not be error-free and that propagation of errors through the inversion should be moderated. This trade-off parameter has commonly been introduced as a factor $\tan\theta \equiv \mu_0$, where θ is the weighting parameter (e.g. PT). However, that notation seems needlessly complicated and there is further reason to avoid it because of the use of the symbol θ for the colatitude in the higher dimensional inversions.

The coefficients $c_i(x_0)$ are obtained by solving the matrix equation

$$\mathbf{A}\mathbf{c}(x_0) = \mathbf{v}(x_0) , \quad (9)$$

where the symmetric matrix $\mathbf{A}$ (of order $M + 1$) has elements

$$A_{ij} = \begin{cases} \int_0^1 K_iK_j\,\mathrm{d}x + \mu\,\mathbf{E}_{ij} & (i, j \le M) \\ \int_0^1 K_i\mathrm{d}x & (i \le M, \quad j = M + 1) \\ \int_0^1 K_j\mathrm{d}x & (j \le M, \quad i = M + 1) \\ 0 & (i = j = M + 1) \end{cases} \quad (10)$$

and vectors $\mathbf{c}$ and $\mathbf{v}$ are given by

$$\mathbf{c} = \begin{pmatrix} c_1 \\ \vdots \\ c_M \\ \lambda \end{pmatrix}, \quad \mathbf{v} = \begin{pmatrix} \int_0^1 K_1\mathscr{T}\mathrm{d}x \\ \vdots \\ \int_0^1 K_M\mathscr{T}\mathrm{d}x \\ 1 \end{pmatrix} . \quad (11)$$

Constraint (4) has been incorporated with Lagrange multiplier 2λ.

Ideally one would like to have each averaging kernel as close as possible to the target form

$$\mathscr{T}(x_0, x) = \delta(x - x_0) \quad (12)$$

(cf. Jeffrey 1988). However, it is unrealistic in general to expect such good localization. As discussed in PT, using a delta function as the target form gives poorly localized averaging kernels. Better results are obtained using a target with finite width. The specific target form $\mathscr{T}$ considered in PT was a Gaussian of width Δ:

$$\mathscr{T} = \frac{1}{f\Delta} \exp\left[-\left(\frac{x - x_0}{\Delta}\right)^2\right] . \quad (13)$$

Here f is a normalization factor ($\approx \sqrt{\pi}$ for small Δ) which is included to make the total integral of $\mathscr{T}$ equal to unity. This function approaches a delta function for small Δ. Of course one can choose other target forms, and we consider some of these in Sects. 4 and 5. In this and the next section, however, we shall use the Gaussian target function (GTF) given by Eq. (13) . In particular, it is our aim in Section 3 to show how the GTF parameter values might best be chosen.

The free parameter Δ should in general be a function of x_0. The choice of Δ at different radii x_0 is an important issue which

© European Southern Observatory • Provided by the NASA Astrophysics Data System

1994A&A...281..231P

will be discussed in the next section. One could also prescribe Δ to vary with x, but we shall not consider that possibility further here.

The c_i's are now obtained by putting the chosen $\mathscr{T}$ into Eq. (9) (through the vector **v**). Minimizing the error term and obtaining a well-localized kernel are opposing aims (e.g. CDST) and the parameter μ_0 is used to obtain a compromise between them.

The important point for the computational efficiency of the SOLA method is that matrix **A** is independent of x_0. Regardless of the form of $\mathscr{T}$, only the vector **v** has to be recomputed for each different x_0. In the more usual formulations of OLA (e.g. CDST) a matrix of order $M+1$ must be inverted at *each* target radius x_0. Both approaches involve the free parameter μ (or its equivalent). It may, however, be perceived as a disadvantage of the SOLA method with Gaussian target function that a second parameter must be chosen at each x_0. In the next section, therefore, we explore how the SOLA parameters should be chosen. In particular we demonstrate that a natural scaling of Δ with radius greatly reduces the number of free parameters.

First, though, we make a technical point about the way in which the SOLA calculations were performed. The construction of the coefficients **c** appears to require inverting a matrix whose order is essentially the number of modes. However, in practice the number of independent pieces of information contained by a large mode set in the presence of data errors is much smaller than the number of modes (e.g. Christensen-Dalsgaard et al. 1993). It has been pointed out by Christensen-Dalsgaard & Thompson (1993) that this redundancy can be exploited to reduce substantially the number of independent functions – they are combinations of modes rather than modes themselves – used to construct averaging kernels. The idea is basically to express the original mode kernels in terms of some basis of functions ϕ_j ($j = 1, \ldots, N$) and then to find the singular values of the matrix C whose elements are $\int K_i \phi_j dx$: the left singular vectors then define linear combinations of the original modes, and only those corresponding to the largest singular values are determinable from erroneous data. These most significant combinations, rather than the original mode kernels, are then used in the OLA matrix equation (9), which can substantially reduce the size of the matrix to be inverted. We have used this technique for all calculations in this paper, with the difference that it was the matrix with elements $\int K_i K_j dx$ to which the singular value decomposition was applied. This reduces the number of 'modes' from 834 (see below) to 66.

