GEOPHYSICS, VOL. 58, NO. 11 (NOVEMBER 1993); P. 1655–1661, 6 FIGS.

Singular value decomposition for cross-well tomography

Reinaldo J. Michelena*

ABSTRACT

I perform singular value decomposition (SVD) on the matrices that result in tomographic velocity estimation from cross-well traveltimes in isotropic and anisotropic media. The slowness model is parameterized in four ways: One-dimensional (1-D) isotropic, 1-D anisotropic, two-dimensional (2-D) isotropic, and 2-D anisotropic. The singular value distribution is different for the different parameterizations. One-dimensional isotropic models can be resolved well but the resolution of the data is poor. One-dimensional anisotropic models can also be resolved well except for some variations in the vertical component of the slowness that are not sensitive to the data. In 2-D isotropic models, "pure" lateral variations are not sensitive to the data, and when anisotropy is introduced, the result is that the horizontal and vertical component of the slowness cannot be estimated with the same spatial resolution because the null space is mostly related to horizontal and high frequency variations in the vertical component of the slowness.

Since the distribution of singular values varies depending on the parametrization used, the effect of conventional regularization procedures in the final solution may also vary. When the model is isotropic, regularization translates into smoothness, and when the model is anisotropic regularization not only smooths but may also alter the anisotropy in the solution.

INTRODUCTION

In ray theoretic traveltime tomography, the solution of a linear system of equations is the heart of the problem. Solving this linear system transforms variations in traveltimes into variations in model parameters. This transformation from data to model depends on the properties of the matrix that describes the linear system, and singular value decomposition (SVD) is the tool for studying such properties.

SVD has been applied in the past to study the structure of the matrices involved in tomographic traveltime inversion problems. White (1989), Bregman et al. (1989), and Pratt and Chapman (1992), among others, present singular values and singular vectors in model space for cross-well geometries. Stork (1992) also shows singular values and the corresponding singular vectors in model space for the problem of reflection tomography. All these studies, however, have not reported the results of the SVD completely because they make no reference to the properties of the singular vectors in data space.

For a small cross-well geometry, this paper presents the complete results of the SVD (singular values and singular vectors in both data and model spaces) of matrices that result from the following four types of parameterization: 1-D isotropic, 2-D isotropic, 1-D anisotropic, and 2-D anisotropic. The anisotropy is assumed to be elliptical. These four models differ as to how they incorporate the prior information that might be available about the medium. My results show that the model that makes more assumptions about the medium (1-D isotropic) is the one that can be estimated better whereas the model that makes fewer assumptions (2-D anisotropic) contains a large null space that may distort the anisotropy as well as the heterogeneities in the solution. Although these results are not surprising, they remind us that when prior information about the medium is available it is important to incorporate it in the parameterization, because otherwise the results may not even contain the expected features or they may be severely distorted, especially when the medium is anisotropic.

The results of SVD of the previous matrices show how damping the matrix inversion affects the solution when the velocity model is isotropic and anisotropic. As expected, when the model is isotropic, damping the solution results in smoothness in the image. However, when the model is anisotropic, damping not only creates smoother images but also may distort the anisotropy or create artificially anisotropic results.

Manuscript received by the Editor May 15, 1992; revised manuscript received February 19, 1993.
*Stanford University, Department of Geophysics, Stanford, CA 94305.

This paper is an example of the type of analysis that can be performed for any recording geometry to gain insight into how data and model parameters relate. This insight can help to improve both the data acquisition and the estimation of the parameters.

THE LINEAR SYSTEM

Regardless of how the model is described, the problem of ray theoretic traveltime tomography always reduces to the solution of a linear system of equations of the form:

$$\underset{\sim}{\mathbf{J}}\mathbf{m} = \mathbf{t}, \tag{1}$$

where $\underset{\sim}{\mathbf{J}}$ is a matrix whose elements depend on the raypaths and on how the model is described, **m** is the vector of model parameters, and **t** is the vector of measured traveltimes. The vectors **m** and **t** may also represent *variations* with respect to a given model and to measured traveltimes, respectively.

The model space represented by **m** consists of two separate models: one for the heterogeneities and one for the velocities. If the model for velocities is isotropic, the elements of **m** usually represent the coefficients of the expansion of the slowness model in a basis function that describes the model of the heterogeneities. Moreover, if the basis function is also orthogonal (i.e., its different elements do not overlap), each component of **m** represents the slowness within one particular region in space. Although other basis functions that don't have the property of orthogonality have been recently proposed (Harlan, 1989; Van Trier, 1988; Michelena and Harris, 1991), those that have such a property are still the most widely used to represent 1-D layered models (large, rectangular cells) and arbitrary 2-D variations (small, square cells). In this paper, I focus on these two types of basis function to describe the model of heterogeneities. In both cases, the isotropic slowness model can be expressed as

$$S(x, z) = \sum_{j=1}^{N} S_j R_j(x, z), \tag{2}$$

where $R_j(x, z)$ is nonzero only at the jth cell, S_j is the slowness within that cell, and N is the total number of cells (either layers or square pixels). The corresponding vector of model parameters is

$$\mathbf{m} = (S_1, S_2, \ldots, S_N)^T. \tag{3}$$

When the model for velocities is anisotropic, our choices for defining the global model space **m** increase substantially because all the models available in the isotropic case for describing the heterogeneities are now combined with all the different models available for describing the anisotropy. The selection of the proper combination velocity-model/heterogeneity-model should be made according to any prior knowledge we might have about the medium. The examples that follow show that introducing prior information in the proper way in each of these two models helps to estimate both of them more accurately but, unfortunately, if one of the models is incorrect or too general, the results obtained with the other model may be also incorrect.

