Reply to comment by M. M. Deal and G. Nolet on 'Estimation of resolution and covariance for large matrix inversions'

Jie Zhang and George A. McMechan

Center for Lithospheric Studies, The University of Texas at Dallas, PO Box 830688, Richardson, TX 75083-0688, USA

Accepted 1996 June 12. Received 1996 May 14; in original form 1995 August 3

We appreciate the comments made by Deal & Nolet (1996, hereafter referred to as D&N) on our paper (Zhang & McMechan 1995, hereafter referred to as Z&M). We first review the nature of the LSQR solution and then address the main points raised by D&N.

The LSQR algorithm of Paige & Saunders (1982) is an iterative method for obtaining a least-squares solution to large sparse linear problems. The main idea behind our LSQRA method is the realization that Paige & Saunders’ LSQR algorithm, when expressed in its analytical form, also provides a formal inverse operator [eqs (6a) and (A8) in Z&M]. Based on this inverse operator and the definitions of resolution and covariance, we approximate resolution and covariance matrices for the iteratively generated solutions [eqs (7)-(9) and (10)-(13) in Z&M; Berryman 1994].

For large linear problems, the number of iterations needed to obtain an acceptable solution is smaller than the full SVD rank of the original matrix. Nonetheless, the extremal Ritz values are often good approximations to the corresponding singular values of the original matrix (Golub & Van Loan 1989, p. 479), as shown by the examples in both Z&M and D&N, and in Scales (1989); the computed Ritz values are spaced across the range of the singular values of the original matrix with a distribution defined by zeros of Jacobi polynomials (van der Sluis & van der Vorst 1990).

To explore the solution subspace more fully, additional iterations are performed beyond that at which an acceptable solution is defined. After a sufficient number of iterations, Ritz values converge to singular values but with the consequences that Lanczos and Ritz vectors are no longer orthogonal; some duplicate Ritz values and vectors are generated. The selective orthogonalization method of Parlett & Scott (1979) addresses this loss of orthogonality at the cost of increased algorithm complexity and reduced numerical efficiency. We choose to address this problem by identifying and eliminating the duplicate Ritz values and Ritz vectors (also called Ritz pairs) (Scales 1989; Parlett 1980, p. 272), and to use the remaining (nearly) orthogonal Ritz pairs to construct our LSQRA solution. It should be noted that it is not necessary to approximate an LSQRA solution to get the resolution and covariance matrices; the Lanczos decomposition is sufficient.

The Ritz vectors computed at a small number of LSQR iterations do not span the entire possible solution space, as pointed out by D&N, but they do support the approximate solution at the current iteration. This is also true in truncated SVD solutions where the less resolvable eigenvectors associated with small singular values are deliberately omitted. In the example in D&N, in which 600 out of 900 singular values are kept, the computed resolution matrix has a structure very similar to that when all 900 were used. Thus, computation of resolution provides an estimate of solution uniqueness in that part of the model space spanned by the basis vectors used. D&N are concerned with the traditional ('global') estimate of resolution, which clearly requires a broad sampling of the model space, whereas the Ritz vector approach provides a 'local' estimate of uniqueness in the neighborhood of the current solution. The approximate resolution matrix is only correct with respect to a small, but relevant, subspace; thus, it may be anticipated that it also provides correct information for data domain projections that make a small angle with the formed subspace (van der Vorst, personal communication, 1996). In Z&M, our assertion was not that the LSQR resolution matrix at a small number of iterations was comparable to the full SVD resolution (for large problems), but that it asymptotically approaches the latter as the number of iterations increases. In general, it will be desirable to perform a substantial (model- and data-dependent) number of iterations beyond that where the loss of orthogonality occurs, to broaden the sampling of the model space for resolution computation. This is why the ability to deal with the loss of orthogonality is important, and why D&N correctly emphasize this point.

Ritz vectors contain information from the observed data vector as well as the sensitivity matrix, so the Ritz pairs from LSQR are necessarily different from the eigenstructures produced by SVD. Indeed, they correspond to a different parameterization of the problem; however, the corresponding resolution can be computed, and is relevant to that parameterization (i.e. those Ritz pairs). It is not unusual to base the model parameterization on information from the data and the survey geometry (e.g. Michelena & Harris 1991; Michelena 1993).

