ON-LINE SUBSPACE ESTIMATION USING A SCHUR-TYPE METHOD

Alle-Jan van der Veen and Jürgen Götze

Stanford University
Dept. Computer Science/SCCM
Stanford, CA 94306-2140

Inst. Network Th. and Circuit Design
Technical University of Munich
Arcisstr. 21, D-80333 Munich, Germany

Abstract

A new method is presented for estimating the column space (signal subspace) of a low rank data matrix distorted by additive noise. It is based on a tangible expression for the set of all matrices of minimal rank that are ε-close to the data matrix in matrix 2-norm. The usual truncated SVD approximant is contained in this set. Features of the algorithm are (1) it has the same computational structure and complexity as a QR factorization of the data matrix, (2) it yields an on-line scheme, amenable to parallel (systolic) implementation, (3) updating and downdating is straightforward, (4) a rank decision (to detect the number of signals) is automatic. It is shown in simulations on a typical direction finding application that the algorithm exhibits similar performance as SVD-based methods, at a fraction of the computational cost.

1. Introduction

Many parameter estimation algorithms in signal processing applications involve a data matrix which is presumably of low rank, but which is distorted by noise. One well-known example occurs in direction finding, where the data model is $X = AS + N$, and it is desired to estimate the rank of AS and the column span $\mathcal{R}(A)$ of the array response matrix (see *e.g.,* [1, 2] for overviews). Another example is in adaptive filtering, where the model consists of an overdetermined set of equations $Xw = d$ (X and d known), and it is desired to find the weight vector w of a transversal filter (see *e.g.,* [3]). A total least squares solution is obtained by approximating $[X \;\; d]$ by a matrix of low rank, and taking $[w^H \;\; -1]^H$ to be a vector in its kernel of minimal norm. Again, it is important to find an estimate of a column space, in this case of the range of $[X \;\; d]^H$. In both applications, the estimate of the principal column span is to be updated continuously, as more data samples are measured.

In principle, the singular value decomposition (SVD) is the appropriate tool for estimating principal subspaces. However, the SVD is computationally expensive to update, rendering it unattractive for on-line applications. A number of alternatives have been developed to replace the SVD by computationally more attractive decompositions. Examples are the URV decomposition [4, 5] and the Rank Revealing QR decomposition [6, 7].

In this paper, we consider, for a given matrix X and tolerance ε, approximants $\hat{X}$ that satisfy

$$\| X - \hat{X} \| \leq \varepsilon$$

and that are of smallest possible rank. $\| \cdot \|$ denotes the matrix 2-norm (spectral norm, or largest singular value of its argument). The truncated SVD solution, obtained by putting all singular values of the SVD of X that are smaller than ε equal to zero, satisfies the conditions and even minimizes $\| X - \hat{X} \|$. However, because we do not search for the minimum of $\| X - \hat{X} \|$ and because the norm is not the Frobenius norm, $\hat{X}$ is not unique; there is an expression which gives all suitable approximants $\hat{X}$ of the same rank as the truncated SVD approximant. In particular, there is a relatively simple expression for the column space of $\hat{X}$. This subspace may be computed using a Schur-type algorithm, or Hyperbolic QR-factorization, with complexity of order $O(1/2\, m^2 n)$ for a matrix X of size $m \times n$ ($m < n$).

The approximation theory referred to in the above is presented and proven in a separate paper [8]. It is based in turn on a specialization of the Hankel-norm approximation theory for time-varying systems, by Dewilde and Van der Veen [9].

2. Minimal rank 2-norm approximations

For a given $m \times n$ data matrix X, denote the SVD of X as

$$X = [U_1 \quad U_2] \begin{bmatrix} \Sigma_1 & \\ & \Sigma_2 \end{bmatrix} \begin{bmatrix} V_1^H \\ V_2^H \end{bmatrix}$$
$$(\Sigma_1)_{ii} > \varepsilon\,, \qquad (\Sigma_2)_{ii} \leq \varepsilon\,.$$

