*Stanford Exploration Project, Report 75, November 17, 1997, pages 103–107*

# Short Note

# Unitary operators in inversion

*Jon F. Claerbout*[1]

***keywords:*** *linear, velocity, modeling, interpolation*

## INTRODUCTION

Geophysical model building springs from a variety of related formulas ranging from inversion to back projection. When the number of unknowns is huge, as when we seek an image of the earth's reflectivity, then using the conjugate operator (also known as back projection or migration) is always a possible starting technique. The conjugate operator can generally be improved by any transformation that pushes it towards a unitary operator. *Unitary operators play an important role in imaging practice but they do not have an obvious place in geophysical inversion theory.* Here I review some examples and point to our need to better understand how to find "preconditioning operators."

## REVIEW

We define abstract vectors for data $\mathbf{d}$, model $\mathbf{m}$ and a linear operator $\mathbf{B}$ and we define "modeling" by

$$\mathbf{d} \quad = \quad \mathbf{B}\mathbf{m} \tag{1}$$

Many practical problems stem from nonlinearity, but these are not the focus of this short note which ignores nonlinearity.

Conventional theoretical analysis generally involves singular-value decomposition which begins from an eigenvalue analysis of $\mathbf{B}'\mathbf{B}$ where $\mathbf{B}'$ is the conjugate transpose of $\mathbf{B}$. Such analysis is meaningless unless the components of $\mathbf{d}$ have identical physical dimensions, likewise for the components of $\mathbf{m}$. When data components have various physical dimensions, the arbitrariness can be resolved by dividing each component by its estimated variance. This still leaves the model scaling unspecified. Variance scaling leads to questions of covariances, and soon other questions arise.

[1]**email:** not available

Setting aside eigen analysis, the well-known theoretical inverse to equation (1) is

$$\mathbf{m}_{\text{theoretical}} \quad = \quad (\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{d} \tag{2}$$

The well-recognized problem with equation (2) is that the matrix may not be invertible. This problem is acknowledged by the so-called "damped" solution.

$$\mathbf{m}_{\text{damped}} \quad = \quad (\mathbf{B}'\mathbf{B} + \epsilon)^{-1}\mathbf{B}'\mathbf{d} \tag{3}$$

where $\epsilon$ hides a multitude of possible models of increasing sophistication, beginning from a small scalar times an identity matrix, going to a diagonal matrix of inverse variances, going on to an inverse covariance matrix that requires considerable imagination to conjure up, or considerable skill to estimate.

Beginning instead from the practical side things, the first step going from data space to model space is given by the conjugate operator

$$\mathbf{m}_{\text{image}} \quad = \quad \mathbf{B}'\mathbf{d} \tag{4}$$

I teach (Claerbout, 1992) that computation of the conjugate is a straightforward adjunct to the operation itself and I advocate that no program that embodies a linear operator is complete without the few extra steps required to make the conjugate option available. There exist in competitive industrial practice, imaging operators that amount to nothing more than the conjugate operator. The conjugate operation, equation (4), is always a starting point for data analysis and how much further we should go toward equation (2) is an issue that needs reexamination in each application.

The unadorned conjugate operator gives an answer that is generally incorrect in its scale factor. In imaging, however, the scale factor is generally chosen according to the image variance, so this is not an important issue. In more general applications, however, scaling can be very important, and to be meaningful, a minimum requirement is that components of $\mathbf{d}$ scaled to each have the the same physical dimension, and likewise for $\mathbf{m}$. More generally, scaling can involve statistical models of increasing complexity. We cannot hope to pursue this complexity through all its logical ramifications, yet we cannot ignore it because of the many opportunities to learn more from our data that may be hidden there, particularly in the covariance structure.

## THE WINNERS

For applications with small numbers of unknowns, say less than dozens or hundreds, the damped inverse operator is generally the most appropriate, whereas for problems when the unknown model is a field of more than thousands of unknowns, say reflectivity or velocity as a function of two coordinates, then the conjugate operator and its various extensions are, in practice, the answer.

