Geophys. J. Int. (1996) **124,** 372–380

TREITEL SECTION

Nullspace shuttles

Michael M. Deal and Guust Nolet

Department of Geological and Geophysical Sciences, Princeton University, Princeton, NJ 08544, *USA*

Accepted 1994 December 5. Received 1994 November 8; in original form 1994 April 14

SUMMARY

In seismic tomography the problem is generally underdetermined. The solution to the tomographic problem depends on the specific optimization condition used and is inherently distorted due to noise in the data and approximations in the theory. Smoothing is often applied to reduce inversion artefacts with short correlation lengths. However, *a posteriori* smoothing generally affects the data fit. For more sophisticated, non-linear filters this effect can be severe. We present a technique to conserve data fit for filters of arbitrary complexity. The difference between the 'optimal' solution and a filtered version is projected onto the nullspace of the model space in order to preserve the data fit. Thus, we only allow changes to the image that do not conflict with the data. We demonstrate the benefits of such conservative filters using several different non-linear filters to reduce noise, smooth the image, and highlight edges.

The method is exact in small-scale experiments, where we can use the method of singular value decomposition: eigenvectors with large eigenvalues are used to project the difference between the original model and the filtered version onto the nullspace. With large-scale tomographic problems, calculation of all of the large eigenvectors is unrealistic. We show how to use the iterative method of conjugate gradients to apply conservative filters to large-scale tomographic problems with minimum computational effort.

Key words: inversion, seismic tomography.

INTRODUCTION

While the images obtained with seismic tomography give us a new and extraordinary view of the Earth, we must keep in mind the limitations and weaknesses of the solutions. A solution to an inverse problem is not a unique solution, but instead is one that attempts to explain or fit the data subject to a criterion that selects an 'optimal' model. For example, in a standard least-squares problem there are many solutions that minimize the sum of the squared errors, and the solution with the smallest Euclidean norm is often chosen as the 'optimal' model. Solving for the minimum-norm solution is computationally easy to implement using back projections (Van der Sluis & Van der Vorst 1987), but the complexity of a given problem may be better suited by using a different definition of optimality.

Many applications in global tomography suffer from an uneven distribution of ray paths. In this case, minimum-norm solutions tend to distribute the velocity anomalies over model cells that are visited by many rays while leaving the other cells untouched. This introduces unwarranted lateral heterogeneity. Su, Woodhouse & Dziewonski (1994) avoid such effects by projecting the model onto a basis of low-order spherical harmonics. However, this forces an extreme smoothness on the model which does not allow us to see small-scale structures such as subduction zones, even when the degree is raised to angular order 12 (Su *et al.* 1994). In shallow seismic investigations, additional information is often available from well logging or surface geology indicating sharp transitions or a limited choice of seismic velocities. None of such *a priori* information is addressed by simple minimum-norm or minimum-gradient solutions.

The criteria used to solve the tomographic problem can be adjusted and modified until all *a priori* information is incorporated into the inversion, but this will undoubtedly lead to an extremely complicated penalty function. Solving a difficult penalty function may not be computationally feasible or justifiable. Stork & Clayton (1991) solve the tomographic problem by including filters in a modified Dines & Lytle (1979) iteration formula. To prevent instability in the inversion, the filters used are limited to those that have a bandpass or averaging structure. Another approach is to solve the problem using a straightforward method such as the minimum-norm, least-squares algorithm and then to modify the image *a posteriori*.

Smoothing or filtering of the solution can be performed after the inversion to enhance the image or impose *a priori* information on the solution. *A posteriori* filtering has the added advantage that knowledge from the optimal solution can be included in the design of the filters. However, any changes to the solution will generally affect the fit to the data. The result is an 'improved' image, in the sense that more *a priori* information has been incorporated, but with a reduced data fit.

We present a method that allows us to filter an image *a posteriori* while retaining the fit to the data. The filter is restricted to operate only on components of the solution that do not affect the data fit. The filter can be restricted by projecting the changes implemented by the filter onto the nullspace of the model space. The minimum-norm solution can be combined with components from the nullspace to provide an infinite family of solutions all with the same data fit. We refer to a nullspace shuttle as the operator that allows us to move from one solution to another without corrupting the data fit. The nullspace shuttle is guided by the *a priori* assumptions or constraints we impose *a posteriori* on our 'optimal' model.

THEORY

The linearized inverse problem is commonly expressed in matrix notation as

$$\mathbf{Ax} = \mathbf{d}, \tag{1}$$

where $\mathbf{A}_{ij}$ is the distance the ith ray travels in cell j, x_j is the magnitude of the slowness deviations from the reference model in cell j, and d_i is the traveltime delay for ray i. We assume N data and M unknown model parameters. The familiar least-squares solution is derived from the Gauss normal equation

$$\mathbf{A}^{\mathrm{T}}\mathbf{Ax} = \mathbf{A}^{\mathrm{T}}\mathbf{d}. \tag{2}$$