3. Choosing parameters with Gaussian target functions

In this section we consider SOLA averaging kernels constructed using the Gaussian target function [eq. (13)]. The mode set used is Set 1 of CDST, which comprises 834 p-modes in the frequency range of 2–4 mHz, with degrees between 1 and 200. The errors in the data are assumed to be uniform and uncorrelated so that the error variance-covariance matrix is

$$\mathbf{E}_{ij} = \sigma^2 \delta_{ij} \tag{14}$$

where δ_{ij} is the Kronecker δ, σ^2 being the variance of the error in each frequency splitting ω_{1i}. In the definition of μ, Eq. (8), the average variance is scaled out so for this particular **E** the value of σ is immaterial for the construction of the kernels. Because the errors in the data are independent and Gaussian distributed with equal variances σ^2, the corresponding error in the rotation rate is also Gaussian distributed, with variance

$$\sigma^2(x_0, \Omega) = \Lambda^2(x_0)\sigma^2(\omega) , \tag{15}$$

the *error magnification* Λ being defined by

$$\Lambda(x_0) \equiv \left[\sum_{i \in \mathscr{M}} c_i(x_0)^2 \right]^{\frac{1}{2}} . \tag{16}$$

First we consider the effect of choosing different values for the parameters μ and Δ when constructing kernels at fixed target radius $x_0 = 0.5$ (Fig. 1). It is well known that reducing μ tends to make the averaging kernels narrower but at the expense of increasing the error magnification. This behaviour is exhibited at constant Δ in each column of the figure, reading from top to bottom. The effect of reducing Δ at fixed μ is seen in each row, reading from right to left. At relatively large values of Δ ($\gtrsim 0.05$) the averaging kernels approximate very well the target form, so that the effect of reducing the target width is simply to reduce the width of the averaging kernel. As Δ decreases, however, the kernels show increasingly pronounced wings and oscillatory behaviour away from the radius x_0. This is reminiscent of the Gibbs phenomenon in Fourier theory and arises for essentially the same reason, namely that we are trying to use a set of differentiable functions to approximate a function that varies locally on a smaller length-scale than any of the members of the set. As demonstrated by PT, for small values of Δ the kernels constructed with a Gaussian target function are virtually indistinguishable from kernels constructed with target function $\mathscr{T} = \delta(x - x_0)$ (see Fig. 2).

Pronounced wings and other structure away from the target radius mean that the averaging kernels will produce a nonlocal average of the rotation rate. This is undesirable, and yet one does not generally wish to inspect each kernel by eye to see when this is occurring. Instead one needs some measure to indicate when a kernel is no longer a good approximation to the target form $\mathscr{T}$. Accordingly, we define

$$\chi \equiv \int_0^1 \left[\mathscr{K}(x_0, x) - \mathscr{T}(x_0, x) \right]^2 \mathrm{d}x . \tag{17}$$

Figure 1 shows how, in general, well-localized kernels correspond to small values of χ. By making sure that Λ is small one effectively limits the influence of random data-errors. By making sure that χ is small one limits the influence of 'systematic' errors due to nonlocal averaging by the kernel. Ideally, then, the parameters μ_0 and Δ are to be so chosen to ensure that both Λ and χ are small, insofar as this can be achieved.

We shall assume that the aim in constructing an averaging kernel is to obtain a well-localized function that will provide

© European Southern Observatory • Provided by the NASA Astrophysics Data System

1994A&A...281..231P

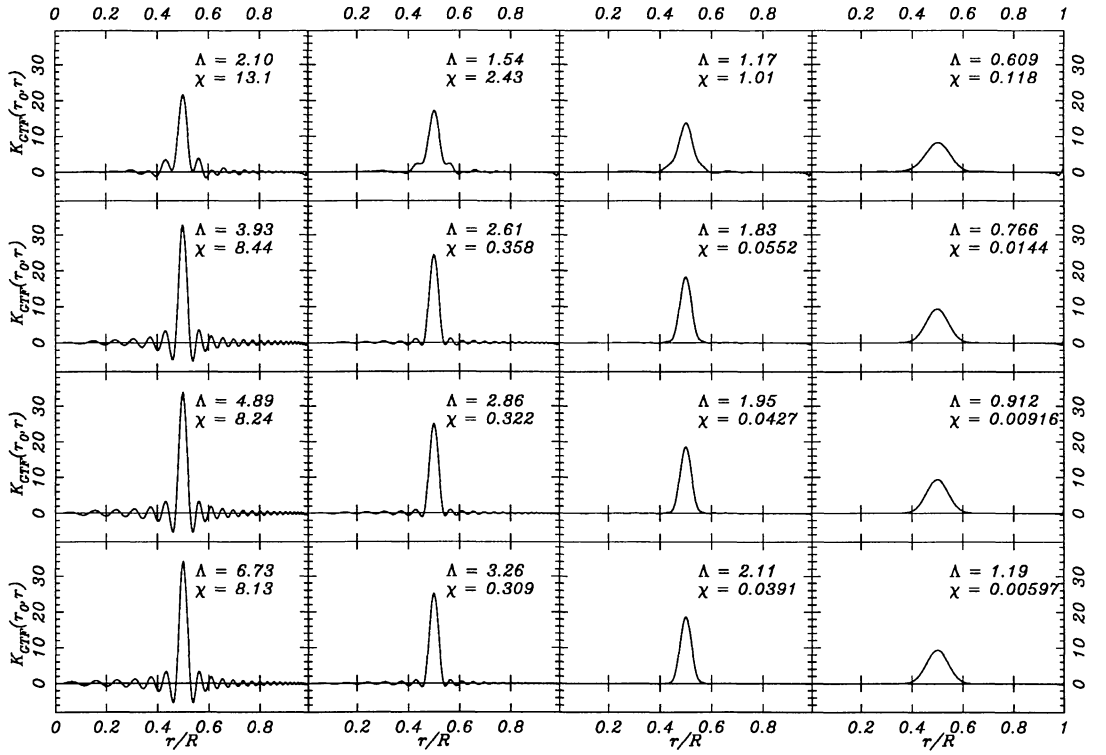

Fig. 1. The behaviour of the GTF kernels at target radius $x_0 = 0.5$, for various values of the width Δ and error weighting μ_0. The value of Δ increases from the left column to the right with values 10^{-2}, 2×10^{-2}, 3×10^{-2}, and 6×10^{-2} (in units of the solar radius); μ_0 increases from the bottom row up with values 3×10^{-3}, 10^{-2}, 6×10^{-2}, and 1.3. For each kernel the appropriate error magnification and χ (see text) are given

