REVISITING ADAPTIVE SIGNAL SUBSPACE ESTIMATION BASED ON RAYLEIGH'S QUOTIENT

Samir Attallah

Centre for Wireless Communications
National University of Singapore
20 Science Park Road
02-34/37, Teletech Park
Singapore 117674
Email: cwcsa@leonis.nus.edu.sg

ABSTRACT

In this paper, we propose a new adaptive algorithm for subspace estimation and tracking that is based on Rayleigh's quotient. This algorithm allows the estimation of the signal subspace of a vector sequence. It has a number of interesting properties such as a low computational complexity, a fast convergence, orthogonality of the subspace vectors which is ensured at each iteration and a good numerical stability. As will be shown, the proposed algorithm outperforms Oja's algorithm.

1. INTRODUCTION

Fast estimation and tracking of the signal subspace of a sequence of random vectors is the key-stone to many applications that span a variety of areas of information processing such as data compression, parameter estimation, pattern recognition, neural networks and, in particular, wireless communications, among others [1]-[2].

Subspace estimation can be performed using the batch eigenvalue decomposition (EVD) of the (estimated) correlation matrix or the singular value decomposition (SVD) of the data matrix [3]. However, these two approaches are not suitable for adaptive applications where the required repetitive estimation of the subspace can be a real computational burden.

In the literature, many fast algorithms have been proposed [4]. These can be classified, depending on their computational complexities, as [7] $O(N^2P)$, $O(N^2)$ or $O(NP)$ where $N \times P$ represents the size of the signal (or noise) subspace weight matrix to be estimated and $P \leq N$. Among the least complex algorithms (i.e. $O(NP)$), we can mention OJA algorithm which was first developed in neural networks area [8]. One of the main advantages of OJA is its ability to estimate both signal (or principal) and noise (or minor) subspaces by simply reversing the sign of the learning parameter, say β. However, unless β is arbitrarily small, OJA algorithm diverges. Moreover, the orthogonality of the weight matrix, i.e. the orthogonality between the P subspace column vectors, is not (perfectly) ensured as shown in the simulations. The latter, however, is a very desirable property in many applications [10].

In this paper, we first present a fast review of OJA algorithm. Then, we introduce Rayleigh's quotient based adaptive subspace estimation algorithm as developed by Yang and Kaveh in [1]. The latter uses Gram-Schmidt orthogonalization at each iteration. Accordingly, it has a very high computational complexity. Next, we derive a fast algorithm based on Rayleigh's quotient followed with its normalized version. Finally, we present the simulation results of the proposed algorithm along with some comments and concluding remarks.

2. REVIEW OF OJA ALGORITHM

Consider the problem of extracting the signal (or noise) subspace spanned by the sequence $\{\mathbf{r}(k)\}$ of dimension $P < N$ which is assumed to be the span of the P signal (or noise) eigenvectors of the covariance matrix $\mathbf{C} = E[\mathbf{r}(k)\mathbf{r}^H(k)]$. To solve this problem, several subspace extraction algorithms have so far been proposed [5]-[9]. The minor subspace extraction algorithm by Oja *et al.* [8] can be formulated as

$$\begin{aligned}\mathbf{W}(i+1) &= \mathbf{W}(i) - \beta\left[\mathbf{r}(i)\mathbf{y}^H(i) - \mathbf{W}(i)\mathbf{y}(i)\mathbf{y}^H(i)\right] \\ &= \mathbf{W}(i) - \beta\mathbf{p}(i)\mathbf{y}^H(i) \qquad (1)\end{aligned}$$

where $\mathbf{W}(i) \in \mathbb{R}^{N\times P}$ is the minor subspace estimate, $\mathbf{y}(i) \triangleq \mathbf{W}^H(i)\mathbf{r}(i)$, $\mathbf{p}(i) \triangleq (\mathbf{r}(i) - \mathbf{W}(i)\mathbf{y}(i))$, and $\beta > 0$ is a learning parameter. Reversing the sign of the adaptive gain, i.e., replacing $-\beta$ in (1) by $+\beta$, yields a principal (signal) subspace extraction algorithm.