When $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ is singular or nearly singular, an inverse can be constructed using singular value decomposition (SVD), and the slowness model $\mathbf{x}$ can be represented as a linear combination of the first k eigenvectors of $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ belonging to the k largest eigenvalues (Wiggins 1972; Jackson 1972):

$$\mathbf{x} = \sum_{j=1}^{k} a_j \mathbf{v}^{(j)} = \mathbf{V}_k \mathbf{a}, \tag{3}$$

where $\mathbf{V}_k$ is a matrix with the eigenvectors of $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ as columns. Assuming that R is the rank of matrix $\mathbf{A}$, the last $R-k$ eigenvalues are smaller than a given tolerance value and are ignored. By increasing the tolerance value we decrease the number of eigenvectors and effectively damp the solution. We define the 'generalized' nullspace as the space of all vectors that satisfy

$$\frac{|\mathbf{Ax}|}{|\mathbf{x}|} < \varepsilon'. \tag{4}$$

This leads to a relationship between the eigenvalue cut-off and a tolerance value ε'. If $\mathbf{x}$ is an eigenvector with eigenvalue λ, then

$$|\mathbf{Ax}| = |\lambda|\,|\mathbf{x}|, \tag{5}$$

and thus the generalized nullspace is spanned by all eigenvectors with

$$|\lambda| < \varepsilon'. \tag{6}$$

Increasing ε' allows us to expand the nullspace at the expense of the data fit. The effect of the tolerance ε' on the geometry of the generalized nullspace may depend strongly on the type of experiment. In a teleseismic inversion such as that by Nolet (1985), the eigenvalues tend to decrease exponentially (Van der Sluis & Van der Vorst 1987). In a more controlled set-up, the eigenvalues may decrease much more slowly and the dimension of the nullspace is a much weaker function of the data misfit. As an example, the synthetic experiment presented later is based on a borehole-type problem. The eigenvalues associated with the matrix $\mathbf{A}$ are shown in Fig. 1. The first 100 eigenvalues decrease quickly, but the eigenvalues decrease very gradually in the last three-quarters of the eigenvalue spectrum. A plot of the root-mean-squared (RMS) misfit associated with our synthetic experiment, Fig. 2, is similar in shape to the eigenvalue spectrum. Use of a tolerance value of $\varepsilon' = \lambda_1 \varepsilon$, where λ_1 is the largest eigenvalue and $\varepsilon = 1 \times 10^{-6}$, as a threshold in deciding which eigenvalues to ignore produces a solution which is a linear combination of the first 785 out of 900 eigenvectors. The solution has a variance reduction of 99.0 per cent. By incorporating only the first 250 eigenvectors into the solution, we obtain a variance reduction of 96.8 per cent. The improvement in the data fit with each successive eigenvector after the first 250 eigenvectors is small. Since the slope in the misfit plot is very close to zero, decreasing the value of k in eq. (3) results in only a slight decrease in the data fit.

Once a solution $\mathbf{x}$ has been obtained we can perform any type of smoothing or filtering operation that brings the image closer to what we expect based on *a priori* information or other geological evidence. The filtering stage can just as easily include the addition of energy to the solution as the removal or smoothing of energy. A variety of non-linear filters can be designed that highlight certain aspects of an image based on *a priori* information. Examples of this may include imposing a temperature profile on a subducting plate, emphasizing sharp boundaries associated with structures such as faults, dykes, salt domes or magma bodies, or constraining the velocity in certain regions based on known velocities from well logs. Since changes made to the solution will have a detrimental effect on the fit to the data, we propose an extra step that allows us to keep the data fit constant. We refer to these filters as conservative filters because the data fit is preserved.

If $\mathbf{F}$ is a linear or non-linear operator of arbitrary complexity, then

$$\mathbf{x}_{\mathrm{f}} = \mathbf{F}(\mathbf{x}), \tag{7}$$

where $\mathbf{x}_{\mathrm{f}}$ is a filtered solution that may violate the data. Allowing $\Delta\mathbf{x} = \mathbf{x}_{\mathrm{f}} - \mathbf{x}$ to be defined as the action of the filter, we can project the changes made to $\mathbf{x}$ onto the nullspace of $\mathbf{A}^{\mathrm{T}}\mathbf{A}$. Projecting $\Delta\mathbf{x}$ onto the nullspace removes all of the components in the range of $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ that violate the fit to the data. The conservative filtered solution can be constructed using eigenvectors of the model space and $\Delta\mathbf{x}$, the changes suggested by the filter. The information contained in $\mathbf{F}$ gives the conservative filtered solution, $\mathbf{x}_{\mathrm{c}}$:

$$\mathbf{x}_{\mathrm{c}} = \mathbf{x} + (\mathbf{I} - \mathbf{V}_k \mathbf{V}_k^{\mathrm{T}})\Delta\mathbf{x}, \tag{8}$$

where $\mathbf{I}$ is the identity matrix of order k. The root-mean-square (RMS) residual will be nearly identical for $\mathbf{x}_{\mathrm{c}}$ and the original solution $\mathbf{x}$. By reducing the value of k in eq. (5) we increase the size of the generalized nullspace. Since the nullspace shuttle only chooses solutions with components in the nullspace, a larger effective nullspace allows the filter $\mathbf{F}$ to