as fine a resolution as possible, consistent with the magnification of random errors being kept below some chosen value. The acceptable error magnification will vary from one problem to another, because it depends on the size of the data errors and on the size of the variations in the solution that are being investigated. For the purposes of our discussion we shall assume that it is of order unity. To obtain fine resolution with the GTF kernels it is obviously desirable for Δ to be as small as possible. But, as can clearly be seen in Fig. 1, when Δ is reduced too far the kernels develop nonlocal structure. This transition can be seen to take place in Fig. 1 between $\Delta = 0.03$ (column 3) and $\Delta = 0.02$ (column 2); moreover, the transition depends only weakly on μ_0. Clearly, then, in this instance the target width Δ (which we may think of as the target resolution) cannot be reduced much below 0.03, given our aim of producing well-localized kernels. It can also be seen from Fig. 1 that χ is sensitive to the departure of the kernel from the target form. The figure enables us to calibrate χ (cf. the appendix), since it is evident by eye that the transition between well-localized kernels (column 3) to kernels with nonlocal structure (column 2) corresponds to χ increasing from a few times 10^{-2} to a few times 10^{-1}, at least in rows 2-4. (The χ measure in column 3, row 1 is already large, because the kernel has substantial shoulders that the Gaussian target does not have. Thus although in this case the kernel is reasonably well localized, the χ value indicates that the effective resolution of this kernel is rather different from what was intended.) Thus,

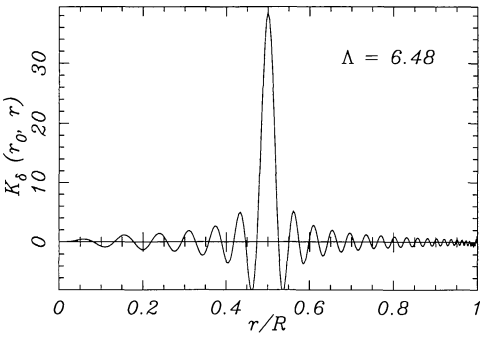

Fig. 2. The kernel obtained using $\mathscr{T} = \delta(x - x_0)$ with $\mu_0 = 10^{-2}$

while it is desirable to choose Δ to be small, it is also desirable that $\chi \lesssim 0.1$. The increase in χ to values of order 0.1 as Δ is reduced from 0.03 to 0.02 can also be seen in Fig. 3. If we suppose that the largest acceptable error magnification at this radius is Λ_m, and assuming that the largest tolerable χ is 0.1, then the values of the parameters Δ and μ_0 corresponding to these criteria can be read of from the intersection of the $\chi = 0.1$ and $\Lambda = \Lambda_m$ contours in Fig. 3. There is of course an entire domain in Fig. 3 where $\chi < 0.1$ and $\Lambda < \Lambda_m$; but any point in that domain has a larger value of Δ and therefore has poorer resolution.

A classical tool for determining the choice of parameter values in an inversion is the "trade-off diagram" (e.g. CDST)

© European Southern Observatory • Provided by the NASA Astrophysics Data System

1994A&A...281..231P

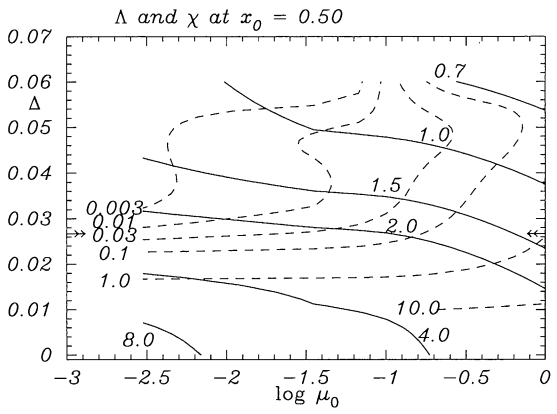

Fig. 3. The behaviour of Λ and χ at radius $r_0 = 0.5\,R$ as the parameters $\log \mu_0$ and Δ are varied. Contours of constant Λ are shown as solid curves; contours of constant χ are dashed

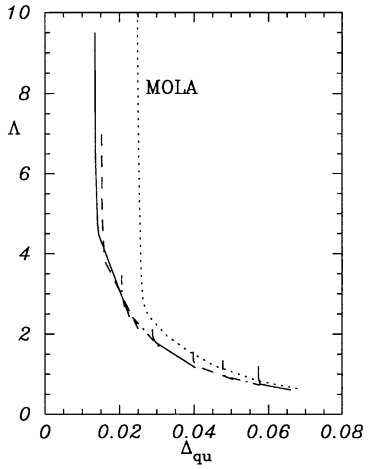

Fig. 4. Trade-off diagram of error magnification Λ against resolution width (defined to be the distance Δ_{qu} between the quartile points of the averaging kernel), at radius $r_0 = 0.5R$. Curves join points with the same target width Δ; values used were $\Delta = 0.00, 0.01, 0.02, 0.03, 0.0416, 0.05, 0.06$. For each curve, μ_0 increases from right to left, from 0.003 to 1.26. For comparison the MOLA trade-off curve (cf. CDST) is also shown (dotted curve)

wherein error magnification is plotted against resolution at fixed radius. In a one-parameter inversion such as the MOLA (eq. (7)) the trade-off diagram consists of a single curve and an inversion's position on that curve is determined by the parameter value. Figure 4 shows a trade-off diagram for the two-parameter SOLA inversion with Gaussian target function. Points corresponding to the same target width Δ are joined up, so each curve corresponds to a single value of Δ and along each such curve μ_0 varies from small values (top left) to large values (bottom right). The resolution measure employed, following CDST, is the distance Δ_{qu} between the lower quartile point (to the left of which lies one quarter of the area under the kernel) and the upper quartile point (to the right of which lies one quarter of the area). The corresponding interquartile distance for the *target* function is $\Delta^{\mathcal{T}}{}_{qu} \approx 0.68\sqrt{2}\Delta = 0.96\Delta$; thus when the target is well-approximated by the kernel, the values of Δ and Δ_{qu} are very similar. As μ_0 is reduced at constant Δ the error magnification increases but the the resolution width cannot get smaller than the corresponding target width; thus each curve at constant Δ asymptotes to $\Delta_{qu} \approx \Delta$ (provided Δ is not too small) as μ_0 is reduced, and thus there is eventually an increase in error magnification with no significant reduction in width Δ_{qu}. (Of course, χ tends to get smaller, but if χ is already very small this is of no great benefit.) This can be seen from Fig. 3, by considering the behaviour of Λ and χ on a horizontal line at constant Δ.

