Stanford Exploration Project, Report 94, November 11, 1997, pages 1–137

Stanford Exploration Project, Report 94, November 11, 1997, pages 1–137

Short Note

Diagonal weighting: An elementary challenge to mathematicians

Jon Claerbout [1]

keywords: *adjoint, imaging, inversion, least-squares, regression, weighting*

INTRODUCTION

Geophysical mapping and imaging are applications where we seek an approximate pseudo inverse of a matrix of very high order. Say, $\mathbf{d} = \mathbf{Fm}$ constructs theoretical data $\mathbf{d}$ from model parameters $\mathbf{m}$ using a linear operator $\mathbf{F}$. Experience shows that the transpose of the simulation operator provides a useful image. Thus the suggested image is $\hat{\mathbf{m}} = \mathbf{F}'\mathbf{d}$ where $\mathbf{d}$ is the observed data, and $\mathbf{F}'$ is the transpose (Hilbert adjoint) of $\mathbf{F}$. Mathematically, this means that when the dimensionality is very high, we often approximate an inverse by a transpose $\mathbf{F}^{-1} \approx \mathbf{F}'$.

Experience shows that a "better" image is often created by a unitary operator. Practitioners often discover improvements in the form of scaling diagonal matrices, say $\mathbf{F}^{-1} \approx \mathbf{D}_1\mathbf{F}'\mathbf{D}_2$ where $\mathbf{D}_1$ and $\mathbf{D}_2$ are the scaling diagonal matrices. We seek an operator $\mathbf{G} = \mathbf{D}_1\mathbf{F}'\mathbf{D}_2$ that is as unitary as possible, i.e., $\mathbf{G}'\mathbf{G} \approx (\mathbf{G}'\mathbf{G})^2$. We believe that $\mathbf{G}$ is also in some sense ideally preconditioned for linear solvers. A basic, recurring problem is a lack of straightforward theory for finding $\mathbf{D}_1$ and $\mathbf{D}_2$.

PROBLEM SETTING

A reflectivity model $\mathbf{m}$ in geophysics is a function defined in a physical volume sampled densely enough to fill a computer's random access memory. The volume typically has somewhat fewer than about $(10^3)^3 = 10^9$ elements (voxels). The data $\mathbf{d}$ upon which the model is built is a hundred to a thousand times larger, say 10^{11} elements and generally requires many magnetic tapes. Thus the matrix $\mathbf{F}$ relating the data to the model would have 10^{20} elements (which is a futile way to think about the problem). What we can readily do is manufacture synthetic data from any model by running a

[1]**email:** jon@geo.stanford.edu

program that amounts to a matrix multiply

$$\mathbf{d}_{\text{theoretical}} \quad = \quad \mathbf{F}\mathbf{m} \tag{1}$$

By paying attention to some details it is straight-forward (in all the problems that I have encountered) to write another program that applies the transpose $\mathbf{F}'$ operator to the data $\mathbf{d}$ giving us a first guess at the model. This first guess is called an image $\hat{\mathbf{m}}$ of the model $\mathbf{m}$.

$$\hat{\mathbf{m}} \quad = \quad \mathbf{F}'\mathbf{d}_{\text{observed}} \tag{2}$$

The method of least squares offers the standard solution

$$\hat{\mathbf{m}} \quad = \quad (\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\mathbf{d}_{\text{observed}} \tag{3}$$

which is wholly impractical because of the size of $\mathbf{F}'\mathbf{F}$. This suggests we seek a diagonal matrix, say $\mathbf{D}_m$ as an approximation for $(\mathbf{F}'\mathbf{F})^{-1}$.

$$\hat{\mathbf{m}} \quad = \quad \mathbf{D}_m\mathbf{F}'\mathbf{d}_{\text{observed}} \tag{4}$$

At my request, Bill Symes offered a suggestion for $\mathbf{D}_m$. His suggestion uses the **diag** operator of matrix algebra which simply takes a vector and distributes its components along the diagonal of a matrix:

$$\mathbf{D}_m \quad = \quad \frac{\mathbf{diag}\ \mathbf{F}'\mathbf{d}}{\mathbf{diag}\ \mathbf{F}'\mathbf{F}\mathbf{F}'\mathbf{d}} \quad = \quad \mathbf{diag}\ \frac{a}{b} \tag{5}$$