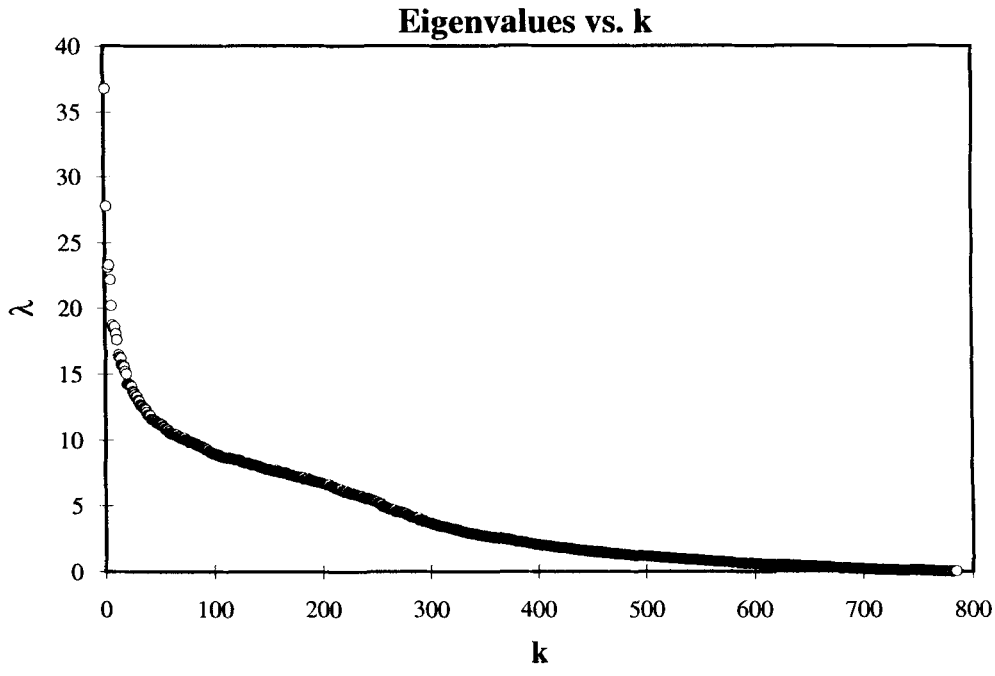


Figure 1. The decrease in eigenvalues relative to the number of eigenvectors determined using SVD. In our synthetic experiment we have 900 model parameters. Using a threshold value of $\varepsilon' = \lambda_1 \varepsilon$, where λ_1 is the largest eigenvalue and $\varepsilon = 10^{-6}$, SVD produces 785 non-zero eigenvalues.

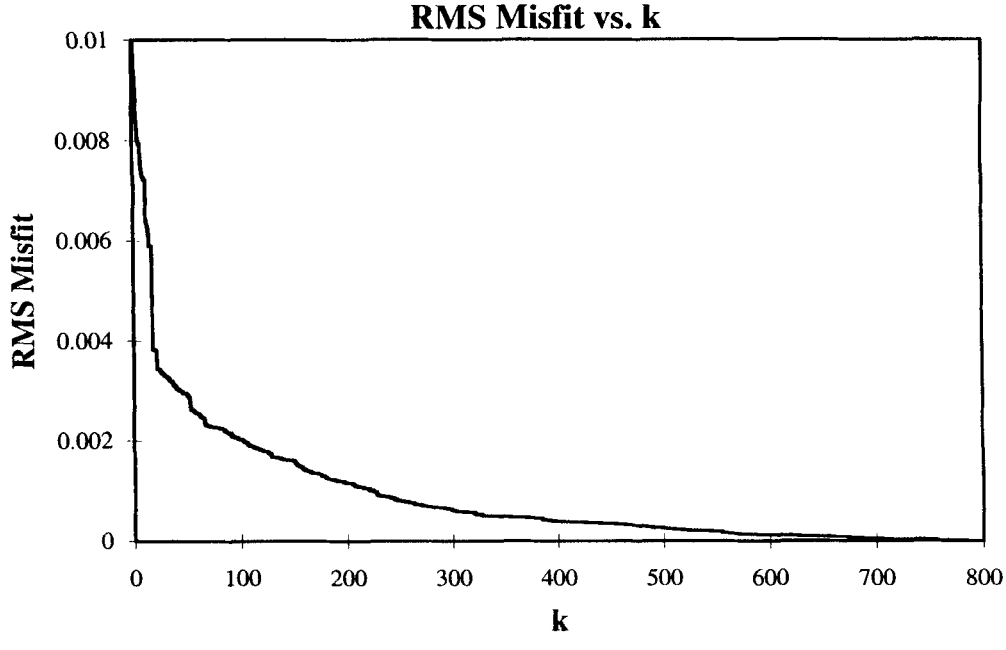
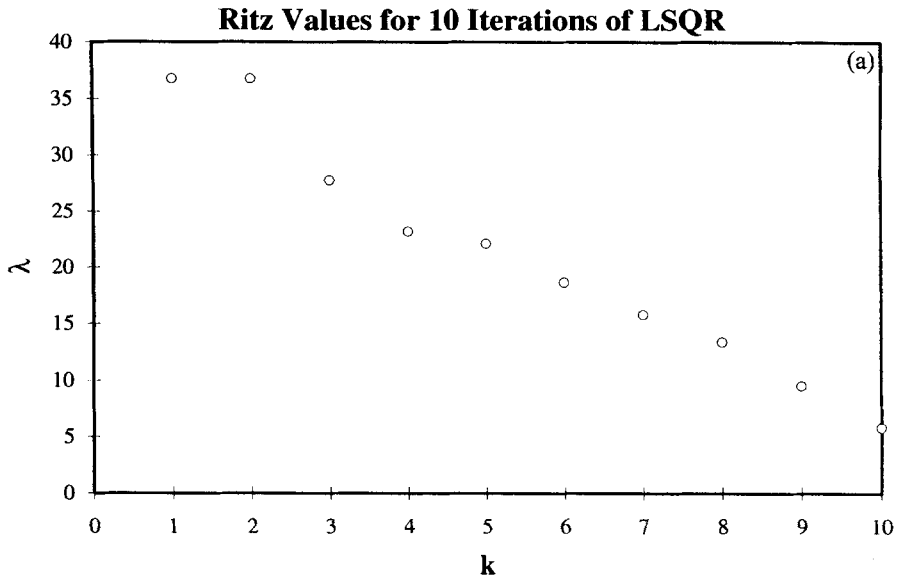
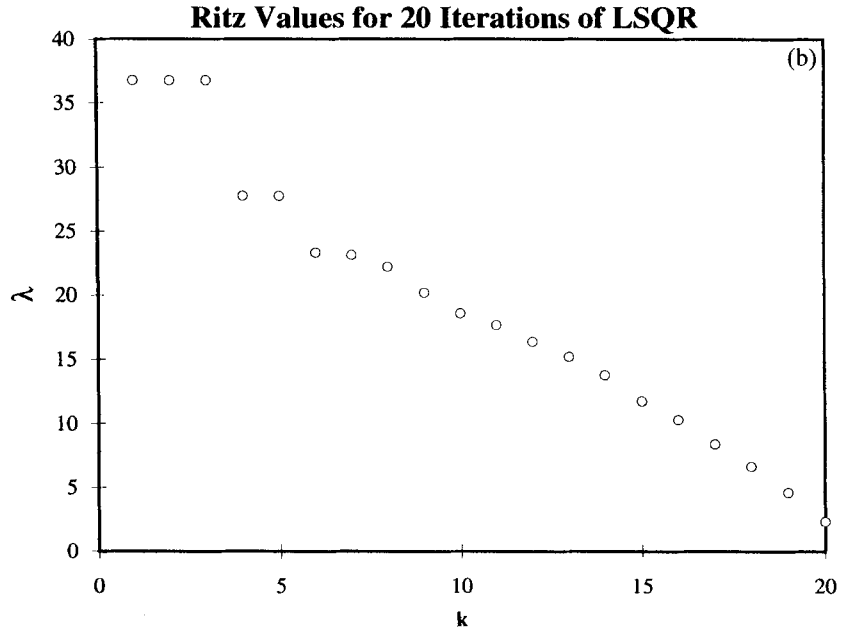
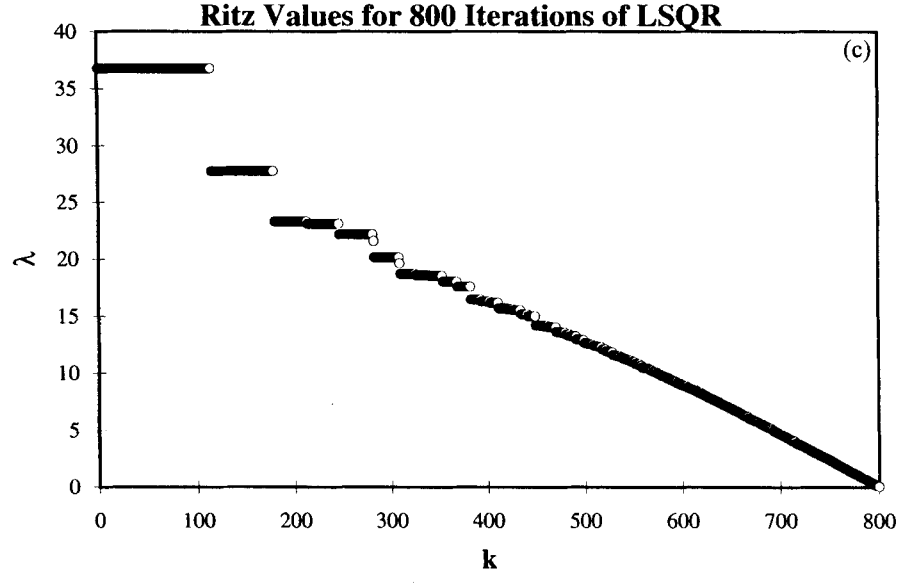