Suppose that d singular values of X are larger than ε, and none are equal to ε. Our approximation theory is based on an implicit factorization of

$$XX^H - \varepsilon^2 I \;=\; BB^H - AA^H\,. \tag{1}$$

This is a Cholesky factorization of an indefinite Hermitian matrix. A and B are chosen to have full column rank. They

are not unique, but their dimensions are well-defined. Using the SVD of X, we obtain one possible decomposition as

$$XX^H - \varepsilon^2 I = U_1(\Sigma_1^2 - \varepsilon^2 I)U_1^H + U_2(\Sigma_2^2 - \varepsilon^2 I)U_2^H ,$$

where the first term is positive semidefinite and has rank d, and the second term is negative semidefinite and has rank $m-d$. Hence, B has d columns, and A has $m-d$ columns.

To obtain an implicit factorization that avoids computing XX^H, we make use of the properties of J-unitary matrices. A 2×2 block matrix Θ is said to be J-unitary, with

$$\Theta = \begin{array}{c} \\ m \\ n \end{array} \begin{array}{c} \begin{array}{cc} m & n \end{array} \\ \begin{bmatrix} \Theta_{11} & \Theta_{12} \\ \Theta_{21} & \Theta_{22} \end{bmatrix} \end{array}, \quad J = \begin{bmatrix} I_{m\times m} & \\ & -I_{n\times n} \end{bmatrix}$$

if $\Theta^H J\Theta = J$, $\Theta J\Theta^H = J$. Two of the properties of J-unitary matrices which we will use are

1. If $[C \;\; D] = [A \;\; B]\Theta$ then

$$\begin{aligned} AA^H - BB^H &= [A \;\; B]J[A \;\; B]^H \\ &= [A \;\; B]\Theta J\Theta^H[A \;\; B]^H \\ &= CC^H - DD^H . \end{aligned}$$

2. $\Theta_{12}^H\Theta_{12} - \Theta_{22}^H\Theta_{22} = -I$
$\Rightarrow\; I - \Theta_{22}^{-H}\Theta_{12}^H\Theta_{12}\Theta_{22}^{-1} \;=\; \Theta_{22}^{-H}\Theta_{22}^{-1} \;>\; 0$
$\Rightarrow\; \| \Theta_{12}\Theta_{22}^{-1} \| \;<\; 1.$

Theorem 1. *Let X have d singular values larger than ε, and none equal to ε. Then there exists a J-unitary matrix Θ such that*

$$[\varepsilon I \quad X]\,\Theta \;=\; [A' \quad B'] \tag{2}$$

where $A' = [A \;\; 0_{m\times d}]$, $B' = [B \;\; 0_{m\times n-d}]$, A: $m\times(m-d)$, B: $d\times(n-d)$, and $[A \;\; B]$ has full rank.

Note that, by the first mentioned property of J-unitary matrices, equation (2) implies (1). Using the second property, we obtain the following theorem.

Theorem 2. *$\hat{X} = B'\Theta_{22}^{-1}$ is a rank-d 2-norm approximant.*

Proof (outline). $\hat{X}$ has rank d because $B' = [B \;\; 0]$ has rank d. Equation (2) yields

$$\begin{aligned} & \varepsilon\Theta_{12} + X\Theta_{22} = B' \\ \Rightarrow \quad & \| X - \hat{X} \| = \varepsilon \| \Theta_{12}\Theta_{22}^{-1} \| < \varepsilon . \end{aligned}$$

□

The 'signal subspace' $\mathcal{R}(\hat{X}) = \mathcal{R}(B)$ is obtained directly from the factorization (2). The factorization is related to a factorization

$$[\varepsilon I \quad X]\,\Theta\Pi \;=\; [(A,B) \quad 0], \tag{3}$$

where Π is a permutation matrix and (A,B) is of full rank and contains the columns of A and B, possibly in permuted order. Equation (3) can be viewed as a 'hyperbolic QR-factorization' (it reduces to a QR factorization if $\varepsilon = 0$). The computation of this factorization using elementary (hyperbolic) rotations is known as the Generalized Schur algorithm in systems theory, and is considered in section 3.