A remarkable aspect of Fig. 4 is that apart from the peeling off at small μ_0 for fixed Δ, discussed above, the data all lie on roughly the same trade-off curve. For larger values of μ_0 many different pairs of values of μ_0 and Δ can correspond to essentially the same point on that curve. They will not in general produce the same averaging kernel, however. As can be seen from Fig. 3, if μ_0 is increased at constant Δ, the value of χ will increase. This is precisely the behaviour seen reading up the third and (less obviously) the fourth columns of Fig. 1. As μ_0 is increased, previously well-localized kernels tend to develop extended shoulders and thus become less well localized. (These broader kernels are reminiscent of the MOLA kernels in PT.) Thus different parameter values may give kernels with similar error magnification and resolution Δ_{qu}, but the kernels will be different in detail, just as different inversion methods produce kernels that differ in detail although the methods look very similar in terms of their trade-off curves. Insofar as the target form has been well chosen and one wishes to match it, it is best to arrange that χ is small, which will be the case if the constant-Δ curve is just turning off the overall trade-off curve. This occurs at χ between 0.05 and 0.1 which is consistent with the visual impression of Fig. 1.

The SOLA trade-off curve in Fig. 4 is essentially the same curve as for other inversion methods. The MOLA trade-off curve is shown there for comparison, and trade-off curves for other methods can be seen in CDST. At small Δ_{qu}, in common with other methods, the error magnification becomes large for little gain in resolution. Here the interquartile measure of resolution is rather flattering to poorly localized kernels and the χ measure reveals what is really happening. The point is well made by the leftmost columns of Fig. 1 : as Δ is reduced (in this case below about 0.025) the central peak of the kernels gets narrower but the kernels develop oscillatory nonlocal structure. This nonlocal structure is not particularly detrimental to Δ_{qu} but means that the inversion will be sensitive to regions far from the target location. Thus it is valuable to inspect the χ measure. A method like regularized least squares (e.g. CDST) produces similar oscillatory kernels in this regime. The MOLA method is in that sense more honest, because it tends not to produce kernels with such small Δ_{qu} measures.

The discussion thus far has related to a single target radius r_0, but of course the inversion must be performed at various radii. How then should the parameters be chosen? There are two issues: first, that Δ must be chosen appropriately for each target radius; and secondly that the full benefit of the SOLA cost-saving is only achieved if the same μ_0 is used for many radii.

© European Southern Observatory • Provided by the NASA Astrophysics Data System

1994A&A...281..231P

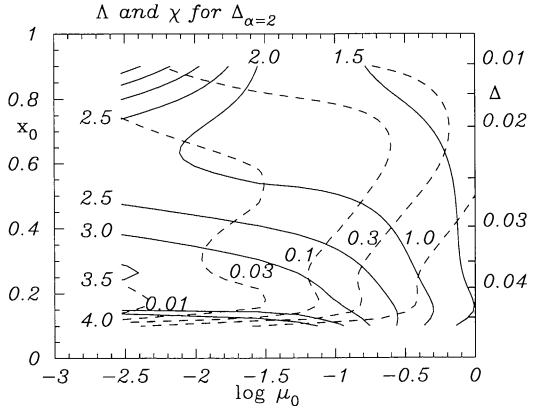

Fig. 5. Contours of constant χ (dashed) and error magnification Λ (solid), as functions of x_0 and μ_0. Δ has been chosen at each radius according to eq. (17) , with $\alpha = 2$: its value is shown on the right axis

The values of Δ could be chosen completely independently at each radius where it is desired to construct an averaging kernel, but such an approach is laborious. Fortunately, it is possible to use physical arguments to choose how Δ should scale with r_0, over a wide range of radii. It is reasonable that the smallest lengthscale over which details of the rotation can be resolved at any particular radius is related to the smallest spatial scale at that radius in the set of rotational kernels. This in turn is essentially inversely proportional to the highest local vertical wavenumber of the eigenfunctions for the given mode set. After due acount is taken of the various constants of proportionality, this argument predicts that the ultimate limit to the resolution width $\Delta_{\rm qu}$ at radius r_0 is approximately $c_S(r_0)/(8R\nu_{\rm max})$, where $c_S(r_0)$ is the adiabatic sound speed at r_0 and $\nu_{\rm max}$ is the highest frequency mode in the set, assuming that r_0 lies within that mode's acoustic cavity (Thompson 1993). It can be seen for a regularized least-squares inversion with the same modeset as used here that not only is the agreement very good between the limiting resolution and this prediction, but also that the resolution still scales in the same way with radius when the amount of regularization in the inversion is increased (Thompson 1993; cf.Christensen-Dalsgaard et al.1993). Thus Δ should be chosen to be

$$\Delta \propto \frac{1}{2}\min(\lambda_i) \qquad i \in \mathscr{M} \tag{18}$$

where λ_i is the wavelength of the ith kernel, at the target radius, and from the previous discussion this can be written as

$$\Delta(x_0) = \frac{\alpha}{8}\frac{c_S(x_0)}{R\nu_{\rm max}} \tag{19}$$

where α is a constant of proportionality, independent of radius. Hence the strategy is to optimize Δ at one radius and then use the relation (19) to scale Δ for all other x_0.

