APPROXIMATE MINIMUM NORM SUBSPACE PROJECTION OF LEAST SQUARES WEIGHTS WITHOUT AN SVD

M J Smith and I K Proudler

Defence Evaluation and Research Agency,
St. Andrews Road, Malvern,
Worcestershire, WR14 3PS,
UK.

E-mail: mjs@signal.dera.gov.uk

ABSTRACT

A QR based technique is presented for estimating the approximate numerical rank and corresponding signal subspace of a matrix together with the subspace projection of the least squares weights. Theoretical difficulties associated with conventional QR factorisation are overcome by applying the technique of Row-Zeroing QR to the covariance matrix. Thresholding is simplified compared with the use of the data matrix as the diagonal value spectrum is sharpened and the subspace estimate is improved. An approximation to the minimum norm solution for the projection of the least squares weight onto the signal subspace of the data is obtained simply, without performing an SVD.

1. INTRODUCTION

A common requirement in adaptive processing of sensor array data is the estimation of the number of strong signals present and the projection of both the data and the least squares adaptive weight vector onto the signal subspace. Projection of the least-squares weight vector onto a subspace of reduced dimension is an established technique for reducing the number of adaptive degrees of freedom used by an adaptive sensor array. Conventional algorithms for subspace estimation based upon eigenvalue decomposition (EVD) or singular value decomposition (SVD) are, however, both expensive to compute and difficult to make recursive. By contrast, algorithms based upon ordinary QR factorisation have established pipelineable architectures[6] for calculation of the least squares weight vector but are generally unreliable for rank and subspace estimation[2]. Chan's rank-revealing QR factorisation[1] (RRQR) transforms QR factorisation to a form guaranteed to reveal the rank of a rank-deficient matrix by the use of column permutation operations. Despite the much reduced computational load, it is not simple to design real-time hardware for the RRQR algorithm. It suffers from the same problems as the EVD or SVD in that the need to permute the columns of the matrix makes it a two-sided process and it also requires an indeterminate number of iterative steps.

A computationally simpler technique for estimating the number of strong signals and projecting the least-squares weight vector onto the signal subspace has previously been presented in both block and recursive forms[7][8]. Based upon QR factorisation, the technique is ideally suited to the adaptive null-steering problem because the least-squares weight is readily obtained from the QR factorisation by backsubstitution i.e. no matrix inversion is required. This one-sided projection and rank estimation technique (OSPRE) [7], however, lacks a formal guarantee to reveal the rank of an arbitrary rank-deficient matrix. In fact, there is a well-known counter-example[5] for which the algorithm fails. In this paper, it is shown that this same technique applied to the QR factorisation of the covariance matrix produces a much better estimate of the signal subspace and also overcomes the problem case. The least squares weight vector can still be obtained by backsubstitution and an approximation for the minimum norm solution to the projected weight is obtained for only a little further effort.

We begin by reviewing the Row-Zeroing QR factorisation technique which forms the basis of OSPRE in section 2. In section 3 we discuss the numerical linear algebra technique known as Orthogonal Iteration. The new algorithm is introduced in section 4. How the counter example due to Kahan [5] is overcome when the new technique is used is discussed in section 5. The results of a computer simulation are presented in section 6.

2. ROW-ZEROING

The technique of Row-Zeroing in the QR factorisation[7] enables rank and subspace estimation without the need for iteration or permutation. Consequently it is faster and is more readily made recursive than Chan's RRQR algorithm although it produces a more approximate estimate of the rank and the signal subspace[8]. The approach is essentially QR factorisation computed using a version of modified Gram-Schmidt orthogonalisation. It proceeds in the usual way, constructing the R matrix R_0, say, row by row, but when a small diagonal element below some threshold α, say, is detected, that entire row of R_0 is set to zero. When r diagonal entries of the n column matrix R_0 are left un-zeroed, the $(r+1)^{th}$ singular value, $\sigma_{r+1}(X)$ of the $m \times n$ data matrix X is then bounded by:

$$\sigma_{r+1}(X) \leq \alpha\sqrt{n-r} \tag{1}$$

Hence, for small enough α the rank of X is revealed to be at most r. The subspace angle[3] θ_X between the signal subspace estimated using Row-Zeroing QR factorisation upon X and that obtained from the SVD is bounded by[7]:

$$\left|\sin(\theta_X)\right| \leq \frac{\sigma_{r+1}(X)}{\sigma_r(R_{X,S})} \tag{2}$$

in which $R_{X,S}$ is obtained from R_0 by deleting both rows and columns which contain a zeroed diagonal element.

Where it arises the occurrence of a small diagonal element on the leading diagonal element of R is a reliable indicator of rank-deficiency. Moreover, Row-Zeroing then overcomes the corruption of the remainder of R which arises with conventional QR factorisation. This corruption would normally prevent the degree of rank deficiency from being determined. There is however, no guarantee of any small diagonal element in R when X is rank-deficient. Consider, for example, the $n \times n$ upper triangular matrix K_n due to Kahan[5]:

$$K_n = \mathrm{diag}(1, s, s^2, \ldots, s^{n-1}) \begin{bmatrix} 1 & -c & -c & \ldots & -c \\ & 1 & -c & \ldots & -c \\ & & 1 & \ldots & -c \\ & 0 & & & : \\ & & & & 1 \end{bmatrix} \qquad (3)$$

where, $c^2 + s^2 = 1$. Chan[1] for example, uses the value $n = 50$ and $c = 0.2$. For this choice the diagonal values of K_n decrease smoothly from 1.0 to 0.368 and the singular values from σ_1 to σ_{49} decay similarly smoothly from 4.635 to 0.411. The value of σ_{50}, however, is 9.287×10^{-5}. The matrix is therefore rank-deficient but has no small entry on its leading diagonal. For this reason, the Row-Zeroing approach will clearly fail for Kahan's matrix.

In conventional QR-based least-squares minimisation we write the $m \times n$ data matrix X as $X = QR$. The least squares weight vector $\underline{\omega}$, i.e. the solution to $\min_{\underline{\omega}} E\{|X\underline{\omega} + \underline{y}|^2\}$ is obtained from:

$$R\underline{\omega} = -Q^H \underline{y} \qquad (4)$$

by backsubstitution. Now, it is not difficult to show that the subspace projection of the least-squares weight vector $\underline{\omega}$ is equivalent to the weight vector obtained from the least-squares problem with the projected X matrix and $\underline{y}$. Thus, in OSPRE, Row-Zeroing is applied to X in order to replace R in (4) with the Row-Zeroed R matrix R_0. This matrix is an approximation to the "R" matrix in the QR factorisation of the projection of X onto the signal subspace. The resulting equation for $\underline{\omega}$ can then be written:

$$\begin{bmatrix} R_{11} & R_{12} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \underline{\omega}_1 \\ \underline{\omega}_2 \end{bmatrix} = -\begin{bmatrix} \underline{u}_1 \\ 0 \end{bmatrix} \qquad (5)$$

For simplicity, the first r rows of R are shown here as non-zero (and $\underline{u}_1$ as the first r elements of $Q^H \underline{y}$) but it in general this could be any set of r rows. Notice that $\underline{\omega}_2$ is left undetermined. One solution for $\underline{\omega}$ is clearly:

$$\underline{\omega}_2 = 0 \qquad \text{and} \qquad R_{11}\underline{\omega}_1 = -\underline{u}_1 \qquad (6)$$

where $\underline{\omega}_1$ is obtainable simply by back-substitution. The general solution, however, is:

$$\underline{\omega}_1 = \underline{\omega}_{1,0} - R_{11}^{-1} R_{12} \underline{\omega}_2 \qquad (7)$$