Figure 2. A plot of the RMS misfit value as a function of the number of eigenvectors used in calculating the minimum-norm, least-squares solution. The x-axis corresponds to the variable k in eq. (3).

Figure 3. The effects of increasing the number of iterations in LSQR. (a) shows the eigenvalue spectrum for 10 iterations, (b) for 20 iterations and (c) for 800 iterations. LSQR does not produce a set of unique Ritz values that have a one-to-one correspondence to the eigenvalues calculated using SVD. In our experiments, the repetition of Ritz values is a function of the total number of iterations and it is computationally infeasible to use LSQR to calculate a set of eigenvectors that span the range of $\mathbf{A}^T\mathbf{A}$.







have a greater effect on the solution in exchange for only a slight reduction in data fit (as long as the slope of the misfit plot, Fig. 1, is nearly 0).

LARGE-SCALE PROBLEMS

Separating the range and the nullspace of $\mathbf{A}^T\mathbf{A}$ is simple when the problem is solved using SVD because the eigenvectors that span the range are generated. However, realistic tomographic problems are too large to be solved with SVD. Large systems of equations can be efficiently solved using an iterative method such as LSQR (Paige & Saunders 1982; Nolet 1985; Scales 1987; Nolet & Snieder 1990). After k iterations, the LSQR algorithm approximates the matrix $\mathbf{A}$ as (Paige & Sanders 1982)

$$\mathbf{A} \approx \mathbf{U}_k \mathbf{T}_k \mathbf{V}_k^T, \tag{9}$$

where $\mathbf{T}_k$ is a symmetric, tridiagonal matrix. The eigenvalues of $\mathbf{T}_k$ are known as 'Ritz' values, each associated with a 'Ritz' vector. As k increases, the Ritz values converge to the eigenvalues of $\mathbf{A}$ and the Ritz vectors are transformed eigenvectors (Van der Sluis & Van der Vorst 1987). Generally, LSQR is extremely efficient and a large variance reduction is obtained with only a few iterations. In global tomographic problems, k is easily 3 or 4 orders of magnitude smaller than N or M. However, the knowledge of the (minimum-norm) solution does not yet define the nullspace. With a small-scale example, we shall demonstrate that the computation of the full nullspace is not feasible, as it would require many more than M iterations.