The set of all minimal-rank 2-norm approximants will be parametrized by matrices S_L, with block partitioning

$$S_L = \begin{array}{c} \\ m-d \\ d \end{array} \begin{array}{c} \begin{array}{cc} d & n-d \end{array} \\ \begin{bmatrix} (S_L)_{11} & (S_L)_{12} \\ (S_L)_{21} & (S_L)_{22} \end{bmatrix} \end{array}.$$

Theorem 3. *All rank d 2-norm approximants $\hat{X}$ of X are given by*

$$\hat{X} \;=\; (B' - A'S_L)(\Theta_{22} - \Theta_{21}S_L)^{-1} ,$$

where S_L satisfies $(i) : \| S_L \| \leq 1$, and $(ii) : (S_L)_{12} = 0$.
The approximation error is

$$X - \hat{X} \;=\; \varepsilon(\Theta_{11}S_L - \Theta_{12})(\Theta_{22} - \Theta_{21}S_L)^{-1} .$$

By this theorem, the estimated signal subspace is given by $\mathcal{R}(\hat{X}) = \mathcal{R}(B' - A'S_L) = \mathcal{R}(B - A(S_L)_{11})$. The first condition on S_L ensures that $\| X - \hat{X} \| < \varepsilon$, whereas the second condition is required to have $\hat{X}$ of rank d.

The choice of a particular approximant $\hat{X}$, or subspace estimate $\mathcal{R}(\hat{X})$, boils down to a suitable choice of the parameter S_L. Various choices are interesting:

1. The approximant $\hat{X}$ in theorem 2 is obtained by taking $S_L = 0$. This is the simplest approximant; the signal subspace estimate is equal to the range of B. The error is given by $\varepsilon\| \Theta_{12}\Theta_{22}^{-1} \|$. Note that, even if all singular values of X are larger than ε so that it is possible to have $\hat{X} = X$, the choice $S_L = 0$ does not give zero error. Hence, this simple choice of S_L could lead to biased estimates. This is confirmed in the simulation example in section 4, and occurs in cases where σ_d is close to ε.

2. As the truncated SVD solution satisfies the requirements, there is an S_L which yields this particular solution and minimizes the approximation error. However, computing this S_L requires an SVD itself.

3. If we take $S_L = \Theta_{11}^{-1}\Theta_{12}$, then we obtain $\hat{X} = X$ and the approximation error is zero. Although this S_L is contractive, it does not satisfy the condition $(S_L)_{12} = 0$. To satisfy both conditions, we propose to take

$$S_L = \Theta_{11}^{-1}\Theta_{12}\begin{bmatrix} I & \\ & 0 \end{bmatrix} = \begin{bmatrix} (\Theta_{11}^{-1}\Theta_{12})_{11} & 0 \\ (\Theta_{11}^{-1}\Theta_{12})_{21} & 0 \end{bmatrix} .$$

The corresponding signal subspace estimate is given by the range of $B' - A'\Theta_{11}^{-1}\Theta_{12}\left[\begin{smallmatrix} I & 0 \\ 0 & 0 \end{smallmatrix}\right]$, and can be computed by a Schur complement formula. If X is square, then this S_L gives an approximation error $X - \hat{X}$ which has rank $m - d$, *i.e.,* the same as a truncated SVD solution would give.

4. If $d \geq m/2$, then it is possible to take S_L such that $S_L S_L^H = I_m$, $(S_L)_{12} = 0$ (*e.g.,* $S_L = [I \;\; 0]$), which gives a uniform approximation error: all singular values of $X - \hat{X}$ are equal to ε.

3. Computation of Θ

Elementary rotations

If $[\varepsilon I \;\; X]$ satisfies certain regularity conditions, then it is possible to compute Θ such that (3) holds, using elementary plain rotations only. At an elementary level, we are looking for 2×2 matrices $\tilde{\theta}$ such that $[a \;\; b]\tilde{\theta} = [* \;\; 0]$. The matrices $\tilde{\theta}$ are J-unitary, but with respect to *unsorted* signature matrices $\tilde{j} = \mathrm{diag}[\pm1]$:

$$\tilde{\theta}\tilde{j}_2\tilde{\theta}^H = \tilde{j}_1\,, \quad \tilde{\theta}^H\tilde{j}_1\tilde{\theta} = \tilde{j}_2\,.$$