Between the conjugate operator and the inverse operator lies an operator (often overlooked) that is sometimes a guide to the most effective compromise. This operator is a certain unitary operator. The solution and the unitary operator are given by

$$\mathbf{m}_{\text{unitary}} \quad = \quad [(\mathbf{B}'\mathbf{B})^{-1/2}\,\mathbf{B}']\mathbf{d} \tag{5}$$

If $\mathbf{B}'\mathbf{B}$ tends to a singular matrix, then $(\mathbf{B}'\mathbf{B})^{-1/2}\,\mathbf{B}'$ is like $0/0$ whereas $(\mathbf{B}'\mathbf{B})^{-1}\,\mathbf{B}'$ is like $0/0^2$ so equation (5) is much safer than (2). Equation (5) is exceedingly ambiguous because of the huge multiplicity of ways of representing the square root of an operator. For an operator of $N$ eigenvalues there are $2^N$ square roots corresponding to choice of sign of each. Generally, there exists an application-specific reason for selecting one of the square roots. We might seek a symmetric phaseless operator, for example, or we might seek a causal one. A clean and illustrative example of such an operator is unitary normal moveout with linear interpolation (Claerbout, 1986) or (1992), page 118. This example begins from the observation that linear interpolation causes smoothing which suppresses high frequencies and noticing that it would be nice if the transformation conserved energy. The linear interpolation is represented mathematically by an NMO transformation matrix $\mathbf{B}$ containing adjacent elements that sum to unity (linear interpolation) leading to $\mathbf{B}'\mathbf{B}$ being a tridiagonal matrix. To build unitary NMO, the tridiagonal matrix $\mathbf{B}'\mathbf{B}$ is factored into a bidiagonal matrix and its conjugate, say $\mathbf{A}'\mathbf{A}$ by an obvious method that is known as the Cholesky decomposition. Then (5) reduces to

$$\mathbf{A}\mathbf{m}_{\text{factored}} \quad = \quad \mathbf{B}'\mathbf{d} \tag{6}$$

which is easily solved for $\mathbf{m}$. by recursion. Formally, the unitary operator is $\mathbf{A}^{-1}\mathbf{B}'$.

More typically, in practice the square-root operator is not easily found, nor is it needed precisely. This leads to

$$\mathbf{m}_{\text{precond}} \quad = \quad \ll \mathbf{B}'\mathbf{B} \gg^{-1/2} \;\; \mathbf{B}'\mathbf{d} \tag{7}$$

which has the same form as (5) except that we build various simplifications into the definition of $\ll \mathbf{B}'\mathbf{B} \gg$. A notable example of this concept arises in time-series analysis where the symbols $\ll\ \gg$ denote ensemble smoothing and the factorization is not done by the Cholesky algorithm but by that of Kolmogoroff described in (Claerbout, 1992) on page 230.

One of the most important linear operators in seismic exploration is velocity analysis (summing along hyperbolic trajectories of various velocities). After having defined such a transform, you can push data into velocity space and pull it back again with the operator $\mathbf{B}\mathbf{B}'$. Unless by chance or design you have defined your operator to be already unitary, you will find, when comparing the original data $\mathbf{d}$ to $\mathbf{B}\mathbf{B}'\mathbf{d}$ that the processing weakens events at various times, locations, or frequencies. The remedy for this is to redefine the operator with various scale factors, such as $\sqrt{-i\omega}\sqrt{x/v}\sqrt{x/t}\sqrt{1/t}$ that you can uncover either experimentally, theoretically, or by looking at Claerbout (1992) All this activity amounts to looking for a unitary representation such as equation (7). In my experience in geophysics, such efforts are not as systematic as they should be and could be. We should make a greater effort to digest the literature on preconditioning. Further, the Fortran language itself inhibits segregation of numerical-analysis concepts from application-specific concepts. Thus successful applications of advanced numerical analysis concepts are many fewer than we would like to see. This motivates a C++ study in this report (Nichols et al., 1992).

Finally, I wish to cite another example of operator factorization for the inverse needed by (7). That example is the two-dimensional prediction-error operator in Abma and Claerbout, (Abma and Claerbout, 1992). The confusion caused by that paper at the 1992 SEP sponsor meeting is one reason I prepared this short note. The other reason, is to bring up again our lack of systematic use of preconditioning when we attack huge inverse problems. To begin with, I believe that the practical search for unitary operators is closely related to the practical search for preconditioners, though I have not formally confirmed it.

## REFERENCES

Abma, R., and Claerbout, J. F., 1992, Finding the amplitudes of migration hyperbolas from the kinematics: SEP–**73**, 235–242.

Claerbout, J. F., 1986, Pseudo unitary NMO tutorial: SEP–**48**, 347–350.

Claerbout, J. F., 1992a, Anti aliasing: SEP–**73**, 371–390.

Claerbout, J. F., 1992b, Earth Soundings Analysis: Processing Versus Inversion: Blackwell Scientific Publications.

Nichols, D., Dunbar, G., and Claerbout, J. F., 1992, The C++ language in physical science: SEP–**75**, 455–472.