The LSQR algorithm builds a solution based on orthogonal Ritz vectors. Each Ritz vector is associated with a Ritz value that is an approximation to an eigenvalue. The order in which the Ritz values are generated by LSQR is based on arithmetic round-off errors (Cullum & Willoughby 1985, Ch. 4). According to the misfit plot in Fig. 2, use of only the first 300 eigenvectors adequately define the range of $\mathbf{A}^T\mathbf{A}$. However, despite the fact that we can obtain an acceptable least-squares solution with only 10 iterations of LSQR, even 300 iterations of LSQR will not produce the Ritz values corresponding to the first 300 eigenvectors derived from SVD. The method of LSQR first finds the Ritz vectors at each end of the eigenvalue spectrum as well as those associated with clusters of eigenvalues. Each iteration of LSQR produces a new Ritz vector. However, orthogonality is not perfectly maintained in finite precision, and the new vector is often a linear combination of previously determined Ritz vectors (Scales 1989). Increasing the number of iterations increases the amount of repetition among Ritz vectors (Simon 1984).

Fig. 3 demonstrates the effects of increasing the number of iterations. In the case of only 10 iterations (Fig. 3a), the largest Ritz value is duplicated once and the other eight Ritz values are distributed mostly towards the upper end of the eigenvalue spectrum. When the number of iterations is increased to 20 (Fig. 3b) the amount of duplication among the largest Ritz values increases. Ritz values from the lower end of the eigenvalue spectrum also tend to show up with increasing frequency. In the extreme case of 800 iterations (Fig. 3c), 497 of the Ritz values are duplicates and another 100 Ritz values correspond to eigenvectors that have very little effect on the solution and may be included in the generalized nullspace. In connection with the massive repetition of Ritz values, another drawback is that the algorithm cannot recognize when two or more different Ritz vectors should have the same Ritz value. The duplication of Ritz values is caused by the Ritz vectors' loss of orthogonality. There has been considerable amount of work done concerning partial or selective re-orthogonalization of the Ritz vectors. However, this re-orthogonalization can slow down the rate of convergence (Parlett & Scotts 1979; Partlett 1980; Simon 1984). Our experience demonstrates that solving for the eigenvectors that span the range of the model using LSQR is computationally infeasible (Deal & Nolet 1996). Since LSQR does not provide us with a complete set of eigenvectors that span the range of $\mathbf{A}^T\mathbf{A}$ we need to implement a different scheme to perform conservative filtering. For a more complete discussion of conjugate gradients and the related Lanczos algorithm the reader is referred to Parlett (1980) and Cullum & Willoughby (1985).

We can think of $\Delta\mathbf{x}$ as being made up of two vectors, one that lies in the range of $\mathbf{A}^T\mathbf{A}$, and another that lies in the nullspace:

$$\Delta\mathbf{x} = \Delta\mathbf{x}_{\text{range}} + \Delta\mathbf{x}_{\text{null}}. \tag{10}$$

The portion in the nullspace is the desirable vector because this is the component of the filter that does not affect the data fit. At this point, $\Delta\mathbf{x}_{\text{null}}$ and $\Delta\mathbf{x}_{\text{range}}$ are unknowns. In order to preserve the solution's data fit we solve for $\Delta\mathbf{x}_{\text{range}}$. Multiplying both sides by $\mathbf{A}$ yields

$$\mathbf{A}\Delta\mathbf{x} = \mathbf{A}\Delta\mathbf{x}_{\text{range}} + \mathbf{A}\Delta\mathbf{x}_{\text{null}}. \tag{11}$$

By definition, $\mathbf{A}\Delta\mathbf{x}_{\text{null}} = 0$ and $\mathbf{A}\Delta\mathbf{x}$ can be evaluated exactly. If we let $\mathbf{h} = \mathbf{A}\Delta\mathbf{x}$, then our problem reduces to

$$\mathbf{A}\Delta\mathbf{x}_{\text{range}} = \mathbf{h}, \tag{12}$$

where $\Delta\mathbf{x}_{\text{range}}$ is the only unknown. This system of equations can be solved by the method used to solve the original inverse problem, LSQR. Since we are solving for $\Delta\mathbf{x}_{\text{range}}$ it is essential that the algorithm used yields the minimum-norm solution. Once the minimum-norm solution for $\Delta\mathbf{x}_{\text{range}}$ has been determined, the evaluation of $\Delta\mathbf{x}_{\text{null}}$ is trivial, and the final conservatively filtered solution is

$$\mathbf{x}_c = \mathbf{x} + \Delta\mathbf{x}_{\text{null}}, \tag{13}$$

or, equivalently,

$$\mathbf{x}_c = \mathbf{x}_f - \Delta\mathbf{x}_{\text{range}}. \tag{14}$$

Again, $\mathbf{x}_c$ will have the same RMS residual value as the original, unfiltered solution $\mathbf{x}$.

EXAMPLES

We demonstrate the effects of conservative filtering using the synthetic 'data' generated for the model shown in Fig. 4. Our model consists of a uniform background with a 10 per cent velocity anomaly in the shape of the letter P. The synthetic experiment is designed to mimic a cross-hole tomographic problem. The region is divided into a grid of 30×30 square cells; there are 30 sources evenly distributed along the left vertical edge and 30 receivers on the right side. Since the techniques discussed are irrespective of ray geometry, we use a constant-velocity starting model with straight rays in our example. The data consist of 900 traveltime delays on the order of 1 s.

