WEIGHTED SUBSPACE FITTING
USING SUBSPACE PERTURBATION EXPANSIONS

Richard J. Vaccaro

Department of Electrical Engineering
University of Rhode Island
Kingston, RI 02881 USA
(vaccaro@ele.uri.edu)

ABSTRACT

This paper presents a new approach to deriving statistically optimal weights for weighted subspace fitting (WSF) algorithms. The approach uses a formula called a "subspace perturbation expansion," which shows how the subspaces of a matrix change when the matrix elements are perturbed. The perturbation expansion is used to derive an optimal WSF algorithm for estimating directions of arrival in array signal processing.

1. INTRODUCTION

A variety of parameter estimation problems in signal processing and system identification can be solved using "subspace" methods. These methods rely on the fact that a rank deficient matrix can be formed from noise-free data. Furthermore, information about the signal or system parameters is embedded in the column space and/or row space of this matrix. With noisy data, the appropriate subspace is estimated, usually with the singular value decomposition, and the parameters are extracted from the estimated subspaces.

In [7], a formula is derived which shows how much of a subspace perturbation is induced by additive noise in the data. This formula is called a "subspace perturbation expansion." In this paper, this expansion is used to derive an optimally weighted subspace fitting algorithm for estimating directions of arrival in array signal processing. The resulting cost function is different than the original WSF cost function derived in [8]; however, the two cost functions yield the same performance.

Weighted subspace fitting was introduced as a method for parameter estimation in [8, 5]. All of the previous work on weighted subspace fitting uses the asymptotic (in the number of data snapshots) distribution of sample eigenvectors to derive statistically optimal weights. In contrast, the

subspace perturbation expansion is based on a data matrix of finite size. Nevertheless, it will be seen that for the DOA problem, the subspace perturbation approach yields results that are essentially identical to those obtained by the asymptotic approach. The subspace perturbation approach may have advantages in other applications in which the data matrix is structured (e.g. Hankel structured data matrices in signal modeling and system identification).

Throughout this paper, the superscript 'T' refers to the matrix transpose, the superscript 'H' refers to the complex-conjugate transpose, and the superscript '*' refers to the complex conjugate.

2. SUBSPACE PERTURBATION EXPANSION

Let $\mathbf{Y}$ be an $m \times n$ matrix of rank p, where $p < \min(m, n)$. The singular value decomposition of $\mathbf{Y}$ can be partitioned as follows:

$$\mathbf{Y} = \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^H \\ \mathbf{V}_2^H \end{bmatrix}.$$

We are interested in the subspaces $\text{col}(\mathbf{U}_1)$ and $\text{col}(\mathbf{U}_2)$, where $\text{col}(\mathbf{A})$ denotes the column space of the matrix $\mathbf{A}$. Let $\mathbf{Y}$ be perturbed as follows

$$\tilde{\mathbf{Y}} = \mathbf{Y} + \mathbf{N}.$$

The SVD of $\tilde{\mathbf{Y}}$ can be partitioned as follows:

$$\tilde{\mathbf{Y}} = \begin{bmatrix} \tilde{\mathbf{U}}_1 & \tilde{\mathbf{U}}_2 \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{\Sigma}}_1 & \mathbf{0} \\ \mathbf{0} & \tilde{\boldsymbol{\Sigma}}_2 \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{V}}_1^H \\ \tilde{\mathbf{V}}_2^H \end{bmatrix}.$$

It is shown in [7] that we can find an orthonormal basis for $\text{col}(\tilde{\mathbf{U}}_1)$ in the following form

$$\tilde{\mathbf{X}}_1 = (\mathbf{U}_1 + \mathbf{U}_2 \mathbf{P})(\mathbf{I} + \mathbf{P}^H \mathbf{P})^{-\frac{1}{2}} \quad (1)$$

where $\mathbf{P}$ is a coefficient matrix. In addition, an orthonormal basis for $\text{col}(\tilde{\mathbf{U}}_2)$ can be found in the following form:

$$\tilde{\mathbf{X}}_2 = (-\mathbf{U}_1 \mathbf{P}^H + \mathbf{U}_2)(\mathbf{I} + \mathbf{P} \mathbf{P}^H)^{-\frac{1}{2}}. \quad (2)$$