where $\underline{\omega}_{1,0}$ is the solution for $\underline{\omega}_1$ from (6). There is therefore an element of choice here for $\underline{\omega}_2$. One choice is to select the elements of $\underline{\omega}_2$ in such a way as to minimise the norm of $\underline{\omega}$. This solution can be written:

$$\underline{\omega}_2 = (F^H F + I)^{-1} F^H \underline{\omega}_{1,0} \qquad (8)$$

where $F = R_{11}^{-1} R_{12}$. Now whilst $\underline{\omega}_{1,0}$ can be obtained simply from the OSPRE algorithm, in general, it does not result in a beam pattern with low sidelobes. To achieve this the minimum norm solution (equation (8)) can be used which entails extra computation. In what follows we show that this difficulty is overcome by applying the Row-Zeroing technique to the covariance matrix or powers thereof. Before introducing this approach, however, we discuss the numerical linear algebra procedure known as Orthogonal Iteration.

3. ORTHOGONAL ITERATION

Orthogonal Iteration[3] is a technique for estimating the "dominant invariant subspace" $D_q(A)$ of a square complex $n \times n$ matrix A i.e. the subspace spanned by the leading q eigenvectors $\underline{u}_i$ of A associated with the q largest eigenvalues λ_i. The technique proceeds as follows: given a matrix A and some initial $n \times q$ matrix Q_0 with orthogonal columns:

$$\begin{aligned} &\text{FOR } k=1,2,\ldots \\ &\quad Z_k = XQ_{k-1} \\ &\quad QR_k = Z_k \qquad \text{(QR decomposition)} \\ &\quad Q_k = [Q]_{1:q} \\ &\text{END} \end{aligned} \qquad (9)$$

where the last step means select the first q columns of the matrix Q. For $q = n$ and at iteration k Orthogonal Iteration generates square matrices Q_k such that T_k given by:

$$T_k = Q_k^H A Q_k \qquad (10)$$

converges to upper triangular form with increasing k. Hence, Orthogonal Iteration converges to a Schur decomposition of X (provided, theoretically, that Q_0 is not deficient in any of the q eigendirections). That is to say that:

$$\lim_{k \to \infty} T_k = \mathrm{diag}\{\lambda_i : i=1, 2, \ldots n\} + \bar{R} \qquad (11)$$

where $\bar{R}$ is strictly upper triangular. Note, however, that the eigenvalues λ_i do not necessarily arise in decreasing magnitude order.

A general matrix A is said to be "normal" if:

$$A^H A = A A^H \qquad (12)$$

Furthermore, it is readily shown that an upper triangular matrix which is normal has to be diagonal. Therefore, at convergence and assuming that A is a normal matrix, $\bar{R} = 0$ and

$$Q_k^H A Q_k = \mathrm{diag}\{\lambda_i : i=1, 2, \ldots n\} \qquad (13)$$

The columns of Q_k will then be the (right) eigenvectors $\underline{u}_i$ (in the appropriate order).

Finally note that it is easy to show from equation (9) that with $Q_0 = I$ (and $p = n$) we have:

$$A^k = Q_k R_k R_{k-1} \ldots R_1 = Q_k R_k' \qquad (14)$$

where the last step defines R_k' and follows because the product of triangular matrices is also triangular. Hence, by virtue of its uniqueness, the QR decomposition of the matrix A^k will produce

the orthogonal matrix Q_k that would appear at iteration k of the Orthogonal Iteration procedure. Thus the QR decomposition of increasing powers of a normal matrix results in a sequence of orthogonal matrices that converge to the orthogonal matrix found in the EVD of the original matrix.

From equations (13) and (14), it is easy to show that[1]

$$R_k' = Q_k^H A^k \approx \mathrm{diag}\{\lambda_i^k : i=1, 2, \ldots n\} Q_k^H \qquad (15)$$

for sufficiently large k. As the rows of the matrix Q_k^H are of unit norm, it can be seen that the norm of the rows of R_k' converge to the k^{th} power of the eigenvalues λ_i. Given these properties of a normal matrix and the QR decomposition of its powers, we are now in a position to formulate the new algorithm.