By congruence, the number of positive entries in $\tilde{j}_1$ is equal to the number of positive entries in $\tilde{j}_2$, and similarly for the negative entries. Hence, it is sufficient to consider the following six cases (where $s^H s + c^H c = 1$):

1. if $|a| > |b|$ and $\tilde{j}_1 = \mathrm{diag}[1 \;\; -1]$:
$$\begin{matrix} [& \overset{+}{a} & \overset{-}{b} &] \end{matrix}\begin{bmatrix} 1 & -s \\ -s^* & 1 \end{bmatrix}\frac{1}{c^*} = \begin{matrix} [& \overset{+}{*} & \overset{-}{0} &] \end{matrix},$$
2. if $|a| < |b|$ and $\tilde{j}_1 = \mathrm{diag}[1 \;\; -1]$:
$$\begin{matrix} [& \overset{+}{a} & \overset{-}{b} &] \end{matrix}\begin{bmatrix} -s & 1 \\ 1 & -s^* \end{bmatrix}\frac{1}{c^*} = \begin{matrix} [& \overset{-}{*} & \overset{+}{0} &] \end{matrix},$$
3. if $|a| < |b|$ and $\tilde{j}_1 = \mathrm{diag}[-1 \;\; 1]$:
$$\begin{matrix} [& \overset{-}{a} & \overset{+}{b} &] \end{matrix}\begin{bmatrix} -s & 1 \\ 1 & -s^* \end{bmatrix}\frac{1}{c^*} = \begin{matrix} [& \overset{+}{*} & \overset{-}{0} &] \end{matrix},$$
4. if $|a| > |b|$ and $\tilde{j}_1 = \mathrm{diag}[-1 \;\; 1]$:
$$\begin{matrix} [& \overset{-}{a} & \overset{+}{b} &] \end{matrix}\begin{bmatrix} 1 & -s \\ -s^* & 1 \end{bmatrix}\frac{1}{c^*} = \begin{matrix} [& \overset{-}{*} & \overset{+}{0} &] \end{matrix},$$
5. $$\begin{matrix} [& \overset{+}{a} & \overset{+}{b} &] \end{matrix}\begin{bmatrix} c & s \\ -s^* & c^* \end{bmatrix} = \begin{matrix} [& \overset{+}{*} & \overset{+}{0} &] \end{matrix},$$
6. $$\begin{matrix} [& \overset{-}{a} & \overset{-}{b} &] \end{matrix}\begin{bmatrix} c & s \\ -s^* & c^* \end{bmatrix} = \begin{matrix} [& \overset{-}{*} & \overset{-}{0} &] \end{matrix}.$$

In the above equations, the signature matrix $\tilde{j}_1$ associates a signature to a and b; the signature $\tilde{j}_2$ assigns a signature to the result $[* \;\; 0]$; where $(\tilde{j}_2)_{11}$ is in fact equal to the sign of $[a \;\; b]\tilde{j}_1[a \;\; b]^H$.

Hyperbolic QR factorization

The elementary rotations are embedded in plane rotations which are applied to the columns of $[\varepsilon I \;\; X]$ in the same way as Givens rotations are used for computing a QR factorization. Each elementary rotation produces a zero entry in X. The difference with QR is that we have to keep track of the signatures associated to the columns of the matrix. The general scheme, however, goes as follows:

$$\begin{matrix} + & + & + & & - & - & - & - \end{matrix}$$
$$\left[\begin{array}{ccc|cccc} \varepsilon & & & \times & \times & \times & \times \\ & \varepsilon & & \times & \times & \times & \times \\ & & \varepsilon & \times & \times & \times & \times \end{array}\right] \quad \tilde{\Theta}_{(1,1)} \rightarrow$$
$$\begin{matrix} - & + & + & & + & - & - & - \end{matrix}$$
$$\left[\begin{array}{ccc|cccc} \times & & & 0 & \times & \times & \times \\ \times & \varepsilon & & \times & \times & \times & \times \\ \times & & \varepsilon & \times & \times & \times & \times \end{array}\right] \quad \tilde{\Theta}_{(2,1)} \rightarrow$$
$$\begin{matrix} - & + & + & & + & - & - & - \end{matrix}$$
$$\left[\begin{array}{ccc|cccc} \times & & & 0 & \times & \times & \times \\ \times & \times & & 0 & \times & \times & \times \\ \times & \times & \varepsilon & \times & \times & \times & \times \end{array}\right] \quad \rightarrow$$
$$\begin{matrix} & & & - & + & - & & + & + & - & - \end{matrix}$$
$$\cdots \quad \tilde{\Theta}_{(m,n)} \rightarrow \quad \left[\begin{array}{ccc|cccc} \times & & & 0 & 0 & 0 & 0 \\ \times & \times & & 0 & 0 & 0 & 0 \\ \times & \times & \times & 0 & 0 & 0 & 0 \end{array}\right].$$

(Except for the first matrix, the signatures of the columns in the above matrices are examples, as they are data dependent.) This scheme ensures that $[\varepsilon I \;\; X]\tilde{\Theta} = [\tilde{X} \;\; 0]$, where $\tilde{X}$ is a resulting lower triangular invertible matrix; it contains the columns of A and B in some permuted order. The columns with a positive signature are the columns of A, the columns with a negative signature are those of B.

Updating and downdating

The Schur method is straightforward to update as more and more columns of X are measured. If $[\varepsilon I \;\; X_n]\tilde{\Theta} = [(A_n, B_n) \;\; 0]$ is the factorization at point n and $X_{n+1} = [X_n \;\; x_{n+1}]$, then, because the algorithm works column-wise,

$$[(A_n, B_n) \;\; 0 \;\; x_{n+1}]\theta^{(n+1)} = [(A_{n+1}, B_{n+1}) \;\; 0 \;\; 0]\,,$$

for some J-unitary matrix $\theta^{(n+1)}$ acting on the columns of A_n, B_n, and on x_{n+1}. Hence, we can continue with the result of the factorization that was obtained at the previous step.

The downdating problem is to compute the factorization for X_n with its first column x_1 removed, from a factorization of X_n. It can be converted to an updating problem, where the old column x_1 is now introduced with a positive signature,

$$[(\overset{+}{A_n}, \overset{-}{B_n}) \quad \overset{+}{x_1}]\theta^{(n+1)} = [(A_{n+1}, B_{n+1}) \;\; 0]\,.$$

This is possible because, implicitly, we factor $\varepsilon^2 I - X_n X_n^H + x_1 x_1^H = A_n A_n^H - B_n B_n^H + x_1 x_1^H$.

4. Application to Direction Finding

In order to assess the applicability of the Schur-based subspace estimation technique, we consider the direction finding problem. Suppose that we have an array of m equispaced omnidirectional sensors, which receives d sinusoidal signals from directions ϕ_k, $k = 1, \cdots d$. A total number of n samples is taken, which gives an $m \times n$ data matrix X which satisfies the model $X = \mathcal{A}S + N$. Here, $\mathcal{A} = \mathcal{A}(\phi_1, \cdots, \phi_d) : m \times d$ is the array response matrix, and $S : d \times n$ contains the n samples of the d source signals. N contains samples of white additive i.i.d. noise sources with variance $\sigma^2 I$, independent of the signals. Given X, the ϕ_k are to be estimated.

The ESPRIT algorithm for estimating the DOAs [10] (in one of its incarnations) works in two steps. The first step is to estimate the signal subspace, which is usually taken to be the d principal left singular vectors of X, $U_S = U_{SF} = U_1$. This leads to the classical SVD-ESPRIT direction finding scheme. We will compare this with the Schur-based subspace estimates, and investigate the choice $U_S = U_{SS1} = \mathcal{R}(B - A(S_L)_{11})$ with $S_L = \Theta_{11}^{-1}\Theta_{12}\left[\begin{smallmatrix} I & 0 \\ 0 & 0 \end{smallmatrix}\right]$, and $U_S = U_{SS2} = \mathcal{R}(B)$. As is well-known, once the signal subspaces are estimated, the DOAs are obtained via a certain eigenvalue decomposition based on these subspaces.