The minimum-norm solution to our synthetic problem is shown in Fig. 5. This solution is obtained by using LSQR with

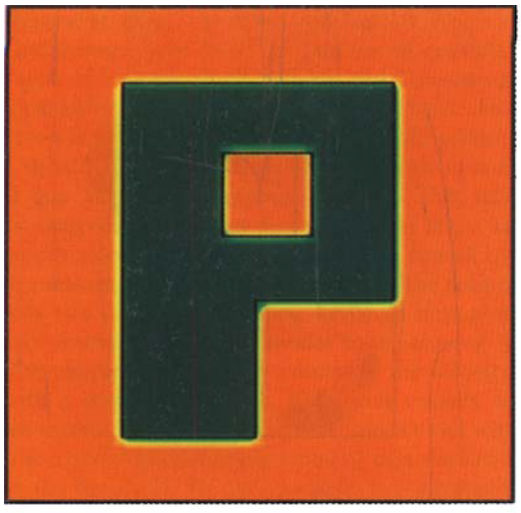


Figure 4: Synthetic starting model

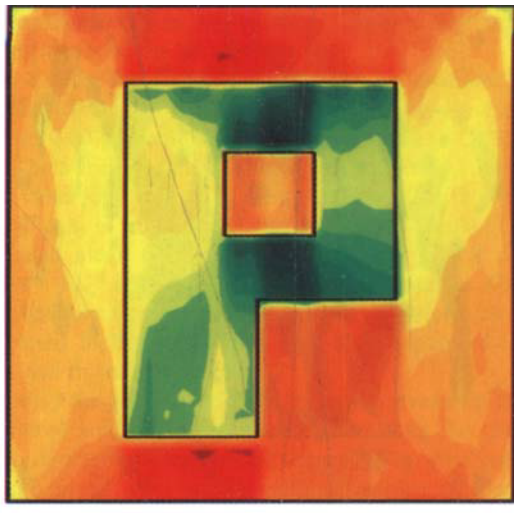


Figure 5: Minimum norm solution

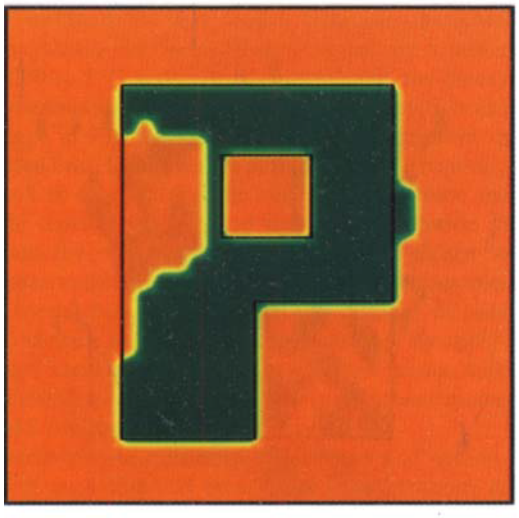


Figure 6: Binary filtered image

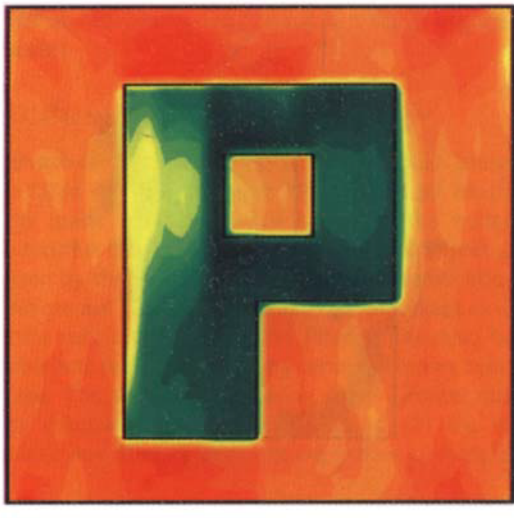


Figure 7: Corrected image

Figure 4. The starting model for the synthetic examples that are based on a cross-hole-type experiment with a constant background velocity and a +10 per cent velocity anomaly. The region of interest is parametrized as a grid of 30 × 30 square cells. The traveltime data are generated using 30 sources and 30 receivers equally distributed along the vertical edges of the model.

Figure 5. A solution to the data generated using the starting model in Fig. 4 with no noise added. This least-squares solution has an RMS residual of 4.44×10^{-5} s, indicating a good fit to the data.

Figure 6. The result from applying a binary filter to the least-squares solution in Fig. 5. The binary filter operates by forcing the velocity in each cell to take on one of two values, the background velocity or the 10 per cent anomaly. Even though the binary image resembles the starting model, its residual value is 9.57×10^{-3} s, indicating a poor fit to the data.

Figure 7. The image obtained using the conservative filtering method and projecting the difference between Figs 5 and 6 onto the nullspace of $\mathbf{A}^{\mathrm{T}}\mathbf{A}$. The conservatively filtered image in Fig. 7 not only resembles the starting model but also has a residual of only 4.44×10^{-5} s, indicating a good fit to the data.