The coefficient matrix $\mathbf{P}$ in (1) and (2) can be expressed as matrix series

$$\mathbf{P} = \mathbf{0} + \mathbf{P}^{(1)} + \mathbf{P}^{(2)} + \cdots \qquad (3)$$

where the superscript (j) refers to a matrix product containing j factors of the perturbation matrix $\mathbf{N}$. It is shown in [7] that

$$\mathbf{P}^{(1)} = \mathbf{U}_2^H \mathbf{N} \mathbf{V}_1 \boldsymbol{\Sigma}_1^{-1}. \qquad (4)$$

Thus, using (2), the first-order expression for a basis for the perturbed subspace $\operatorname{col}(\tilde{\mathbf{U}}_2)$ is

$$\tilde{\mathbf{X}}_2^{(1)} = -\mathbf{U}_1 \mathbf{P}^{(1)H} + \mathbf{U}_2.$$

To first order, this basis is orthonormal. A formula for $\mathbf{P}^{(2)}$ is available in [7].

3. WEIGHTED SUBSPACE FITTING

In this section we consider the direction-of-arrival (DOA) estimation problem in array signal processing and show how to use the subspace perturbation expansion to derive statistically optimal weights in a subspace fitting algorithm. We begin with the standard data model for narrowband DOA estimation. The model for the noise-free signal is

$$\mathbf{Y} = \mathbf{A}(\boldsymbol{\theta}_0)\mathbf{S}$$

where $\mathbf{Y}$ is $m \times n$, $\mathbf{A}(\boldsymbol{\theta}_0)$ is $m \times r$, and $\mathbf{S}$ is $r \times n$. In this application, m is the number of sensors, n is the number of snapshots of array data, and r is the number of narrowband signals impinging on the array. The vector of possible DOAs is $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$ denotes the actual DOAs. The SVD of $\mathbf{Y}$ is

$$\mathbf{Y} = \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^H \\ \mathbf{V}_2^H \end{bmatrix}$$

where $\mathbf{U}_1$ has r columns. The crucial observation about subspaces is that columns of $\mathbf{U}_1$ and $\mathbf{A}(\boldsymbol{\theta}_0)$ span the same subspace, and therefore columns of $\mathbf{U}_2$ are orthogonal to columns of $\mathbf{A}(\boldsymbol{\theta}_0)$.

The observed (noisy) data is

$$\tilde{\mathbf{Y}} = \mathbf{Y} + \mathbf{N}$$

where the elements of $\mathbf{N}$ are taken to be zero mean i.i.d. complex Gaussian random variables with variance σ^2 (real and imaginary parts are uncorrelated). The SVD of $\tilde{\mathbf{Y}}$ is

$$\tilde{\mathbf{Y}} = \begin{bmatrix} \tilde{\mathbf{U}}_1 & \tilde{\mathbf{U}}_2 \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{\Sigma}}_1 & \mathbf{0} \\ \mathbf{0} & \tilde{\boldsymbol{\Sigma}}_2 \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{V}}_1^H \\ \tilde{\mathbf{V}}_2^H \end{bmatrix}$$

where $\tilde{\mathbf{U}}_1$ has r columns. We assume that the number of signals r has been correctly estimated.

The subspace-fitting criterion used in this paper is based on the fact that, in the noise-free case[1],

$$\mathbf{U}_2 \mathbf{U}_2^H \mathbf{A}(\boldsymbol{\theta}_0) = \mathbf{0}.$$

With noisy data the previous expression will not equal zero and we look for the parameter vector $\hat{\boldsymbol{\theta}}$ that minimizes the equation error:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \| \tilde{\mathbf{U}}_2 \tilde{\mathbf{U}}_2^H \mathbf{A}(\boldsymbol{\theta}) \|_{\mathbf{W}}^2 \qquad (5)$$

where the norm is defined as

$$\| \cdot \|_{\mathbf{W}}^2 = \operatorname{vec}(\cdot)^H \mathbf{W} \operatorname{vec}(\cdot)$$

and $\mathbf{W}$ is a weight matrix. We define

$$\mathcal{C}(\boldsymbol{\theta}) \stackrel{\text{def}}{=} (\tilde{\mathbf{U}}_2 \tilde{\mathbf{U}}_2^H \mathbf{A}(\boldsymbol{\theta})), \text{ and } \mathbf{c}(\boldsymbol{\theta}) \stackrel{\text{def}}{=} \operatorname{vec}(\mathcal{C}(\boldsymbol{\theta})). \qquad (6)$$

It will be shown that, to first order, $\mathbf{c}(\boldsymbol{\theta})$ is a zero-mean Gaussian random vector. Thus, the optimal weight matrix to use in (5) is [1]

$$\mathbf{W} = [\mathrm{E}(\mathbf{c}(\boldsymbol{\theta}_0)\mathbf{c}^H(\boldsymbol{\theta}_0))]^{\dagger}$$

where $\dagger$ denotes pseudo inverse. We will deal with the fact that $\mathbf{W}$ is a function of $\boldsymbol{\theta}_0$ at the end of this section.