In this paper, I assume that the anisotropy is elliptical and that the model of heterogeneities is described either by horizontal homogeneous layers or by square pixels (Michelena et al., 1993). Even though most rocks are not elliptically anisotropic, by using elliptical anisotropy I can show some of the difficulties we may encounter when tomographically estimating variations in velocity with both direction and position. When the anisotropy is elliptical, the representation of the slowness images is a generalization of expression (2) as follows:

$$S_x(x, z) = \sum_{j=1}^{N} S_{x_j} R_j(x, z), \tag{4a}$$

$$S_z(x, z) = \sum_{j=1}^{N} S_{z_j} R_j(x, z). \tag{4b}$$

The corresponding vector of model parameters is

$$\mathbf{m} = (S_{x_1}, S_{x_2}, \ldots, S_{x_N}, S_{z_1}, S_{z_2}, \ldots, S_{z_N})^T, \tag{5}$$

where S_{x_j} and S_{z_j} are the horizontal and vertical components of the slowness, respectively.

If the vector **m** is not transformed into an image(s), it is difficult to understand the relations among its different components. The same applies for the vector **t** of traveltimes. A simple method of mapping **t** into an image has been used in recent publications (see, for example, Squires et al., 1992). The mapping consists of plotting each component of **t** at its corresponding source-receiver location in a 2-D space where the axes are source depth and receiver depth. Traveltimes corresponding to sources and receivers at the same depth map into the diagonal of the image and other traveltimes corresponding to rays that travel at nonzero angle with respect to the horizontal map off the diagonal. This transformation of the vector **t** into an image and the transformations (2) and (4) are used extensively in the next sections to visualize the vectors in both data and model space that result from the SVD.

SVD: A SHORT REVIEW

Any $M \times N$-matrix $\underset{\sim}{\mathbf{J}}$ can be decomposed in the following way (Golub and Van Loan, 1989):

$$\underset{\sim}{\mathbf{J}} = \underset{\sim}{\mathbf{U}}\underset{\sim}{\mathbf{L}}\underset{\sim}{\mathbf{V}}^T, \tag{6}$$

where $\underset{\sim}{\mathbf{U}}$ is an $M \times M$ orthonormal matrix of eigenvectors that span the data space, $\underset{\sim}{\mathbf{V}}$ is an $N \times N$ orthonormal matrix of eigenvectors that span the model space, and $\underset{\sim}{\mathbf{L}}$ is an $M \times N$ diagonal matrix whose elements are the singular values of $\underset{\sim}{\mathbf{J}}$. The columns of $\underset{\sim}{\mathbf{U}}$ (**u**) are the eigenvectors of $\underset{\sim}{\mathbf{J}}\underset{\sim}{\mathbf{J}}^T$, and the columns of $\underset{\sim}{\mathbf{V}}$ (**v**) are the eigenvectors of $\underset{\sim}{\mathbf{J}}^T\underset{\sim}{\mathbf{J}}$. The decomposition (6) is called singular value decomposition.

When a singular value is zero, the corresponding singular vector in data space cannot be mapped into model space or vice versa. Data vectors or model vectors with zero singular value belong to the null space and cannot be resolved. When a singular value is not zero but is small compared with the largest one (i.e., the condition number is large), the contribution of the corresponding eigenvectors to the solution must be eliminated or attenuated, that is regularized, because the matrix inversion may become unstable.

SVD: APPLICATION

SVD (Dongarra et al., 1979) was performed on the matrix $\mathbf{J}$ after describing the model space as described by equations (2) and (4). To represent the results of the SVD, I show the singular values simultaneously with the singular vectors in data and model space, both sets of vectors having been transformed into images as explained in the previous section. This representation of the SVD results follows Pratt and Chapman (1992), with the addition of the singular vectors in data space.

The ray geometry used to compute the SVD for the different parameterizations is shown in Figure 1. It is the same as the one used by Pratt and Chapman (1992): five sources and five receivers in separate wells in a constant slowness medium. When the model is isotropic, the matrix $\mathbf{J}$ depends only on the ray geometry, and when the model is anisotropic, $\mathbf{J}$ depends on both the ray geometry and the slowness model (which is constant in this case), which means that even when the rays are straight the tomographic estimation of velocity anisotropy is, in general, a nonlinear problem. Ray bending adds another nonlinearity to the problem.