In the computer simulation experiments, a linear array consisting of $m = 4$ sensors is used. Two sources are impinging on the array. The signal to noise ratio is chosen to be 20dB in all cases. One hundred test runs using $n = 30$ samples are executed. Table 1 lists the statistical results, for three different sets of angles of incidence, and avaraged over the test runs. If the signals are spatially well separated, the difference between the three subspace estimates is negligible. If the signals are moving closer, the variance of the Schur estimate with $S_L = 0$ starts to increase, but the choice $S_L = \Theta_{11}^{-1}\Theta_{12}\left[\begin{smallmatrix} I & 0 \\ 0 & 0 \end{smallmatrix}\right]$ still performs the same as the SVD-based estimate. Finally, part (*c*) of the simulations shows that if the signals are so close that the variance clouds are overlapping even for the SVD-based estimates, then the choice $S_L = 0$ breaks down, but the variance of Schur method 1 is still within reasonable bounds.

5. Conclusions

In this paper, we have applied a new subspace estimation method to the direction finding problem. The algorithm is relatively simple and straightforward to implement, as it has the same structure as a QR factorization. The estimated signal subspace can be tracked adaptively by updating and downdating. For the simplest estimate ($S_L = 0$), only $m^2/2$ elementary rotations are required for each new sample vector of dimension m. It is assumed that a suitable value for the noise threshold, ε, is known on the outset. The precise value of ε is not critical, as long as it lies in the gap between the signal and noise singular values

Table 1. Estimated DOAs for the ESPRIT algorithm.

(*a*)

$\phi_{1,2} = 10^\circ, 70^\circ$	SVD	Schur 1	Schur 2
DOAmean	9.9948	9.9947	10.0138
	70.0160	70.0160	69.9601
DOAstd	0.0122	0.0122	0.0124
	0.1243	0.1242	0.1283

(*b*)

$\phi_{1,2} = 20^\circ, 30^\circ$	SVD	Schur 1	Schur 2
DOAmean	19.9447	19.9417	21.0607
	30.0209	30.0228	28.7274
DOAstd	0.2230	0.2320	1.5937
	0.2384	0.2528	2.3564

(*c*)

$\phi_{1,2} = 20^\circ, 23^\circ$	SVD	Schur 1	Schur 2
DOAmean	19.6720	19.4254	9.1648
	23.3412	23.5344	21.5303
DOAstd	2.6122	4.0261	209.8293
	2.3578	4.4285	2.2607

References

[1] D.V. Bhaskar Rao and K.S. Arun, "Model based processing of signals: A state space approach," *Proc. IEEE*, vol. 80, pp. 283–309, Feb. 1992.

[2] A.J. van der Veen, E.F. Deprettere, and A.L. Swindlehurst, "Subspace based signal analysis using singular value decomposition," *Proceedings of the IEEE*, vol. 81, pp. 1277–1308, Sept. 1993.

[3] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, 1991.

[4] G.W. Stewart, "An updating algorithm for subspace tracking," *IEEE Trans. Signal Processing*, vol. 40, pp. 1535–1541, June 1992.

[5] M. Moonen, P. Van Dooren, and F. Vanpoucke, "On the QR algorithm and updating the SVD and URV decomposition in parallel," *Lin. Alg. Appl.*, vol. 188,189, pp. 549–568, 1993.

[6] T.F. Chan, "Rank revealing *QR* factorizations," *Lin. Alg. Appl.*, vol. 88/89, pp. 67–82, 1987.

[7] C.H. Bischof and G.M. Schroff, "On updating signal subspaces," *IEEE Trans. Signal Processing*, vol. 40, pp. 96–105, Jan. 1992.

[8] A.J. van der Veen, "A Schur method for low-rank matrix approximation," *submitted*, Dec. 1993.

[9] A.J. van der Veen and P.M. Dewilde, "On low-complexity approximation of matrices," *Lin. Alg. Appl.*, vol. 203/204/205, May/June/July 1994.

[10] R. Roy, A. Paulraj, and T. Kailath, "ESPRIT—a subspace rotation approach to estimation of parameters of cisoids in noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 34, no. 5, pp. 1340–1342, 1986.