4. COVARIANCE POWER PROJECTION

Let the covariance matrix of the data be $M = X^H X$ and for some power p write $M^p = Q_M R_M$. Then the normal equations give:

$$R_M \underline{\omega} = -Q_M^H M^{p-1} X^H \underline{y} \qquad (16)$$

and we can again obtain $\underline{\omega}$ by backsubstitution in the usual way. The data matrix X is not generally normal but the covariance matrix $M = X^H X$, and any power of M is normal since M is Hermitian. Thus, using the results of the previous section, the QR decomposition of a suitably large power of the covariance matrix will produce an orthogonal matrix that is approximately the matrix of eigenvectors of M (i.e. right singular vectors of X). Furthermore, the norm of each row of the triangular factor will be the k^{th} power of the eigenvalues. Thus if we proceed with the Row-Zeroing technique of OSPRE, but with thresholding based upon the row norm instead of the diagonal element, we can determine the signal subspace and project the least-squares weight vector onto it. The important point here is that the process is now guaranteed to be rank-revealing in the limit $k \to \infty$ since the eigenvalues of the covariance matrix are the squares of the singular values of the data matrix and only the latter are reliable indicators of rank. This provides encouragement that in practice R_M will tend to be rank-revealing and give good subspace estimates for some finite, and possibly small, k.

Now, it has been observed experimentally that the diagonal element of the triangular factor in the QR decomposition is often one of the largest elements in the row. Thus we can expect that the diagonal elements of the triangular factor will be indicative of the norm of that row i.e. λ_i^k. Thus it may be possible to perform the thresholding based only on the value of the diagonal elements. Indeed, if we use the Row-Zeroing technique upon M^p to give $R_{M,0}$, say, it is easy to see that the analytic error bound upon the subspace angle corresponding to (2) is now given by:

$$\left|\sin(\theta_M)\right| \le \frac{\sigma_{r+1}(M^p)}{\sigma_r(R_{M^p,S})} \qquad (17)$$

This is a tighter bound than that upon θ_X since, $\sigma_{r+1}(M^p)$ is $(\sigma_{r+1}(X))^{2p}$ and, in practice, one finds $\sigma_r(R_{M^p,S}) \sim \sigma_r(R_S)^{2p}$. Hence the subspace estimate is correspondingly more accurate.

1. This may look odd at first sight but remember that it is only true for $k \to \infty$. In general the right hand matrix is dense but with small elements below the diagonal if k is sufficiently large.

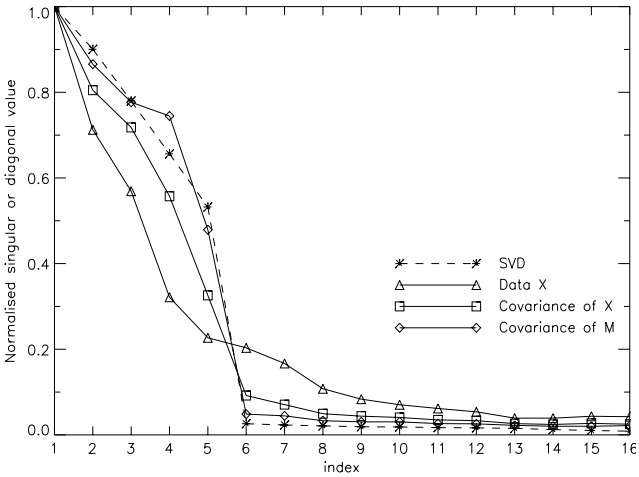

Figure 1. Diagonal Value Spectrum

In addition, the diagonal value spectrum of the R matrix is also sharpened - making thresholding easier. To see how this arises consider the simple case:

$$X = \left[\underline{x}_1\ \underline{x}_2\right]; \quad \underline{x}_1 = \lambda_1 \underline{e}_1, \quad \underline{x}_2 = \lambda_2 \underline{e}_1 + \varepsilon \underline{e}_2 \qquad (18)$$