Isotropic models

Figure 2 shows the SVD when the model is discretized using six horizontal isotropic layers [equation (2)]. The differences among the singular values are small, which means that the problem is well-conditioned. The largest singular values correspond to singular vectors in data and model space ($\mathbf{u}$ and $\mathbf{v}$, respectively) whose components are roughly of the same magnitude. With this parameterization only some "big structures" (averages) in data space can be explained whereas in model space all the parameters can be well resolved. By representing the singular vectors in data space $\mathbf{u}$ as images, it is possible to identify source-receiver pairs whose traveltimes belong to the null space and therefore cannot be resolved or have no influence in the estimation of the model parameters. For this reason, errors in these particular traveltimes (noise) will also have little or no effect in the solution, which means that this parameterization is not too sensitive to errors in the data.

Allowing lateral variations in the previous parameterization results in a matrix $\mathbf{J}$ whose SVD is shown in Figure 3. The largest singular value corresponds roughly to horizontal layers (in model space) and nonhorizontal rays (in data space). As the singular values decrease, the eigenvectors in model space tend to contain more horizontal and high-frequency variations and the eigenvectors in data space tend to span near and far vertical offsets (diagonal and nondiagonal structures in the data images). In model space, the smallest singular values correspond to "pure" horizontal variations, which means that the data is not sensitive to this type of variation in the model. In data space, the smallest singular values correspond to rapid changes among nearby traveltimes that have little or no influence on the model. Rapid changes among nearby traveltimes might be produced by noise that, unfortunately, is not necessarily confined to the less influential part of the data. Therefore, in some applications it might be necessary to damp the effect of singular values larger than those contained in the null space to attenuate the effect of certain components of the noise.

The results shown in the previous two figures are as expected. In both cases the largest eigenvalues correspond to gross features in both model and data space. When few parameters compared with the number of data points are used (Figure 2), the data is not well resolved, and when the number of parameters is increased, some components of the model (pure lateral variations, for example) may be difficult or impossible to retrieve from the given data, even if the problem is overdetermined as in Figure 3.

Even though these results were expected, they have received little attention. The discretization of the model in square pixels assumes that we don't know anything about the spatial variations in the medium, unlike the discretization of the model in layers. Since the discretization of the model in layers is a subset of the discretization of the model in square pixels, we may think that whatever can be estimated by using homogeneous layers can be also estimated by using homogeneous square pixels. What Figures 2 and 3 show is that this statement is not necessarily true. The estimation of the parameters depends on how the data and parameters relate to each other. In problems of tomographic estimation of velocities, 2-D inversions are done often in places that are known to be isotropic and horizontally layered to account for all possible "unexpected" variations in the medium. The extra degrees of freedom (and the null space) introduced in the inversion have to be penalized appropriately in the objective function, which produces an image with less reso-

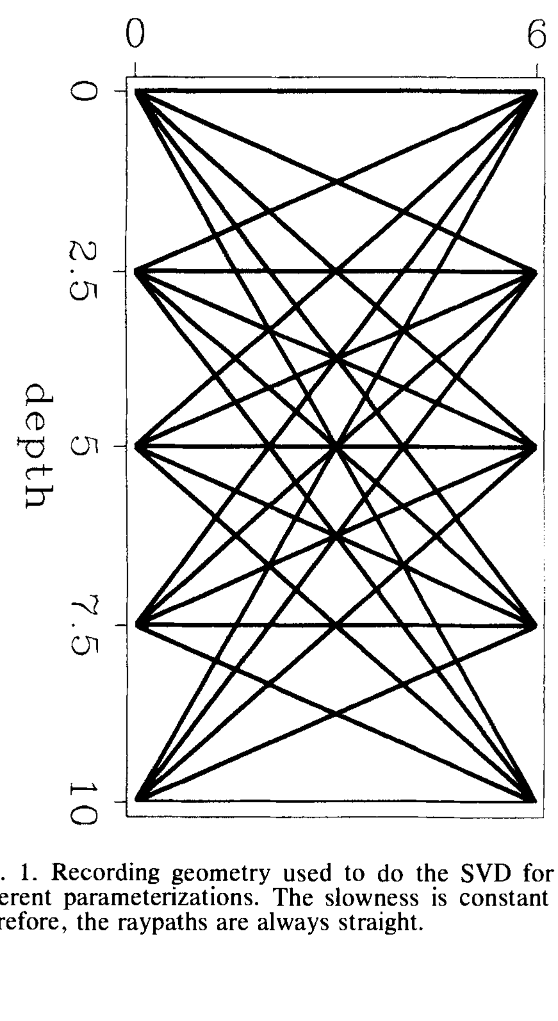

FIG. 1. Recording geometry used to do the SVD for the different parameterizations. The slowness is constant and therefore, the raypaths are always straight.