The second issue is whether a single choice of μ_0 will suffice for all, or at least a wide range of, radii, since changing μ_0 necessitates a new matrix inversion. The two issues together are addressed by Fig. 5, which shows contours of constant χ and of constant Λ as functions of position and of μ_0, assuming that Δ

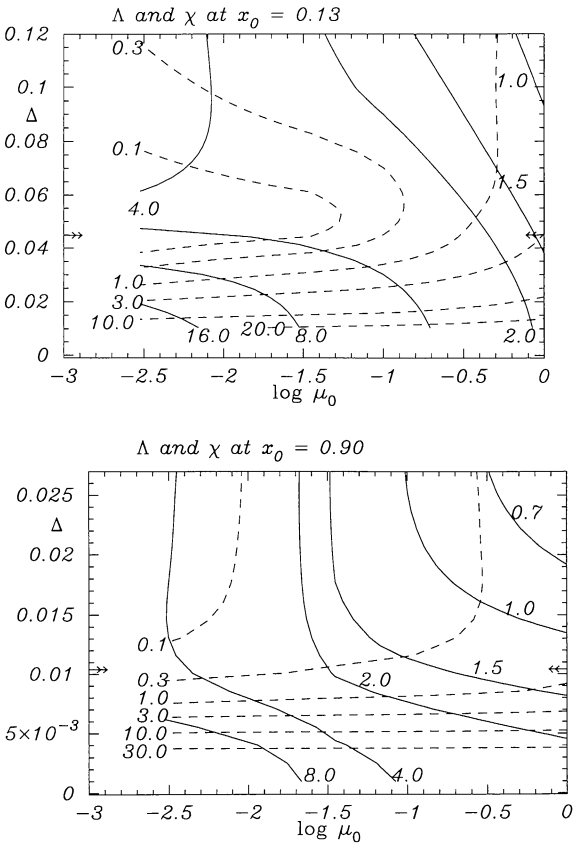

Fig. 6 a and b. As Fig. 3 but for **a** $r_0 = 0.13R$ and **b** $r_0 = 0.9R$

has been chosen according to eq. (19) , with $\alpha = 2$. Note that the contours of constant χ are fairly vertical except at small and large radii, so that over a large range of radii one can construct well-localized kernels using the chosen scaling of Δ without changing μ_0. It is only near the surface and near the centre, where few if any of the modes used have lower turning points, that the behaviour is problematic. This is further illustrated in Fig. 6, which shows χ and Λ contours as functions of Δ and μ_0, analogously to Fig. 3, at radii $r_0 = 0.13R$ and $r_0 = 0.9R$. The automatically scaled Δ, with $\alpha = 2$, is indicated here and in Fig. 3 with a small arrow. The Δ scale has been chosen in each panel so that the arrow is in roughly the same place. At all radii the χ measure increases rapidly if Δ is reduced too much. The anomalous behaviour for $r_0 = 0.13R$, wherein at fixed μ_0 the measure χ can *increase* with Δ is presumably an end-effect, caused by the target having too large a nonzero value at $x = 0$, where essentially all the mode kernels vanish. It is evident from Fig. 6 that near-Gaussian kernels can be constructed even at these small and large radii, by choosing a smaller value of μ_0, provided a large error magnification can be tolerated.

© European Southern Observatory • Provided by the NASA Astrophysics Data System

1994A&A...281..231P

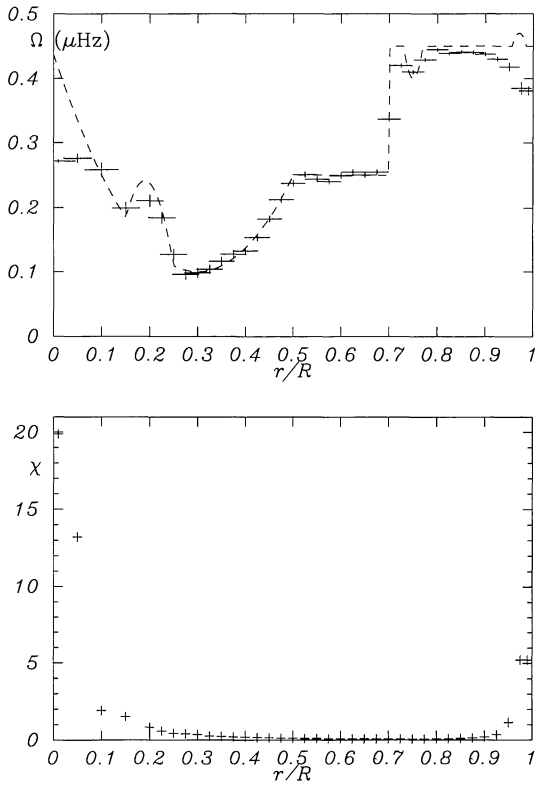

Fig. 7a and b. a The first artificial profile. The input rotation profile is shown as the dashed line. The superposed crosses are the recovered values using a simple Gaussian as target function in the SOLA method, plotted against target radius. The horizontal error bar is 2Δ wide. The vertical error bars are $\pm\Lambda(x_0)\sigma$ in extent. **b** The values of χ for the kernels used in reconstructing Ω

4. The blind tests

4.1. Results using the Gaussian target function

Using the same Set 1 of CDST some blind tests of the SOLA method were performed. The rotation rate was taken to be a function of r/R only. An artificial rotation rate was prepared by one of the authors and frequency splittings were computed. The splittings were then contaminated with random errors sampled from a Gaussian distribution with a standard deviation $\sigma = 2$ nHz, which is comparable to the uncertainty in the best determined a-coefficients (see Libbrecht 1989). The splittings were then passed to the other author for inversion.

Figure 7 shows the results for the inversion of the first test profile. The high values of the χ measure indicate that the determinations at the innermost and outermost few radial points are of poor quality due to nonlocal structure in the kernels. Inspection of the upper panel confirms that these determinations do indeed differ from the input profile. The failure at very small radii is due to the fact that the lowest degree modes in the data set do not penetrate deeply enough to probe this region. The failure near the surface is due to the lack of high degree modes. This causes spurious structure in the kernels above $r = 0.95R$ which influences the determinations as deep as $r = 0.85R$. The small dip just above the discontinuity is barely resolved. The bump near the surface is not reproduced at all, because of the problems with the quality of the kernels this near to the surface.