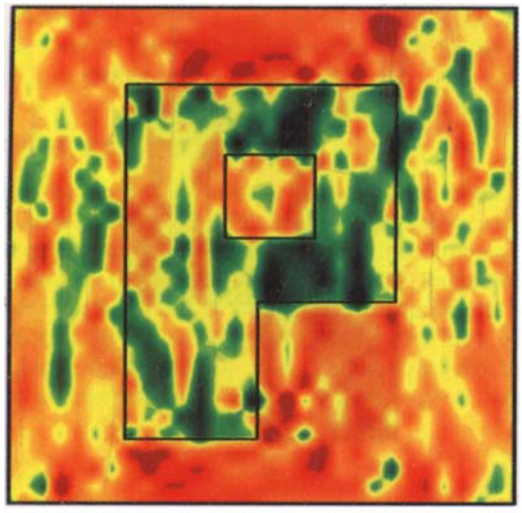


Figure 8: Solution with noise

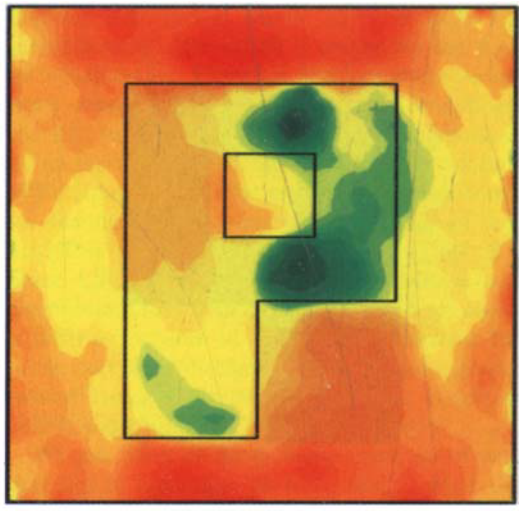


Figure 9: Smoothed solution

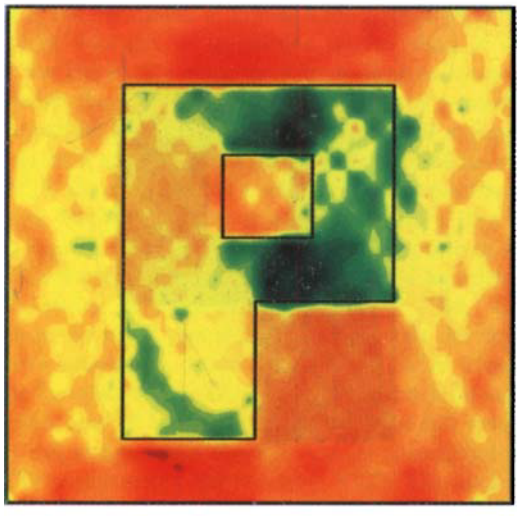


Figure 10: Corrected image

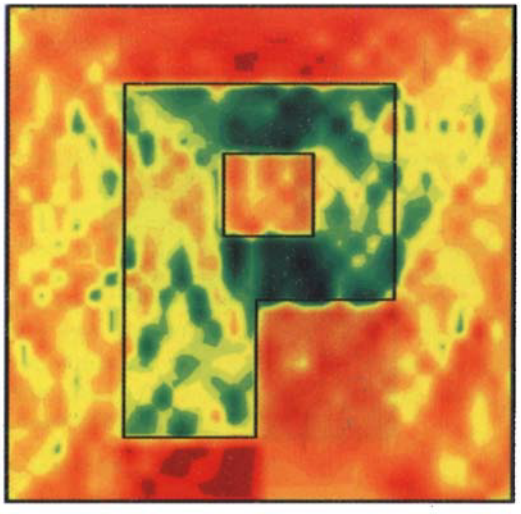


Figure 11: Damped solution

Figure 8. The least-squares solution found if the data generated using the starting model in Fig. 4 are corrupted by Gaussian noise with a standard deviation of 0.20 per cent of the total traveltime. The solution has a residual of 4.44×10^{-4} s.

Figure 9. The α-trimmed mean filtered version of the noisy image in Fig. 8. The α-trimmed mean filter is a combination of a mean and a median filter which is used to reduce noise without excessive blurring of the edges. The filtered image has a residual of 4.90×10^{-3} s, indicating a poor fit to the data.

Figure 10. The result obtained by applying the conservative filtering method to the α-trimmed mean image; the residual is 6.18×10^{-4} s. Most of the noisy artefacts have been removed while preserving a portion on the top half of the 'P'. Since the experiment is based on a pure cross-hole experiment, it is not surprising that the energy in the leg of the 'P' has been smeared out horizontally.

Figure 11. The damped least-squares solution for the synthetic data corrupted by Gaussian noise with one standard deviation equal to 0.20 per cent of the total traveltime.

uncorrupted, noise-free data. The RMS residual for Fig. 5 is 4.44×10^{-5} s, indicating a good fit to the data. As an example, we perform a type of *a posteriori* filtering on Fig. 2 based on *a priori* geological knowledge. For example, we may know that the area of interest is primarily made up of rocks with seismic velocities that should fall in the yellow or green regions and not in the red one. In addition, suppose we know that this region contains sharp discontinuities due to dykes or faults as opposed to smooth velocity variations that may be caused by temperature or pressure gradients. If we knew that the region consisted of only two distinct rock types, we could impose a bimodal velocity distribution on the model by employing a binary filter. The simplest binary filter compares the velocity of each cell with a threshold velocity. If the cell velocity is greater than the threshold then the velocity within the cell will be defined as the *a priori* known upper velocity, and similarly for cells below the threshold. In our example, the threshold velocity used is the average of the background velocity model and a value 10 per cent above the background. This concept can easily be extended to multi-velocity distributions, and the threshold values can be adjusted for different practical situations.

The binary-filtered solution is shown in Fig. 6. Clearly, our binary image resembles the true model. However, the RMS residual for the binary-filter image has increased by a factor of more than 200 to 9.57×10^{-3} s; mathematically, the binary image is an unacceptable solution. If the difference between the binary image in Fig. 6 and the least-squares solution in Fig. 8 is projected onto the nullspace of the model, the resulting components will lie in the nullspace and will not violate the data. If we add these components to the original solution in Fig. 5 we obtain the conservatively filtered image shown in Fig. 7. The conservatively filtered image not only resembles our synthetic model but also has an RMS residual of only 4.44×10^{-5} s, the same as the least-squares solution. By applying the *a priori* knowledge of a binary system along with conservative filtering, a solution resembling the synthetic model with a low RMS residual is quickly obtained.