This suggestion seems reasonable because the denominator has two more powers of $\mathbf{F}$ than does the numerator and the suggestion seems wise because it uses the known data $\mathbf{d}$. Unfortunately, the denominator b can easily vanish. To prevent such trouble we rearrange and smooth

$$\mathbf{D}_m \quad = \quad \mathbf{diag}\ \frac{< ab >}{< b^2 > + \epsilon^2} \tag{6}$$

where ϵ is a positive number for experimentation and the numerator and denominator can be locally averaged as indicated by the "$< >$". (Remember that the vectors $\mathbf{m}$ and $\mathbf{d}$ are sampled representations of continuous functions of space and time.) Our experience with Symes' suggestion was disappointing, probably for the reason suggested next.

Besides the standard pseudo inverse $\hat{\mathbf{m}} = (\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\mathbf{d}$ there is another

$$\hat{\mathbf{m}} \quad = \quad \mathbf{F}'(\mathbf{F}\mathbf{F}')^{-1}\mathbf{d}_{\text{observed}} \tag{7}$$

that arises when a fitting problem is underdetermined. This pseudoinverse solves the problem of minimizing $||\hat{\mathbf{m}}||$ subject to the constraint equations $\mathbf{d} = \mathbf{F}\mathbf{m}$. This pseudoinverse suggests a right side diagonal approximation

$$\hat{\mathbf{m}} \quad = \quad \mathbf{F}'\mathbf{D}_d\mathbf{d}_{\text{observed}} \tag{8}$$

I don't expect this right-side diagonal to work any better than did the earlier left-side diagonal. Experience with special cases suggests we need both the data-space diagonal and the model-space diagonal, namely

$$\hat{\mathbf{m}} \quad = \quad \mathbf{D}_m \mathbf{F}' \mathbf{D}_d \mathbf{d}_{\text{observed}} \tag{9}$$

Our challange for the mathematicians is to give us some guesses for the diagonal matrices $\mathbf{D}_m$ and $\mathbf{D}_d$. This problem is one of a very general nature, far more general than the several applications in which I have found it. It is like the question, "What is an all-purpose preconditioner?"

HINTS

Naturally, if we are given specific information about the operator $\mathbf{F}$ we can hope for better diagonal multiplers. Most of the practical problems I am familiar with are similar to the Radon transform except that straight lines in the Radon transform become various hyperbolas and hyperboloids.

The building blocks, as you see from the Symes guess and our follow-on guesses, include using the data vector $\mathbf{d}$. We may also use vectors of ones and vectors of random numbers in both data space and model space. We can take spatial derivatives in the geophysical model space $\mathbf{m}$, and likewise space and time derivatives in the data space $\mathbf{d}$. (The Radon transform special case shows that we might even need half-order derivatives.) We also have square roots of scalars and we have smoothing.

UNITARITY

Believe it or not, we often have problems where we are unsure of the precise form of $\mathbf{F}$. This can arise when we know a travel time but are uncertain about an amplitude. It also arises when we are unsure about a source radiation pattern. It also arises when we scan a parameter like in a velocity spectrum.

We begin from an operator $\mathbf{F}$ that is somehow wrong in that it should really have pre and post diagonal operators applied, say $\mathbf{G} = \mathbf{D}_d \mathbf{F} \mathbf{D}_m$. Although we may be unsure how to define the $\mathbf{D}_d$ and the $\mathbf{D}_m$, we are quite sure we will be happy with an operator that conserves the energy in the original data. Thus we seek approximate idempotence,

$$\mathbf{G}'\mathbf{G} \quad \approx \quad (\mathbf{G}'\mathbf{G})^2 \tag{10}$$

ROW PARTITIONED OPERATORS

A related problem arises with partitioned operators. Fitting observed data to theoretical data where there are two classes of model parameters $\mathbf{m}_1$ and $\mathbf{m}_2$ we seek to minimize the norm of the residual $\mathbf{r}$ defined by

$$\mathbf{0} \quad \approx \quad \mathbf{r} \quad = \quad [\mathbf{A} \quad \mathbf{B}] \begin{bmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \end{bmatrix} - \mathbf{d} \tag{11}$$

The problem of interest arises when we seek the best diagonal scaling matrices before and after $\mathbf{A}$ and $\mathbf{B}$.