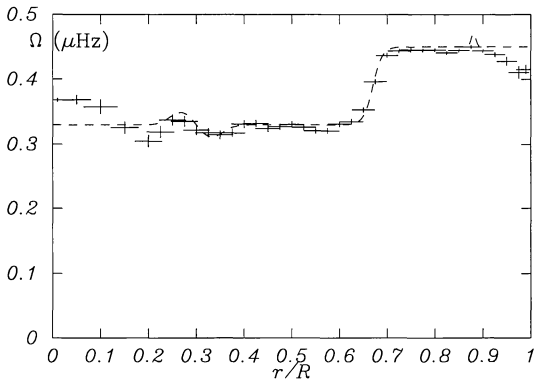

Fig. 8. The second artificial profile. The use of the symbols is the same as in Fig. 7

Figure 8 shows a second artificial profile which is smooth and with some features that are smaller in amplitude or spatial extent. Again the SOLA reconstruction of the profile is generally quite good, but the small bump just below $r = 0.9R$ cannot be recovered. Below $r = 0.2R$ the resolution with this mode set becomes rather poor and nonlocal effects in the kernels start to dominate the inversion. In both Fig. 7 and Fig. 8 any agreement of the inversion with the input profile below $r \approx 0.2R$ is probably fortuitous.

4.2. Determination of the first derivative

It might be of interest to determine whether the rotation profile has highly localized strong gradients or 'discontinuities'. Although this can be inferred from the determination of Ω itself, because of finite resolution and errors this will not always yield the desired quality for the result. Target form (13) may be generalized to

$$\mathcal{T} = \frac{1}{f\Delta}\mathcal{P}_n\left((x-x_0)/\Delta\right)\exp\left[-\left(\frac{x-x_0}{\Delta}\right)^2\right], \qquad (20)$$

where $\mathcal{P}_n$ is an arbitrary polynomial of degree n :

$$\mathcal{P}_n(u) = \sum_{k=0}^{n} a_k u^k \qquad (21)$$

By choosing appropriate coefficients a_k one may design a target which will produce kernels that are sensitive to strong gradients in the rotation. Taking

$$\begin{cases} a_1 = \dfrac{2}{\Delta(x_0)}, \\ a_k = 0, \qquad \text{all other } k \end{cases} \qquad (22)$$

produces an antisymmetric kernel. This kernel has a zeroth moment that is essentially zero and a first moment

$$M_1 = \int_0^1 (x-x_0)\mathcal{T}(x_0,x)\mathrm{d}x \qquad (23)$$

© European Southern Observatory • Provided by the NASA Astrophysics Data System

1994A&A...281..231P

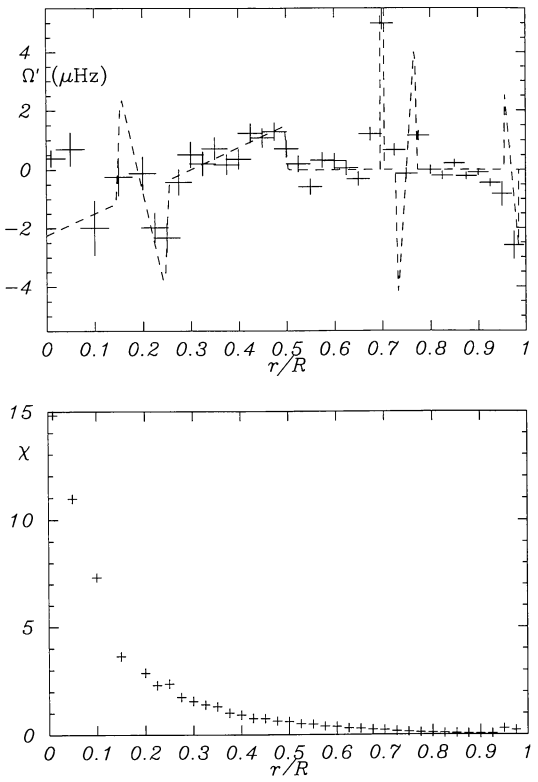

Fig. 9. The inversion for the gradients in the first artificial profile. The use of the symbols is the same as in Fig. 7

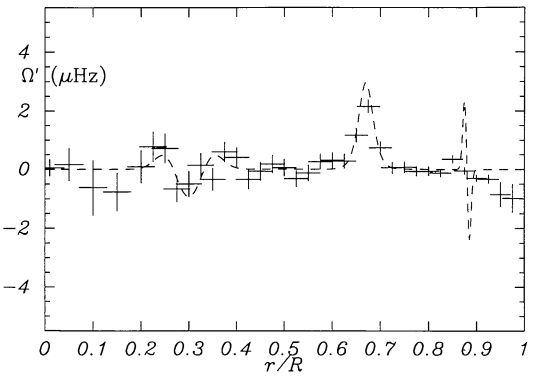

Fig. 10. The gradients in the second artificial profile. The use of thesymbols is the same as in Fig. 9

that is equal to unity (the factor f being chosen appropriately). The kernel should therefore not be sensitive to the rotation rate itself, but only to the odd terms in a Taylor expansion of Ω around x_0. If the rotation rate has a strong gradient around x_0 this will show up as a signal for this inversion kernel.

Figure 9 shows the results for this inversion. The zeroth moment of the kernel was made exactly zero by setting the right-hand side of constraint (4) to 0. Note that the χ-measure has been multiplied by Δ^2 for consistency with the measure for the simple Gaussian kernel. With this normalization the integral of the square of the target function is the same for both types of kernels. Figure 10 show the results for the second profile. Although the finite resolution of the kernels causes the smearing out of the signature of the discontinuity over three points, it still clearly stands out in the inversion. The smoother behaviour around $r \approx 0.7R$ of the second profile is reflected in the much smaller gradient. This method can therefore distinguish the two cases provided the jump in rotation rate is large enough or the errors are small enough.