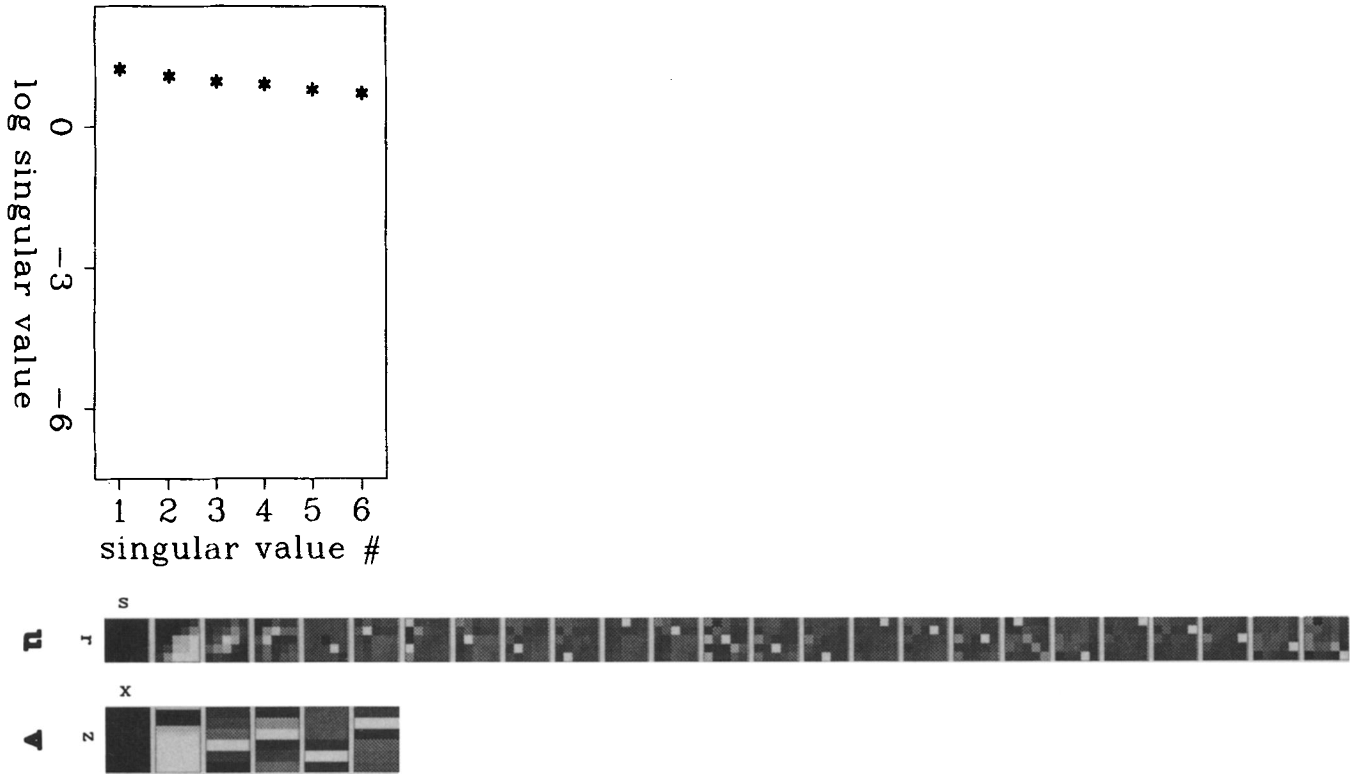

FIG. 2. Singular value decomposition when the model is described as a superposition of six horizontal isotropic layers (1-D isotropic). Vector **u** represents the singular vectors in data space and **v** represents the singular vectors in model space. The origin of the axes is at the upper left corner of each image. Parameter r is the receiver axis, s is the source axis, x is the horizontal distance, and z is the depth. Most singular vectors in data space are in the null space. The gray scale goes from black (negative) to white (positive).