where, $\underline{e}_1$ and $\underline{e}_2$ are unit vectors and λ_1, λ_2 and ε are scalars. For this case:

$$R = \begin{bmatrix} \lambda_1 & \lambda_2 \\ 0 & \varepsilon \end{bmatrix} \qquad (19)$$

whereas for the covariance matrix:

$$R_M = \begin{bmatrix} \lambda_1\sqrt{(\lambda_1^2+\lambda_2^2)} & \lambda_2\sqrt{(\lambda_1^2+\lambda_2^2)} \\ 0 & \varepsilon^2 \end{bmatrix} \qquad (20)$$

Hence, for $\lambda_1 \approx \lambda_2$ the ratio of the first to the second diagonal element in (20) is the square of that for (19) i.e. the gap between the diagonal elements is widened. A similar argument holds for successive powers of the covariance matrix.[2]

The effectiveness of this approach is illustrated in figure 1. The spectrum of diagonal values for the different techniques is compared with the singular value spectrum for a particular case of 5 signals of 20dB power with respect to noise. For a fair comparison the values in the diagonal value spectra obtained for M and for M^2 are raised to the powers $1/2$ and $1/4$ respectively. All the spectra were then normalised by their largest values. The ordinary QR diagonal value spectrum of the data shows no discernable gap at all. Using the covariance sharpens the diagonal spectrum as predicted. A further (albeit smaller) improvement in the diagonal value spectrum is then obtained using $M^H M$ in place of M.

Finally note that, as in section 2, the projected least-squares weight vector produced by the above method (OSPRE with M^p as input)

2. These results are not due simply to a "squaring" of the data matrix. Whilst the net information content of the matrix is the same, columns of the covariance contain information from all sensors and not just from one sensor alone as in the data.

is not the minimum norm solution. However, the Q matrix from the decomposition contains estimates of the right singular vectors of the data matrix X. Thus, a projection operator can be constructed that is an approximation to that which would result from having calculated the SVD of X. This is used to approximate the minimum norm solution by projecting the weight produced from OSPRE. This process constitutes the proposed "Covariance Power Projection Algorithm". The algorithm is clearly not suited to implementation in a pipelined form. However, it is possible to form the covariance matrix from the triangular factor R of the data matrix since $X^H X = R^H Q^H Q R = R^H R$. Thus, a practical implementation can consist of a recursive updating of R (standard QRD-RLS[6]) and then, whenever subspace projection is required, the R matrix can be used, off-line, to form the covariance matrix (which involves fewer operations than forming $X^H X$) or a power thereof, to use as the input to OSPRE.

5. RESOLVING THE PROBLEM CASE

Whereas a small diagonal element on the leading diagonal of the "R" matrix in a QR factorisation does indicate rank-deficiency *when it arises*, there is *no guarantee of this arising* when a matrix is rank-deficient. This can now be seen as a consequence of the difference in general between the eigenvalues and singular values of a matrix. The reason that Kahan's matrix is difficult to handle is because its eigenvalues are different from its singular values and in particular because the former reveal no gap. Secondly, an additional difficulty with K_n for algorithms based on QR decomposition is the upper triangular structure itself - K_n is its own triangular factor. Upset this structure by a simple column interchange and the rank-deficiency soon becomes apparent by one or two steps of Orthogonal Iteration. Such an operation changes the row subspace and the eigenvalues of the matrix but leaves the column subspace and the singular values unaltered. Note however that the covariance of K_n *is* normal. Taking the covariance is an extreme example of an operation which renders the result non-upper-triangular but which leaves the singular value structure fundamentally unaltered (apart from a squaring, the values are unchanged).