This determination is in general better than simply taking the difference of the zero order inversion at adjacent points. The combination of coefficients that corresponds to the effective averaging kernel of that procedure (i.e. the difference of the two coefficients for each mode) is not likely to be the combination that is optimal according to the general recipe of minimizing (6) with $\mathscr{T}$ defined using Eq. (20), (21), and (22). It will generally yield a determination that is less accurate and also with a larger uncertainty due to data errors.

However it is important to remember that the determinations of Ω and its gradient use the same mode set and are therefore not completely independent. The small negative gradient near the surface appears to confirm the decrease in Ω near the surface which is in fact spurious. In all cases the χ measure should be monitored to guard for nonlocal effects in the kernel. If χ is large in one or both of the zero order and first order types of inversion then consistency in the results is meaningless.

4.3. Higher order terms

The generalization to choosing the coefficients of $\mathscr{P}_n$ for determining higher order terms is straightforward. The only further target function of this kind that we have considered has zeroth and first moments equal to 0 and a second moment [cf. eq. (28)] equal to -1. Such a kernel has a central peak with negative sidelobes on either side. It is sensitive to small scale 'bumps' and 'dips' that introduce large absolute values of the second derivative of Ω in a Taylor series. We have tried these kernels but the errors become quite large, so that the real variation in the second derivative of our artificial rotation rate was of the order of the one standard deviation error bars.

5. Discussion

We have previously demonstrated (PT) that the SOLA method produces kernels that are of as good a quality as those obtained with the more traditional OLA formulations and shown that the SOLA method is computationally more efficient. Here we have sought to show how the parameters for SOLA with a Gaussian target function might best be chosen and have related the SOLA results to the classical trade-off diagram of error magnification against resolution. For inverting for the solar rotation rate, SOLA is certainly as good as MOLA or other methods, as our tests with artificial data show. With different choices of target forms, it is possible to estimate directly the first or higher derivatives of the rotation rate, which illustrates the versatility

© European Southern Observatory • Provided by the NASA Astrophysics Data System

1994A&A...281..231P

of this approach. In those cases where the test inversion results diverge from the input rotation profile, the independent accuracy measure χ does indeed show a rapid increase and therefore indicates that the determinations are unreliable. Although the inversions in this paper all use artificial data, we have presented elsewhere the results of inverting real data (Pijpers & Thompson 1993).

There are several directions in which one might further develop the SOLA method as we have used it here. One, which is not exclusive to SOLA, is that one might work with some independent variable other than the radius. A second is that one could design target forms that deliberately have zero higher moments, since it has been argued (CDST) that kernels with vanishing higher moments are advantageous for reproducing the underlying function. As final thoughts, therefore, we discuss each of these ideas in turn.

It has been suggested that the acoustical depth, defined as

$$\tau(r) \equiv \int_r^R \frac{\mathrm{d}r'}{c_S(r')} \tag{24}$$

is a more natural independent variable for the inversions than r/R. Thus one might envisage working throughout in terms of integrals with respect to τ rather than x, and choosing a Gaussian function of the form (13) with x, x_0, Δ replaced with τ, τ_0, Δ_τ. It is simple to prove that this formulation is equivalent to continuing to work with x, but with a different target function. Specifically, the necessary target function is the Gaussian target $\mathscr{T}$ multiplied by an extra weighting function W that has a value close to unity over a large part of the solar interior :

$$W = \frac{1}{c_S(x)\langle c_S^{-1}\rangle} ; \tag{25}$$

and Δ is related to Δ_τ by

$$\Delta = \frac{\Delta_\tau}{R\langle c_S^{-1}\rangle} . \tag{26}$$

In Eqs. (25) and (26) the inverse mean sound speed $\langle c_S^{-1}\rangle$ is defined as

$$\langle c_S^{-1}\rangle \equiv \frac{1}{x - x_0}\int_{x_0}^{x} \frac{\mathrm{d}x'}{c_S(x')} . \tag{27}$$

The weighting function only changes rapidly as a function of r above $0.9R$. It will therefore have an influence on inversions only if the peak of the Gaussian falls in that range of r. In all other cases the rapid decrease of the Gaussian quenches all effects of this weighting. For inversions in which near-surface structure is of particular interest the acoustical depth may indeed be a more appropriate independent variable. As the results of the artificial profile show, the current inversions of Ω above $r \approx 0.9R$ may suffer somewhat from spurious surface structure in the kernels. The use of $\tau(r)$ as independent variable may alleviate these problems to some extent.

We turn finally to the question of designing kernels with vanishing higher moments. It was argued in CDST that the reason that the regularized least squares inversion with second derivative smoothing reproduces a smooth underlying function so well is that the kernels for that method tend to have small second moments

$$\frac{1}{2!}\int (x - x_0)^2 \mathscr{K}\, \mathrm{d}x. \tag{28}$$

By a proper choice of the coefficients a_k one can construct targets of the form (20) with second or higher moments that are equal to zero. For example, the choice

$$\begin{cases} a_0 = \frac{3}{2}, \\ a_2 = -1, \\ a_k = 0, \text{ all other } k \end{cases} \tag{29}$$

yields targets $\mathscr{T}$ with a second moment equal to zero. The second moment of the constructed kernels will generally not be equal to 0 even if visually they resemble the target function very closely, because the higher moments are quite sensitive to the residual structure near the surface. Averaging kernels produced in this way resemble the regularized least squares kernels closely and have similar second moments to those reported by CDST (see their Table 1.).

Acknowledgements. This work was supported by the UK Science & Engineering Research Council under grant no. GR/H33596.