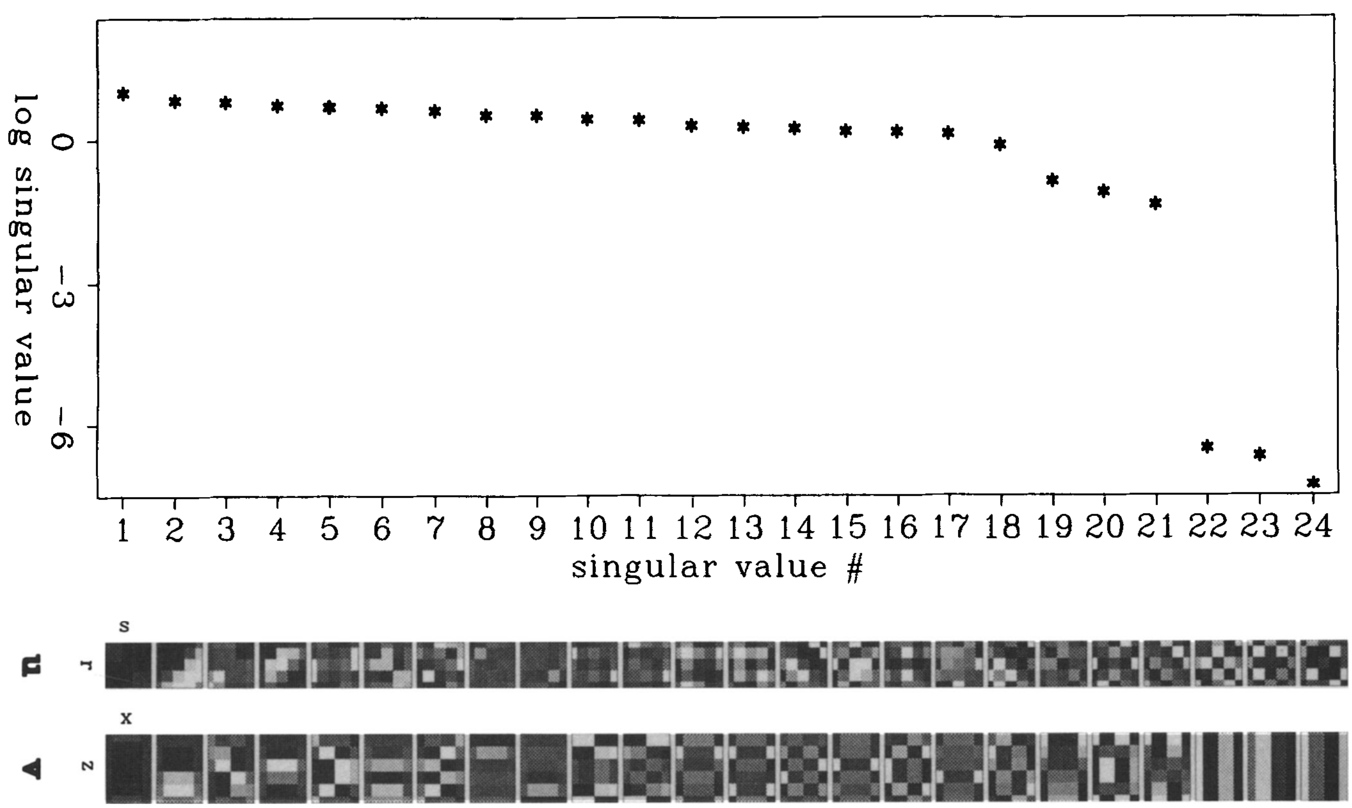

FIG. 3. Singular value decomposition when the model is described as a superposition of 6 × 4 homogeneous isotropic squared regions (2-D isotropic). The amount of lateral variation in the model space increases as the size of the singular values decreases.

lution overall than the one obtained by directly estimating the parameters of a 1-D model. Of course, if well logs are available and the medium is known to be isotropic and horizontally layered, 1-D inversions are not interesting, and the intrinsic advantages of the parameterization (fewer unknowns and better conditioning) are not useful. However, as the next section shows, if the medium is anisotropic and known to be horizontally layered, using a model of heterogeneities that appropriately incorporates such prior information can make the difference between retrieving or not (accurately) variations of velocity as a function of direction.

The large number of data singular vectors **u** contained in the null space in Figure 2 also suggests that, if the medium is known to be isotropic and horizontally layered, the data acquisition can be optimized to increase the number of measurements that influence the solution, which results in a better estimation of the velocities.

Anisotropic models

The SVD for a 1-D anisotropic parameterization [equation (4)] is shown in Figure 4. The upper half of each eigenvector in model space corresponds to $S_x(x, z)$, the lower half to $S_z(x, z)$. As expected, the largest singular values correspond to gross features in both model and data space. In descending order of singular value, the corresponding singular vectors in data space first span $S_x(x, z)$, then $S_z(x, z)$. The least sensitive part of the model (singular values 11 and 12) is spanned by vectors that contain only information about $S_z(x, z)$. In data space, the behavior for the largest singular values is similar to the isotropic 1-D case.

Figure 5 shows the SVD when the model is 2-D anisotropic. The result for this parameterization is roughly a combination of the SVD for the 2-D isotropic and the 1-D anisotropic model (Figures 3 and 4 respectively); that is, vertical variations in $S_x(x, z)$ correspond to the largest singular values and horizontal and high-frequency variations in $S_z(x, z)$ to the smallest ones. Nearly half the vectors in the null space (Figure 6) contain information about $S_z(x, z)$ only, and the other half contain information about both $S_x(x, z)$ and $S_z(x, z)$. These vectors cannot be estimated from the data.

Figures 4, 5, and 6 show that when we introduce anisotropy in the model, the sensitivity of the data to the vertical component of the slowness is lower than the sensitivity to the horizontal component, which is not a surprise for cross-well geometries that don't adequately sample the vertical direction. This limitation, however, doesn't impede an accurate estimation of variations of velocity anisotropy with position if we use the proper model to describe the heterogeneities at the same time, as shown in Figure 4, where most singular vectors in model space correspond to large singular values.