3. YANG AND KAVEH ADAPTIVE SUBSPACE ESTIMATION ALGORITHM

As shown in [1] the iterative maximization of the cost function

$$J_{\mathbf{W}} = tr(\mathbf{W}^H \mathbf{C}\mathbf{W}) \tag{2}$$

subject to

$$\mathbf{W}^H \mathbf{W} = I \tag{3}$$

converges to the signal subspace of **C**. This maximization can be achieved by using the gradient-descent technique, that is

$$\mathbf{W}(i+1) = \mathbf{W}(i) + \beta \nabla_J(i) \tag{4}$$

where $\beta > 0$ and the gradient is given by

$$\nabla_J = 2\mathbf{C}\mathbf{W} \tag{5}$$

By replacing the gradient and the correlation matrix by their corresponding instantaneous estimates, (4) becomes

$$\mathbf{W}(i+1) = \mathbf{W}(i) + 2\beta \mathbf{r}(i)\mathbf{r}^H(i)\mathbf{W}(i) \tag{6}$$

Condition (3) can be satisfied by using Gram-Schmidt orthogonalization at each iteration [1]. However, this algorithm presents a number of drawbacks [4]:

- Extremely high computational complexity
- Slow convergence
- Stability problems.

In the following, we derive a new algorithm which is still based on Rayleigh's quotient, but without the previously mentioned drawbacks. We shall call it the fast Rayleigh's quotient based (FRQ) algorithm.

4. FRQ ADAPTIVE SUBSPACE ESTIMATION ALGORITHM

The proposed algorithm consists of (6) plus an orthogonalization step of the weight matrix to be performed at each iteration. Using informal notation, we can write:

$$\mathbf{W}(i+1) := \mathbf{W}(i+1)(\mathbf{W}^H(i+1)\mathbf{W}(i+1))^{-1/2} \tag{7}$$

where $(\mathbf{W}^H(i+1)\mathbf{W}(i+1))^{-1/2}$ denotes an inverse square root of $(\mathbf{W}^H(i+1)\mathbf{W}(i+1))$. To compute the latter, we use the updating equation of $\mathbf{W}(i+1)$. Keeping in mind that $\mathbf{W}(i)$ is now an orthogonal matrix and setting $\mathbf{y}(i) = \mathbf{W}^H(i)\mathbf{r}(i)$, we have

$$\begin{aligned}\mathbf{W}^H(i+1)\mathbf{W}(i+1) &= \mathbf{I} + 4\mathbf{y}(i)\mathbf{y}^H(i)[\beta + \beta^2\|\mathbf{r}(i)\|^2] \\ &= \mathbf{I} + \mathbf{x}\mathbf{x}^H.\end{aligned} \tag{8}$$

where $\mathbf{I}$ is the identity matrix, and $\mathbf{x} \triangleq 2\sqrt{\beta + \beta^2\|\mathbf{r}(i)\|^2}\mathbf{y}(i)$. Let us set $\rho = 4\beta(1 + \beta\|\mathbf{r}(i)\|^2)$. Then, using [11]

$$(\mathbf{I} + \mathbf{x}\mathbf{x}^H)^{-1/2} = \mathbf{I} + (\frac{1}{\sqrt{1 + \|\mathbf{x}\|^2}} - 1)\frac{\mathbf{x}\mathbf{x}^H}{\|\mathbf{x}\|^2},$$

we obtain

$$(\mathbf{W}^H(i+1)\mathbf{W}(i+1))^{-1/2} = \mathbf{I} + \tau(i)\mathbf{y}(i)\mathbf{y}^H(i), \tag{9}$$

where $\tau(i) \triangleq \frac{1}{\|\mathbf{y}(i)\|^2}(\frac{1}{\sqrt{1 + \rho(i)\|\mathbf{y}(i)\|^2}} - 1)$. Substituting (9) into (7) and using the updating equation of $\mathbf{W}(i+1)$ leads to

$$\begin{aligned}\mathbf{W}(i+1) &= (\mathbf{W}(i) + 2\beta\mathbf{r}(i)\mathbf{y}^H(i))(\mathbf{I} + \tau(i)\mathbf{y}(i)\mathbf{y}^H(i)) \\ &= \mathbf{W}(i) + \beta\overline{\mathbf{p}}(i)\mathbf{y}^H(i),\end{aligned} \tag{10}$$

where $\overline{\mathbf{p}}(i) \triangleq \tau(i)\mathbf{W}(i)\mathbf{y}(i)/\beta + 2\mathbf{r}(i)[1 + \tau(i)\|\mathbf{y}(i)\|^2]$.