6. SIMULATIONS

In figure 2 a number of beampatterns are plotted for the case of two 30dB sources at $-60°$ and $+30°$. In each case the beamformer is in the form of a generalised sidelobe canceller[4] with a look direction of $0°$. The quiescent weight vector was designed to give a -30dB Chebyshev weighted pattern. The upper curve was calculated using ordinary QR factorisation to determine the unconstrained contribution to the weight vector. The lower (solid) curve gives the corresponding result when the unconstrained weight vector is projected onto the signal subspace obtained with the SVD. The sidelobe level is much reduced (and there is less variation form batch to batch i.e. less "jitter"). The middle curve is the result obtained using the Covariance Power Projection algorithm with the covariance matrix but without the final projection to produce the minimum norm solution. This curve has lower sidelobes than ordinary QR but does not do as well as the SVD (although it is found to greatly reduce the jitter). However, a fourth curve is present on the same axes although it does not show clearly because it is so close to the SVD result. This curve is the approximate minimum norm solution obtained by using the full Covariance Power Projection algorithm.

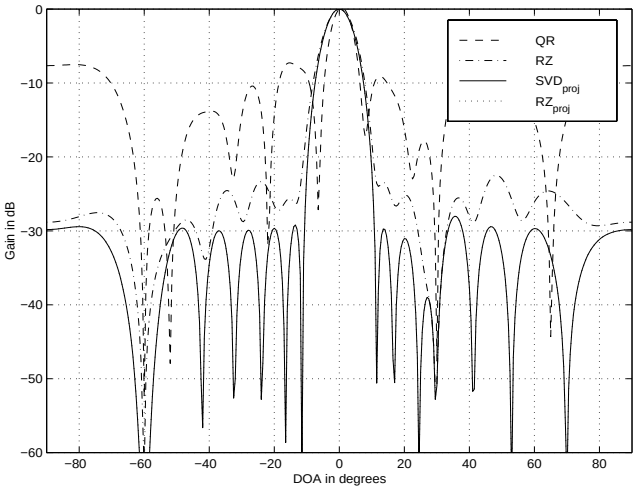

Figure 2. Adapted Beampatterns

7. CONCLUSIONS

A QR based technique is presented which obtains a good approximation to the minimum norm solution of the projection of the least squares weight onto the signal subspace of the data matrix, without performing an SVD. Theoretical difficulties associated with conventional QR factorisation of the data are overcome by applying the technique of Row-Zeroing QR to the covariance matrix. Thresholding is simpler, the subspace estimate is improved and the signal subspace estimate and least-squares weight are obtained simultaneously without matrix inversion**.**

8. REFERENCES

[1] Chan T.F., "Rank revealing QR factorisations", *Lin. Alg. and its App.*, Vol. 88, No. 89, pp 67-82, 1987.

[2] Foster L.V., "The probability of large diagonal elements in the QR factorisation", *SIAM J. Sci. Stat. Comput.*, Vol. 11, No. 3, pp531-544, 1990.

[3] Golub G.H. and Van Loan C.F., "*Matrix Computations*" (Second Edition), John Hopkins University Press, Baltimore, USA, 1989.

[4] Griffiths L.J. and Jim C.W., "An Alternative Approach to Linearly Constrained Adaptive Beamforming", *IEEE trans. Antennas and Propagation*, Vol. 30, pp27-34, 1982.

[5] Kahan W., "Numerical Linear Algebra", *Canadian Math. Bull.*, Vol. 9, No.6, pp757-8011, 1966

[6] McWhirter J.G., "Recursive Least Squares Minimisation using a Systolic Array", *Proc. SPIE: Real-Time Signal Processing IV*, Vol. 431, pp 105-112, 1983

[7] Smith M.J. and Proudler I.K., "A one-sided algorithm for subspace projection beamforming", *Proc. SPIE: Advanced Signal Processing Algorithms, Architectures and Implementations VI*, pp100-111, Denver, Colorado, 4th -9th August 1996.

[8] Smith M.J. and Proudler I.K., "A recursive one-sided algorithm for subspace projection beamforming", *Proc. SPIE: Advanced Signal Processing Algorithms, Architectures and Implementations VII*, San Diego, California, 28th -30th July 1997.