When using a model that assumes nothing about the heterogeneities (square pixels), estimating spatial variations in slowness anisotropy may become a very difficult task because we have to deal with the features of the medium about which the data give less information: horizontal and high-frequency variations in the vertical component of the slowness. Even if the inversion can retrieve the singular vectors corresponding to the smallest singular values, the

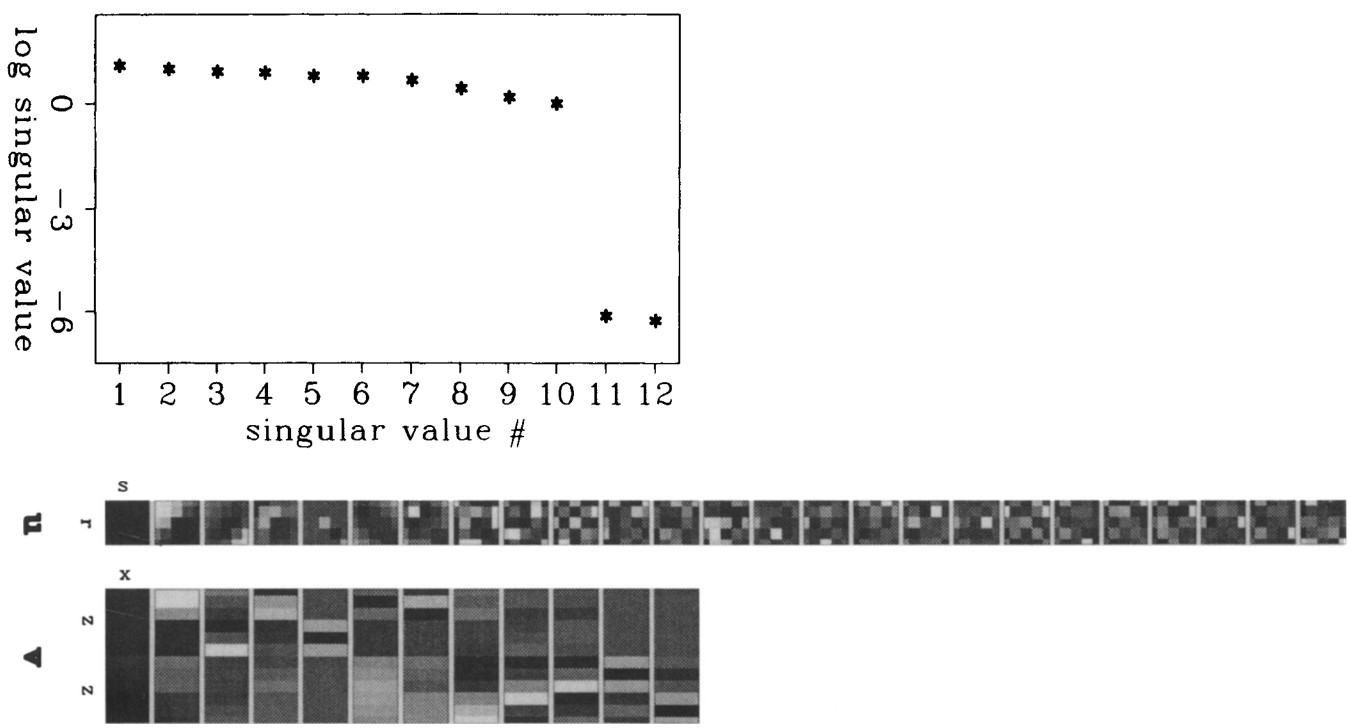

FIG. 4. Singular value decomposition when the model is horizontally layered and anisotropic (6 × 2 model parameters). The upper half of each image in the model space corresponds to $S_x(x, z)$, and the lower half corresponds to $S_z(x, z)$. The origin in the data space is at the upper left corner of each image.

result can still be images with different resolutions for the horizontal and the vertical components of slowness because most vectors in the null space are related to the vertical component (Figure 6).

For these reasons, performing 2-D inversions in places known to be 1-D may create serious problems, in particular when the model is anisotropic. From Figure 4 we see that all variations in $S_x(x, z)$ can be retrieved from the data because the smallest singular values are related to $S_z(x, z)$ only. However, when we allow 2-D variations in the model, several components of $S_x(x, z)$ go to the null space, as Figure 6 shows. This fact has two implications. First, features that could be easily recovered with one parameterization have become more difficult or impossible to recover with another parameterization that is more general. Second, taking 1-D averages or smoothing 2-D images across the horizontal direction is not necessarily the same as performing true 1-D inversions, because the 2-D images may be less accurate and contain more artifacts than the images obtained using 1-D parameterizations.

When the velocity model is isotropic, a common way to deal with the noise and ill-conditioning when solving the system (1) is by damping the least-squares solution or by truncating the SVD. The purpose of these two techniques is to attenuate or eliminate the effect of the smallest singular values of the problem. When the model is isotropic, damping translates into smoothing because what is being attenuated are the high-frequency and horizontal variations in the model. However, when the velocity model is anisotropic, damping out the smallest singular values affects not only the smoothness of the model but also its anisotropy (or isotropy) because the effect of the vertical component of the slowness compared with the horizontal has also been reduced. Therefore, common techniques used to regularize the problem in isotropic media may not be adequate in anisotropic media because they may introduce artificial anisotropy.