In this context, the trivial numerical example given in D&N can be understood as follows. As the two equations in the example are completely decoupled, our LSQRA method can be applied to each separately to obtain the two solutions (1, 0, 0) and (0, 1, 0), and each of these solutions is uniquely resolved. If we couple the two equations into one linear system, we get the results in D&N. The equal values of the first two unknowns in the solution (1, 1, 0) and their weights (1/2, 1/2) in the resolution matrix for the coupled system suggest that the problem should be reparametrized by combining the first and second unknowns into one parameter; i.e. the same medium has been sampled by two (redundant) rays. The problem then becomes...
A = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix},

\text{with}

d = \begin{pmatrix} 1 \\ 1 \end{pmatrix}

which has the solution \((1, 0)^T\), of which the first element is uniquely resolved, and the second (which is the third, unsampled, parameter in the original problem) is not resolved:

\[ R = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}. \]

The two original bases \((1, 0, 0)\) and \((0, 1, 0)\) are reduced to one basis \((1, 0)\) in the new parametrization, equivalent to \((1, 1, 0)\) in the original parametrization.

The model parametrization is one form of a priori knowledge that is normally included into the construction of matrix A. Jackson (1979) suggests the equivalence of the effects of a priori constraints and of data on a model. By analogy, the inclusion of the data vector into the problem, as in LSQRA, provides a priori constraints that implicitly reduce the effective dimension of the model space, and hence the number of basis vectors that are needed to characterize the model [see eq. (2.7) in Paige & Saunders 1982]. The Ritz vectors are the reduced basis set and correspond to a nearly unique (resolved) solution in their subspace. When projected into the (larger) original parametrization space, as done by ourselves and by D&N, the apparent resolution is proportionally reduced by decomposition in that space, as indicated by the lower values of the diagonal elements and the presence of non-zero off-diagonal values. The apparent resolution is lower in any space that is overparametrized; the numerical example above illustrates this relation. Generally speaking, such low resolution suggests the need for more efficient model parametrization. Possibilities include using variable pixel sizes (larger where the resolution is lower) or a spatial Fourier or wavelet representation (which corresponds to appropriate resolution of the features actually present at various scales).

D&N made an observation that the diagonal elements of resolution kernels computed using the LSQRA method are often one or two orders of magnitude smaller than those obtained from a full SVD, when the number of LSQRA iterations is much smaller than the dimensions of the original matrix, yet are still able to produce a solution similar to the SVD solution. From the example provided by D&N, it is clear that the average amplitude of the diagonal elements equals the ratio of the number of eigenstructures (or Ritz pairs) used in the solution to the total number of parameters. For the SVD computations in their example, 900/900 = 0.67. Similarly, for their LSQRA computations, 37/900 = 0.04 and 9/900 = 0.01. If the prime objectives in computing resolutions are to examine the relative values of, and the correlations between values of, elements in each kernel of the resolution matrix, these can be achieved with the information in the LSQRA resolution matrix, independent of the magnitudes of its diagonal elements. Normalization can be performed if desired. Moreover, as more independent Ritz pairs emerge with increased iterations (past the loss of orthogonality as the solution converges), the average amplitude of the diagonal elements of the LSQRA resolution matrix continues to approach 1.0, but at the expense of deteriorated parameter variances. Therefore, the values of the LSQRA resolution kernels should also be evaluated in the light of the LSQRA covariance (e.g. example 2 in Z&M).

One may exploit any other iterative method (e.g. Söderström & Stewart 1974; Nakanishi & Suetsugu 1986) that yields an inverse operator for the construction of resolution kernels. For instance, although we used eqs (11)-(13) to estimate parameter resolution kernels for all three of the examples in Z&M, we could use \( R = B^* A = V L^{-1} U^T A \) (see the discussion in Z&M) or utilize \( B^* = V (F F^T)^{-1} F^T U^T \) (M.A. Saunders 1996, personal communication). In the latter, \( F \) is an upper bidiagonal matrix complementary to \( L \); \( F \) is the same as \( B \) in Paige & Saunders (1982). The amount of computation in these formulas is still manageable, limited by the dimension of \( L \) (i.e. the number of iterations).

This topic is clearly one of substantial practical importance. There remain a number of questions, especially regarding the appropriate dimensions of the Krylov subspace. The number of converged Ritz pairs that are adequate for construction of resolution should be defined in this more general context; these questions should be resolved by further research.

ACKNOWLEDGMENTS

This response benefited from constructive comments by H.A. van der Vorst.

REFERENCES