In the next example, we add noise to the data and demonstrate the use of an α-trimmed mean filter (Gersztenkorn & Scales 1988) in conjunction with the conservative technique. The objective of the α-trimmed mean filter is to smooth the velocity image while trying to preserve edges. An α-trimmed mean value is a combination of the arithmetic mean and the median value and can be expressed as

$$x_\alpha = \frac{1}{n - 2[\alpha n]} \sum_{i=[\alpha n]+1}^{n-[\alpha n]} x(i), \tag{15}$$

when n is odd, and

$$x_\alpha = \frac{1}{n - 2[\alpha(n-1)]} \sum_{i=[\alpha(n-1)]+1}^{n-[\alpha(n-1)]} x(i), \tag{16}$$

when n is even (Gersztenkorn & Scales 1988). The square brackets indicate a greatest integer function and $\{x_i\}$ is an ordered set of n numbers such that $x_1 \leq x_2 \leq \ldots x_n$. If m is an interger, then

$$[\varepsilon] = m \quad \text{where} \quad m \leq \varepsilon \leq m + 1. \tag{17}$$

Filtering of the solution can be carried out by passing a rectangular window or kernel over the image, the α-trimmed mean value for each cell is determined by the velocities of the neighbouring cells in the filtering window. The size of the filtering window can be adjusted.

Fig. 8 is the least-squares solution to the synthetic problem after the addition of Guassian noise with a standard deviation equal to 0.20 per cent of the total traveltime ($\sim$20 ms) to the data. The α-trimmed mean filtered image, using a 3×3 window and a value for α of 0.40, is shown in Fig. 9. Much of the noise from Fig. 8 is eliminated, but the RMS residual for the α-trimmed mean filtered solution is 4.90×10^{-3} s, compared to 4.44×10^{-4} s for the least-squares solution in Fig. 8. The conservatively filtered solution, with a misfit of 6.22×10^{-4} s, is depicted in Fig. 10. In this case, the number of iterations performed in solving eq. (8) is 25.

Clearly, the more *a priori* information available the closer the filtered solution will come to resembling the true Earth. The conservative filtered image shown in Fig. 10 is very similar to the damped least-squares solution shown in Fig. 11. The use of the α-trimmed mean filter helps to demonstrate the use of the projection technique. Using the nullspace shuttle algorithm with *a posteriori* smoothing results in a solution which is nearly equivalent to the damped solution. However, this is just one example of conservative filtering. The advantage of the nullspace shuttle is its ability to be utilized with any number of filters or constraints, depending on what is required by the experiment.

CONCLUSION

Conservative filtering allows a tomographic image to be manipulated without changing the data fit. By restricting the changes made to the solution to lie in the nullspace, we can determine which *a priori* constraints or expectations are permitted by the data. Finding out which modifications to the solution are not supported is equally as important as determining those that are. Conservative filtering can also be viewed as a substitute method for solving certain complex optimization problems. The advantages of using the projection method are (1) the computational convenience and (2) the ability to implement any filter or operator regardless of complexity.

ACKNOWLEDGMENTS

We acknowledge support from NSF grant EAR9204386.

REFERENCES

Cullum, J. & Willoughby, R., 1985. *Lanczos Algorithms for Large Symmetric Eigenvalue Computations,* Birkhauser.

Deal, M.M. & Nolet, G., 1996. Comment on 'Estimation of resolution and covariance for large matrix inversions' by Zang and McMechan, *Geophys. J. Int.,* submitted.

Dines, K.A. & Lytle, R.J., 1979. Computerized geophysical tomography, *Proc. IEEE,* **67,** 1065–1073.

Gersztenkorn, A. & Scales, J.A., 1988. Smoothing seismic tomograms with alpha-trimmed means, *Geophys. J. Int.,* **92,** 67–72.

Jackson, D.D., 1972. Interpretation of inaccurate, insufficient, and inconsistent data, *Geophys. J. R. astr. Soc.,* **28,** 97–109.

Nolet, G., 1985. Solving or resolving inadequate and noisy tomographic systems. *J. Comput. Phys.,* **61,** 463–482.

Nolet, G. & Snieder, R., 1990. Solving large linear inverse problems by projection, *Geophys. J. Int.,* **103,** 565–568.

Paige, C. & Saunders, M., 1982. LSQR: An algorithm for sparse linear equations and sparse least squares, *ACM Trans. Math. SFTW.,* **8,** 43–71.

Parlett, B., 1980. *The symmetric Eigenvalue Problem,* Prentice Hall.

Parlett, B.N. & Scott, D.S., 1979. The Lanczos algorithm with selective orthogonalization, *Math of Comput.,* **33,** 217–238.

Scales, J., 1987. Tomographic inversion via the Conjugate Gradient method, *Geophysics,* **52,** 179–185.

Scales, J.A., 1989. On the use of conjugate gradient to calculate the eigenvalues and singular values of large, sparse matrices, *Geophys. J. Int.,* **97,** 179–183.

Stork, C. & Clayton, R.W., 1991. Linear aspects of tomographic velocity analysis, *Geophys.,* **56,** 483–495.

Su, W.J., Woodward, R.L. & Dziewonski, A.M., 1994. Degree-12 model of shear velocity herterogeneity in the mantle, *J. geophys. Res.,* **99,** 6945–6981.

Simon, H.D., 1984. The Lanczos algorithm with partial reorthogonalization, *Math. of Comput.,* **42,** 115–142.

Wiggins, R.A., 1972. The general linear inverse problem: Implications of surface waves and frequency oscillations for Earth structure, *Rev. Geophys.,* **10,** 251–285.

Woodhouse, J.H. & Dziewonski, A.M., 1984. Mapping the upper mantle: Three dimensional modeling of the Earth structure by inversion of seismic waveforms. *J. geophys. Res.,* **89,** 5953–5986.

Van der Sluis, A. & Van der Vorst, R., 1987. Numerical solution of large, sparse linear algebraic systems arising from tomographic problems, in *Seismic Tomography,* pp. 49–83, ed. Nolet, G., Reidel, Dordrecht.