Besides damped least-squares or SVD truncation, conjugate gradients (CG) is another common way to solve the system (1). In practice, if the data energy distribution among the different singular values is even, early CG iterations tend

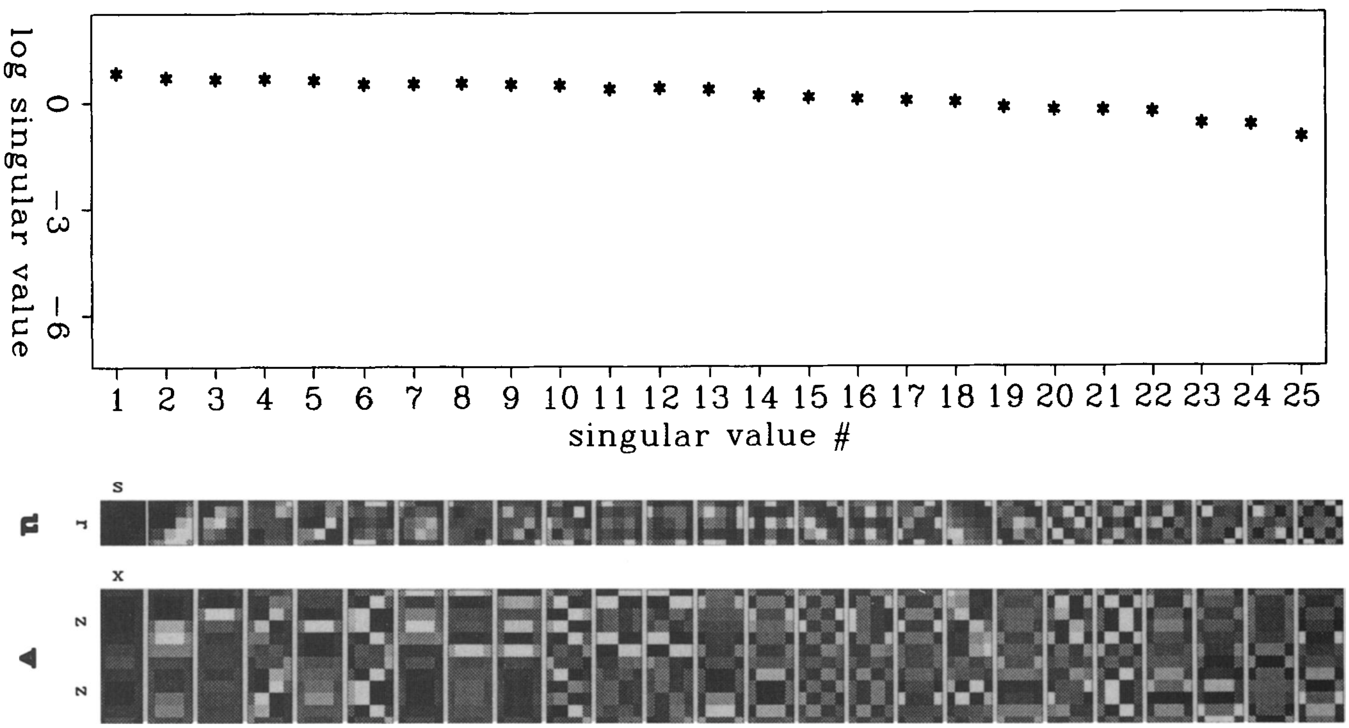

FIG. 5. Singular value decomposition when the model is 2-D anisotropic (24 × 2 model parameters).

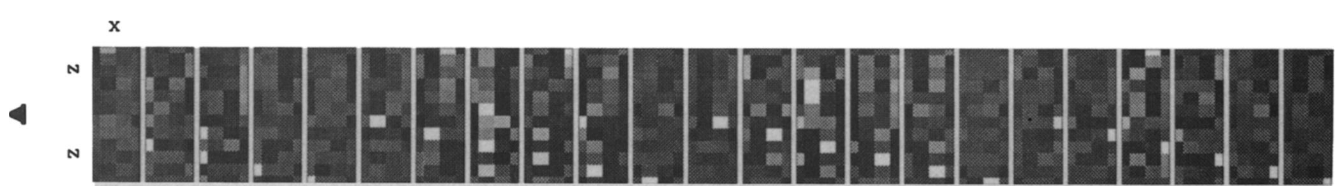

FIG. 6. Vectors that span the null space of the model for the SVD shown in Figure 5. Most vectors contain information about $S_z(x, z)$ (nonzero components in the lower half of each image), and therefore, $S_z(x, z)$ cannot be estimated at the same resolution of $S_x(x, z)$ from cross-well traveltimes alone.

to be more sensitive to the largest singular values, whereas later iterations tend to be affected by both large and small singular values (Stork, 1988). For this reason, stopping the CG iterations after a small number of steps is similar to the effect of damping or of truncating the SVD (Scales and Gersztenkorn, 1988). As a consequence, when the model is anisotropic, early truncation of the CG iterations may also produce artificially anisotropic results because the horizontal component of the slowness converges faster than the vertical, which belongs to the less sensitive part of the model. Michelena et al. (1993) show examples of how the two components of the slowness converge at different speeds. Early truncation of the iterations may be necessary because of noise or ill-conditioning.