5. NORMALIZED FRQ (NFRQ) ALGORITHM

At iteration $(i+1)$, the cost function (2) can be written as

$$J_{\mathbf{W}}(i+1) = tr(\mathbf{W}^H(i+1)\mathbf{C}\mathbf{W}(i+1)) \tag{11}$$

Now injecting (4) into (11) and using a variable stepsize $\beta(i)$, we can write

$$\begin{aligned}J_{\mathbf{W}}(i+1) &= tr([\mathbf{W}^H(i) + \beta(i)\nabla_J^H(i)]\mathbf{C}[\mathbf{W}(i) \\ &+ \beta(i)\nabla_J(i)]) \\ &= J_{\mathbf{W}}(i) + 2\beta(i)tr(\mathbf{W}^H(i)\mathbf{C}\nabla_J(i)) \\ &+ \beta^2(i)tr(\nabla_J^H(i)\mathbf{C}\nabla_J(i))\end{aligned} \tag{12}$$

If **C** is strictly positive definite, then

$$tr(\nabla_J^H(i)\mathbf{C}\nabla_J(i)) > 0,\ \forall\ \nabla_J(i) \neq 0 \tag{13}$$

This means that $J_{\mathbf{W}}(\beta)$, which is a quadratic function of $\beta(i)$, has a global minimum. If we assume that $J_{\mathbf{W}}(i)$ is independent of $\beta(i)$, then

$$\begin{aligned}\frac{\partial J_{\mathbf{W}}(i+1)}{\partial\beta(i)} &= 2tr(\mathbf{W}^H(i)\mathbf{C}\nabla_J(i)) \\ &+ 2\beta(i)tr(\nabla_J^H(i)\mathbf{C}\nabla_J(i))\end{aligned} \tag{14}$$

Hence, the optimal stepsize can be found by setting (14) to zero. This leads to

$$\beta_{opt}(i) = -\frac{tr(\mathbf{W}^H(i)\mathbf{C}\nabla_J(i))}{tr(\nabla_J^H(i)\mathbf{C}\nabla_J(i))} \tag{15}$$

However, we are interested only in the values of β that maximize the cost function. This can be achieved in a suboptimal way by replacing $\beta(i)$ in (12) by the negative value of (15). In this case, we get

$$J_{\mathbf{W}}(i+1) = J_{\mathbf{W}}(i) + 2\frac{tr^2(\mathbf{W}^H(i)\mathbf{C}\nabla_J(i))}{tr(\nabla_J^H(i)\mathbf{C}\nabla_J(i))} + \frac{tr^2(\mathbf{W}^H(i)\mathbf{C}\nabla_J(i))}{tr(\nabla_J^H(i)\mathbf{C}\nabla_J(i))} \quad (16)$$

$$\Longrightarrow J_{\mathbf{W}}(i+1) - J_{\mathbf{W}}(i) = 3\frac{tr^2(\mathbf{W}^H(i)\mathbf{C}\nabla_J(i))}{tr(\nabla_J^H(i)\mathbf{C}\nabla_J(i))} \triangleq \epsilon(i) \quad (17)$$

We can easily notice that $\epsilon(i) \geq 0$ or

$$J_{\mathbf{W}}(i+1) - J_{\mathbf{W}}(i) \geq 0 \quad (18)$$

Therefore, the MSE will increase at each iteration (maximization) by an increment of $\epsilon(i)$. Note that the MSE will increase even if we replace **C** and the gradient by their corresponding estimates. In practice, the correlation matrix **C** and the gradient are replaced by their instantaneous values. In this case, we obtain

$$\widehat{\beta}_{sopt}(i) = \frac{1}{2\|\mathbf{r}(i)\|^2} \quad (19)$$

Thus, NFRQ algorithm can be written as

- Initialization of the algorithm:

$$\mathbf{W}(0) = \text{any arbitrary orthogonal matrix.}$$

- Algorithm at iteration i:

$$\begin{aligned}
\mathbf{y}(i) &= \mathbf{W}^H(i)\mathbf{r}(i) \\
\widehat{\beta}_{sopt}(i) &= \frac{\beta}{2\|\mathbf{r}\|^2 + \gamma} \\
\rho(i) &= 4\widehat{\beta}_{sopt}(i)(1 + \widehat{\beta}_{sopt}(i)\|\mathbf{r}(i)\|^2) \\
\tau(i) &= \frac{1}{\|\mathbf{y}(i)\|^2}\left(\frac{1}{\sqrt{1+\rho(i)\|\mathbf{y}(i)\|^2}} - 1\right) \\
\overline{\mathbf{p}}(i) &= \tau(i)\mathbf{W}(i)\mathbf{y}(i)/\widehat{\beta}_{sopt}(i) + \\
&\quad 2\mathbf{r}(i)(1 + \tau(i)\|\mathbf{y}(i)\|^2) \\
\mathbf{W}(i+1) &= \mathbf{W}(i) + \widehat{\beta}_{sopt}(i)\overline{\mathbf{p}}(i)\mathbf{y}^H(i)
\end{aligned}$$

where $0 < \beta \leq 1$ and γ is a small positive constant which improves the stability of the algorithm.

