Stanford Exploration Project, Report 79, November 16, 1997, pages 241–243

Short Note

Preconditioning

Jon Claerbout and Dave Nichols[1]

keywords: *random, deconvolution, nonlinear*

In our C++ project we plan to supply reusable least-squares solvers. This led to a question that we are only beginning to answer, "are there any helpful general purpose preconditioners?" The one described in this note applies to operators that are representable in partitioned form, such as the regression

$$\begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \quad \approx \quad \mathbf{y} \tag{1}$$

We came up with two possibilities for rescaling $\tilde{\mathbf{x}} = \mathbf{M}\mathbf{x}$ before going to a conjugate-gradient (CG) solver. These two possibilities are:

$$\mathbf{M} \quad = \quad \begin{bmatrix} \|\mathbf{A}_1'\mathbf{y}\| & 0 \\ 0 & \|\mathbf{A}_2'\mathbf{y}\| \end{bmatrix} \tag{2}$$

$$\mathbf{M} \quad = \quad \begin{bmatrix} \|\mathbf{A}_1\mathbf{r}_1\| & 0 \\ 0 & \|\mathbf{A}_2\mathbf{r}_2\| \end{bmatrix} \tag{3}$$

where $\mathbf{r}_1$ and $\mathbf{r}_2$ contain random numbers and where $\|\mathbf{v}\|$ is some norm such as the square root of the sum of the squares. The basic idea of preconditioning is that you solve $\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{M}^{-1}\mathbf{M}\mathbf{x} \approx \mathbf{y}$ in the form $(\mathbf{A}\mathbf{M}^{-1})\tilde{\mathbf{x}} \approx \mathbf{y}$ by the usual CG procedure and then solve $\mathbf{M}\mathbf{x} \approx \tilde{\mathbf{x}}$. The new system should converge faster if $\mathbf{A}\mathbf{M}^{-1} \approx \mathbf{I}$. (In general, everyone is left to his or her own ingenuity for finding $\mathbf{M}$. We believe that gain control and deconvolution operators are good prospects for use in guessing a good $\mathbf{M}$).

We dislike (3) because we don't know how to specify the random numbers. We know the numbers should not be all "1" because d/dx and ∇^2 are important operators and their norm is clearly non zero whereas it might appear to be zero if the random numbers were taken to be all "1". On the other hand should the random numbers be all positive or both positive and negative? We are not sure we can make a prior specification that would be generally satisfactory. While positive random numbers might be safe enough, it is worth considering (2).

[1]**email:** not available

Can it really be that (2) contains an idea not already inside the CG solver? After all, the first step of CG computes the sum of these two terms. Apparently yes, there is an important idea missing from the CG solver which has no prior knowledge that the operator can be partitioned into the $\mathbf{A}_1$ and $\mathbf{A}_2$ parts.

Partitioned problems of the kind discussed here frequently arise from nonlinear optimization where there are two partial derivative matrices with different physical dimensions. Since the problems are nonlinear and the operators themselves change as the iteration proceeds, it is natural to think of changing the preconditioning scalars as the iteration proceeds. In the linear case, however, rescaling with iteration would mess up the CG solver. We asked Michael Saunders about the existence of nonlinear solvers for this type of problem, but he did not have any immediate leads.

An interesting close-to-home example of a partitioned regression problem is found in PVI page 182. This example is a one-dimensional missing data problem. There the $\mathbf{x}_1$ vector contains the 3-5 components of a prediction-error filter (dimensionless transfer function) whereas the $\mathbf{x}_2$ vector would have the typical 1-2 thousand time samples of a seismogram (units of pressure, for example). We have not tested the two alternative preconditioners suggested in this paper because we have not yet done the programming. Also, a more meaningful evaluation will arise in a multidimensional context where preconditioning is more important, and where more programming, and perhaps some more research, are required.

Two apparently related references are (Kennett et al., 1988) and (Golub and Varah, 1974).

REFERENCES

Golub, G. H., and Varah, J. M., 1974, On a characterization of the best L2-scaling of a matrix: SIAM J. Numer. Anal., **11**, no. 3, 472–479.

Kennett, B. N., Sambridge, M. S., and Williamson, P. R., 1988, Subspace methods for large inverse problems with multiple parameter classes: Geophysics, **94**, 237–247.