Appendix : the use of χ as estimator of systematic errors

Given a target averaging kernel $\mathscr{T}$ the appropriate weighted average of the rotation rate is :

$$\langle \Omega(x_0)\rangle_{\mathscr{T}} = \int_0^1 \mathscr{T}(x_0, x)\Omega(x)\mathrm{d}x \tag{A1}$$

In practice it is impossible to reproduce perfectly the target averaging kernel $\mathscr{T}$ from the inversion. This means that there is a 'systematic' error connected to the constructed averaging kernel which is the difference between the weighted average of $\Omega(x_0)$ from (A1) and the estimate obtained from the splittings coefficients. Defining

$$f \equiv \mathscr{K}(x_0, x) - \mathscr{T}(x_0, x) , \tag{A2}$$

it is evident that

$$\langle \Omega(x_0)\rangle = \langle \Omega(x_0)\rangle_{\mathscr{T}} + \int_0^1 f\Omega\mathrm{d}x \tag{A3}$$

so that $\int f\Omega dx$ is the error due to the averaging kernel not reproducing the target form perfectly. It is possible to give an upper limit to the error due to this effect using the measure χ introduced in the paper. Because of the normalization of $\mathscr{K}$ and $\mathscr{T}$, f satisfies

$$\int_0^1 f\mathrm{d}x = 0 . \tag{A4}$$

© European Southern Observatory • Provided by the NASA Astrophysics Data System

1994A&A...281..231P

Therefore

$$\int_0^1 f\Omega \mathrm{d}x \;=\; \int_0^1 f\left(\Omega - \overline{\Omega}\right)\mathrm{d}x\;, \tag{A5}$$

where $\overline{\Omega} = 1/2\,(\Omega_{min} + \Omega_{max})$. It is useful to define $\Delta\Omega = \Omega - \overline{\Omega}$ and $\Delta\Omega_{max} = 1/2\,(\Omega_{max} - \Omega_{min})$. It follows that

$$\left|\int_0^1 f\Delta\Omega \mathrm{d}x\right| \leq \int_0^1 |f\Delta\Omega|\mathrm{d}x \leq \Delta\Omega_{max}\int_0^1 |f|\mathrm{d}x\;. \tag{A6}$$

Unfortunately the calculation of $\int |f|\mathrm{d}x$ is complicated. It is much easier to make use of the parameter

$$\chi \;\equiv\; \int_0^1 f^2\mathrm{d}x\;. \tag{A7}$$

Since, as we prove below,

$$\int_0^1 f^2\mathrm{d}x \geq \left(\int_0^1 |f|\mathrm{d}x\right)^2\;, \tag{A8}$$

it follows that

$$\int_0^1 f\Delta\Omega\mathrm{d}x \;\leq\; \Delta\Omega_{max}\left(\int_0^1 f^2\mathrm{d}x\right)^{1/2} = \Delta\Omega_{max}\chi^{1/2}\;. \tag{A9}$$

Proof :

$$\int_0^1 f(x)^2\mathrm{d}x - \left(\int_0^1 |f(x)|\mathrm{d}x\right)^2 =$$

$$= \int_0^1 f(x)^2\mathrm{d}x - \left(\int_0^1 |f(x)|\mathrm{d}x\right)\left(\int_0^1 |f(y)|\mathrm{d}y\right) =$$

$$= \int_0^1 f(x)^2\mathrm{d}x - \left(\int_0^1\int_0^1 |f(x)f(y)|\mathrm{d}x\mathrm{d}y\right) =$$

$$= \int_0^1 |f(x)|\left[|f(x)| - \int_0^1 |f(y)|\mathrm{d}y\right]\mathrm{d}x\;. \tag{A10}$$

It is clear that

$$\int_0^1 \left(|f(x)| - C\right)^2\mathrm{d}x \geq 0 \tag{A11}$$

for any constant C so in particular for

$$C \;=\; \int_0^1 |f(y)|\mathrm{d}y\;. \tag{A12}$$

By rearranging terms in (A11) and making use of (A12) ,

$$\int_0^1 |f(x)|\left(|f(x)| - C\right)\mathrm{d}x \geq \int_0^1 C\left(|f(x)| - C\right)\mathrm{d}x$$

$$= C\left(\int_0^1 |f(x)|\mathrm{d}x - C\right) \;=\; 0\;. \tag{13}$$

Combining this with (A10) gives

$$\int_0^1 f^2\mathrm{d}x - \left(\int_0^1 |f|\mathrm{d}x\right)^2 \geq 0\;. \tag{A14}$$

Q.E.D.

This means that an upper limit for the error due to imperfect target averaging kernel matching is :

$$\left|\int f(x_0,x)\Omega(x)\mathrm{d}x\right| \leq \chi^{1/2}\frac{(\Omega_{max} - \Omega_{min})}{2}\;. \tag{A15}$$

References

Backus, G. E., Gilbert, J. F., 1968, Geophys.J., 16, 169

Backus, G. E., Gilbert, J. F., 1970, Phil.Trans.R.Soc.Lond.A, 266, 123

Christensen-Dalsgaard, J., Thompson, M. J., 1993, A&A, in the press

Christensen-Dalsgaard, J.,Hansen, P. C., Thompson, M. J., 1993, MNRAS, submitted

Christensen-Dalsgaard, J.,Schou, J., Thompson, M. J., 1990, MNRAS 242, 353 (CDST)

Gough, D. O., 1985, Sol.Phys. 100, 65

Jeffrey, W., 1988, ApJ 327, 987

Libbrecht, K. 1989, ApJ 336, 1092

Pijpers, F. P., Thompson, M. J., 1992, A&A 262, L33 (PT)

Pijpers, F. P., Thompson, M. J., 1993, in: Proc. GONG 1992: Seismic investigation of the Sun and stars, ed. Brown, T. M., PASPC, in the press

Ritzwoller, M. H., Lavely, E. M., 1991, ApJ 369, 557

Thompson, M. J., 1993, in: Proc. GONG 1992: Seismic investigation of the Sun and stars, ed. Brown, T. M., PASPC, in the press

This article was processed by the author using Springer-Verlag TEX A&A macro package 1992.

© European Southern Observatory • Provided by the NASA Astrophysics Data System