6. SIMULATION-RESULTS

In the following, we choose $\mathbf{r}(i)$ to be a sequence of independent jointly-Gaussian random vectors with covariance matrix

$$\mathbf{C} = \begin{pmatrix} 0.9 & 0.4 & 0.7 & 0.3 \\ 0.4 & 0.3 & 0.5 & 0.4 \\ 0.7 & 0.5 & 1.0 & 0.6 \\ 0.3 & 0.4 & 0.6 & 0.9 \end{pmatrix} \quad (20)$$

$P = 2$ and as recommended in [9] $\mathbf{W}(0) = \mathbf{D}$, where $\mathbf{D}_{i,j} = \delta(j-i)$. As in [6], we calculate the ensemble averages of the performance factors

$$\rho(i) = \frac{1}{r_0}\sum_{r=1}^{r_0} \frac{\text{tr}\left(\mathbf{W}_r^H(i)\mathbf{E}_1 * \mathbf{E}_1^H\mathbf{W}_r(i)\right)}{\text{tr}\left(\mathbf{W}_r^H(i)\mathbf{E}_2 * \mathbf{E}_2^H\mathbf{W}_r(i)\right)}, \quad (21)$$

$$\eta(i) = \frac{1}{r_0}\sum_{r=1}^{r_0} \|\mathbf{W}_r^H(i)\mathbf{W}_r(i) - \mathbf{I}\|_F^2, \quad (22)$$

where the number of algorithm runs is $r_0 = 40$, r indicates that the associated variable depends on the particular run, $\|.\|_F$ denotes the Frobenius norm, and $\mathbf{E}_1$ (respectively $\mathbf{E}_2$) is the principal $(N-P)$-dimensional (respectively minor P-dimensional) subspace. Figure 1 compares the performance of OJA with our algorithm. As we can see NFRQ converges practically at the same speed as OJA for $\beta = 0.03$. However, it ensures the (perfect) orthogonality of the signal subspace vectors as shown by η curve. For higher values of β, NFRQ converges much faster at the expense, however, of an increased steady state error. Whereas, when β is increased, OJA algorithm just diverges. In general, our algorithm behaves much better than (1) and do not suffer from numerical instability. Moreover it has a comparable computational complexity, i.e., $O(NP)$.

7. CONCLUSION

In this paper, we proposed a new fast adaptive algorithm for signal subspace estimation that is based on Rayleigh's quotient. The proposed algorithm converges faster than OJA and is numerically much more stable. In addition, it has a lower computational complexity ($O(NP)$) as compared to Yang *et al.* algorithm and ensures the orthogonality of the subspace eigenvectors at each iteration. It is worthwhile to note that the same procedure can be used to develop a fast adaptive noise subspace estimation algorithm based on Rayleigh's quotient as well.

8. ACKNOWLEDGMENT

The author would like to thank K. Abed-Meraim, from Telecom Paris, for helpful discussions.

9. REFERENCES

[1] J. F. Yang and M. Kaveh, ''Adaptive Eigensubspace Algorithms for Direction or Frequency Estimation,'' *IEEE Trans. On Acous., Spe., and Sig. Proc.*, vol. 36, no. 2, pp. 241-251, Feb. 1988.

[2] X. Wang and H. V. Poor, ''Blind Multiuser Detection: A Subspace Approach,'' *IEEE Trans. On Inf. Theory*, vol. 44, no. 2, pp. 677-689, March 1998.