The effect of damping can also be seen in data space. On the one hand, when the damping is large, only gross features in data spaced are resolved. On the other hand, when the damping is small or zero, the high-frequency variations in the data that correspond to the smallest singular values can also be resolved. Therefore, depending on the amount of damping (or, equivalently, where the SVD solution is truncated or when the CG iterations are terminated) some portions of the data may be better resolved than others, which has to be taken into account when interpreting traveltime residuals.

CONCLUSIONS

By performing the SVD of the matrices that result from a small scale numerical experiment, I have shown the relations between data and model space for four different parameterizations. The parameterizations vary according to the amount of prior information that they contain about the medium.

All the results have in common that the largest singular values correspond to gross features in both data and model space. The main differences among the results are in the type of feature in data and model space that the small singular values represent, the size of the null space, and the effect of regularization when dealing with such insensitive parts of the data and the model. When the model is 1-D isotropic, the problem is well conditioned and all the parameters can be resolved well but the resolution of the data is poor. For this type of model, cross-well traveltime tomography performs the best if the medium is also 1-D isotropic.

I have generalized the 1-D isotropic model in three ways: by allowing the layers to vary lateraly, to be anisotropic, and to be both heterogeneous and anisotropic. The effect of lateral heterogeneities in the data was negligible even when the problem was overdetermined. Lateral heterogeneities also introduced into the model high-frequency variations whose influence in the inversion needs to be attenuated. The effect of anisotropy in 1-D was to introduce structures in the vertical component of the slowness that are not sensitive to the data. Most other vertical variations, however, can still be easily retrieved. The effect of lateral variations and anisotropy in the parameterization was to create a large null space in the model related mostly to horizontal and high-frequency variations in the vertical component of the slowness. This means that when 2-D anisotropic models are used anisotropy and heterogeneity cannot be estimated with the same resolution, no matter how simple the real medium is. Hence the importance of using the appropriate parameterization when information about the medium is available beforehand.

Since the singular value distribution is different for the different parameterizations, the effect of conventional regularization procedures such as damping, SVD truncation, or a simple early termination of CG iterations is also different when each of these parameterizations is used. When the model is isotropic, regularization translates into smoothness in the resolution of both data and model spaces. When the model is anisotropic, diminishing the effect of the smallest singular values in the solution not only creates smoother images but may also introduce anisotropy where it doesn't actually exist or, more generally, may distort the anisotropy of the medium.

ACKNOWLEDGMENTS

I wish to thank Larry R. Lines for his help while I did this work at the Amoco Research Lab during the summer of 1991. I also want to thank Jerry M. Harris for many useful comments. The comments and suggestions of Associate Editor William S. Harlan are also gratefully acknowledged.

REFERENCES

Bregman, N. D., Bailey, R. C., and Chapman, C. H., 1989, Ghosts in tomography: the effects of poor angular coverage in 2-D seismic traveltime inversion: Can. J. Expl. Geophys., **25**, 7–27.

Dongarra, J. J., Bunch, J. R., Moler, C. B., and Stewart, G. W., 1979, Linpack: Users' Guide: Soc. Ind. Appl. Math., Chapter 11.

Golub, G. H., and Van Loan, C. F., 1989, Matrix computations: Johns Hopkins Univ. Press.

Harlan, W. S., 1989, Tomographic estimation of seismic velocities from reflected raypaths: 59th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 922–924.

Michelena, R. J., and Harris, J. M., 1991, Tomographic traveltime inversion using natural pixels: Geophysics, **56**, 635–644.

Michelena, R. J., Muir, F., and Harris, J. M., 1993, Anisotropic traveltime tomography: Geophys. Prosp., **41**, 381–412.

Pratt, R. G., and Chapman, C. H., 1992, Traveltime tomography in anisotropic media: Application: Geophys. J. Intl., **109**, 20–37.

Scales, J. A., and Gersztenkorn, A., 1988, Robust methods in inverse theory: Inverse Problems, **4**, 1071–1091.

Squires, L. J., Blakeslee, S. N., and Stoffa, P. L., 1992, The effects of statics on tomographic velocity reconstructions: Geophysics, **57**, 353–362.

Stork, C., 1988, Comparison of Richardson's iteration with Chebyshev acceleration factors to conjugate gradient iteration: Stanford Expl. Proj. SEP-**57**, 479–503.

——— 1992, Singular value decomposition of the velocity-reflector depth tradeoff, Part 2: High-resolution analysis of a generic model: Geophysics, **57**, 933–943.

Van Trier, J., 1988, Migration velocity analysis using geological constraints: 58th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 897–900.

White, D. J., 1989, Two-dimensional seismic refraction tomography: Geophys. J., **97**, 223–245.