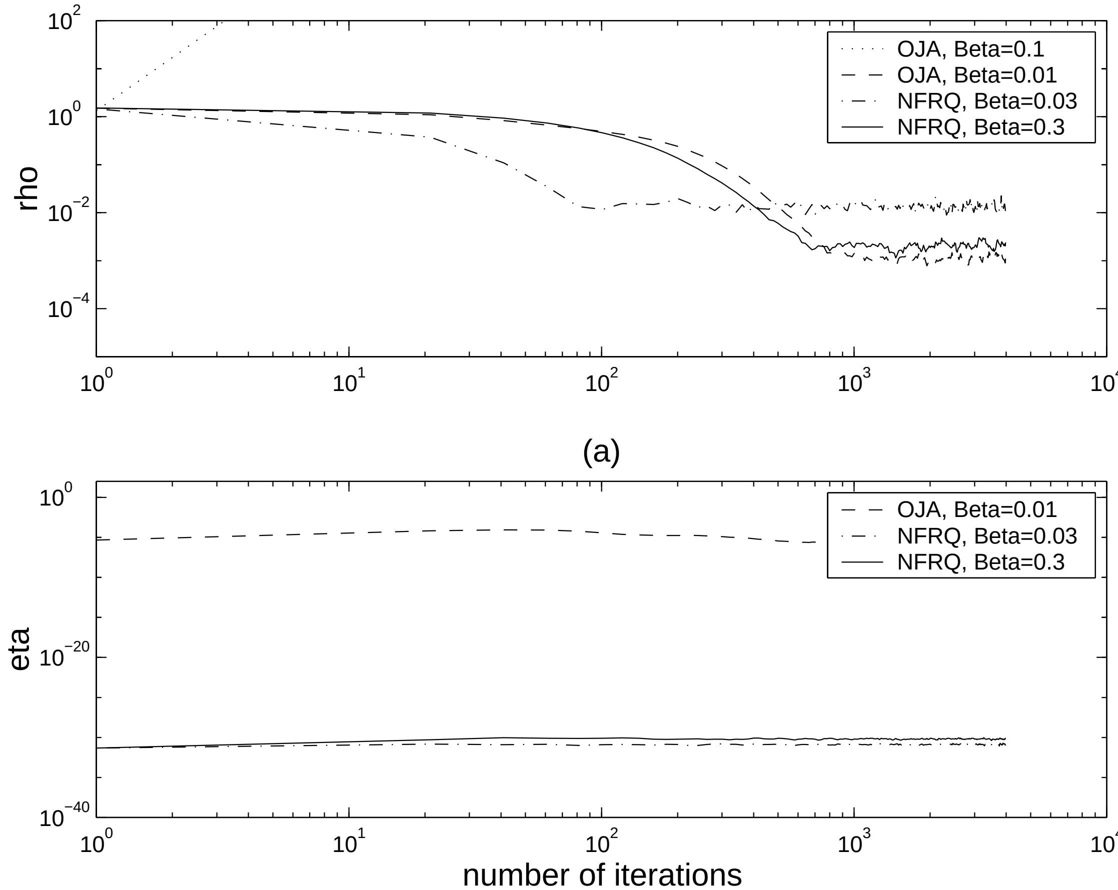

Fig. 1. Average behaviors for signal subspace estimation (Proposed and OJA): evaluation of ρ and η.

[3] G. H. Golub and C. F. Van loan, "Matrix computations," the Johns Hopkins University Press, 1996.

[4] P. Comon and G. H. Golub, ‘‘Tracking a Few Extreme Singular Values and Vectors in Signal Processing,’’ *Proceedings of the IEEE*, vol. 78, no. 8, pp. 1327-1343, August 1990.

[5] T. P. Krasulina, ‘‘Method of Stochastic Approximation in the Determination of the Largest Eigenvalue of the Mathematical Expectation of Random Matrices,’’ *Automat. Remote Contr.*, vol 2, pp. 215-221, 1970.

[6] S. C. Douglas, S.-Y. Kung, and S.-I. Amari, ‘‘A Self-Stabilized Minor Subspace Rule,’’ *Sig. Process. Letters*, vol. 5, no. 12, pp. 328-330, Dec. 1998.

[7] B. Yang, ‘‘Projection Approximation Subspace Tracking,’’ *IEEE Signal Processing*, vol. 43, no. 1, pp. 95-107, Jan. 1995.

[8] E. Oja, ‘‘Principal Components, Minor Components, and Linear Neural Networks,’’ *Neural Networks*, vol. 5, pp. 927-935, Nov./Dec. 1992.

[9] T. Chen, S.-I. Amari, and Q. Lin, ‘‘A Unified Algorithm for Principal and Minor Components Extraction,’’ *Neural Networks*, vol. 11, pp. 385-390, 1998.

[10] S. Marcos, A. Marsal, and M. Benidir, ‘‘The Propagator Method for Source Bearing Estimation,’’ *Sig. Proc.*, vol 42, pp. 121-138, Apr. 1995.

[11] K. Abed-Meraim, S. Attallah, A. Chkeif and Y. Hua, ‘‘Orthogonal OJA Algorithm,’’ *IEEE Signal Processing Letters*, pp. 116-119,