Using the perturbation expansion, the projection matrix $\tilde{\mathbf{U}}_2 \tilde{\mathbf{U}}_2^H$ can be written as follows:

$$\tilde{\mathbf{U}}_2 \tilde{\mathbf{U}}_2^H = (\mathbf{U}_2 - \mathbf{U}_1 \mathbf{P}^H)(\mathbf{I} + \mathbf{P}\mathbf{P}^H)^{-1}(\mathbf{U}_2 - \mathbf{U}_1 \mathbf{P}^H)^H.$$

Approximating $\mathbf{P}$ up to first order and using the fact that, for any matrix $\mathbf{M}$ whose norm is less than one,

$$(\mathbf{I} + \mathbf{M})^{-1} = \mathbf{I} - \mathbf{M} + \mathbf{M}^2 - \cdots$$

we have

$$\begin{aligned} \tilde{\mathbf{U}}_2 \tilde{\mathbf{U}}_2^H &\stackrel{1}{=} (\mathbf{U}_2 - \mathbf{U}_1 \mathbf{P}^{(1)H})(\mathbf{U}_2^H - \mathbf{P}^{(1)} \mathbf{U}_1^H) \\ &\stackrel{1}{=} \mathbf{U}_2 \mathbf{U}_2^H - \mathbf{U}_1 \mathbf{P}^{(1)H} \mathbf{U}_2^H - \mathbf{U}_2 \mathbf{P}^{(1)} \mathbf{U}_1^H \end{aligned} \qquad (7)$$

where $\mathbf{P}^{(1)}$ is given by (4) and '$\stackrel{1}{=}$' means "equal up to first-order perturbation terms." Substituting (7) and (4) into (6) yields

$$\mathbf{c}(\boldsymbol{\theta}_0) \stackrel{1}{=} \operatorname{vec}(-\mathbf{U}_2 \mathbf{U}_2^H \mathbf{N} \mathbf{V}_1 \boldsymbol{\Sigma}_1^{-1} \mathbf{U}_1^H \mathbf{A}(\boldsymbol{\theta}_0)).$$

We can neglect the minus sign and use the properties of Kronecker products [2] to write

$$\begin{aligned} \mathbf{c}(\boldsymbol{\theta}_0) &= [(\mathbf{V}_1 \boldsymbol{\Sigma}_1^{-1} \mathbf{U}_1^H \mathbf{A}(\boldsymbol{\theta}_0))^T \otimes (\mathbf{U}_2 \mathbf{U}_2^H)]\operatorname{vec}(\mathbf{N}) \\ &\stackrel{\text{def}}{=} \mathbf{B}\mathbf{n}. \end{aligned} \qquad (8)$$

---

[1] We use the projection matrix $\mathbf{U}_2 \mathbf{U}_2^H$ instead of just $\mathbf{U}_2^H$ because the perturbation expansion gives a basis for the subspace $\operatorname{col}(\tilde{\mathbf{U}}_2)$, *not* a representation for individual singular vectors.

Then
$$\mathrm{E}[\mathbf{c}(\boldsymbol{\theta}_0)\mathbf{c}^H(\boldsymbol{\theta}_0)] = \sigma^2 \mathbf{B}\mathbf{B}^H \overset{\text{def}}{=} \mathbf{R}.$$

To simplify the expression for the covariance matrix, consider the following SVD:

$$[\mathbf{V}_1 \boldsymbol{\Sigma}_1^{-1} \mathbf{U}_1^H \mathbf{A}(\boldsymbol{\theta}_0)]^T = \mathbf{U}_3 \boldsymbol{\Sigma}_3 \mathbf{V}_3^H. \qquad (9)$$

Substituting this into (8) and using the properties of Kronecker products yields

$$\mathbf{B} = (\mathbf{U}_3 \otimes \mathbf{U}_2)(\boldsymbol{\Sigma}_3 \otimes \mathbf{I})(\mathbf{V}_3^H \otimes \mathbf{U}_2^H)$$

and

$$\mathbf{R}^\dagger = (\mathbf{U}_3 \otimes \mathbf{U}_2)(\boldsymbol{\Sigma}_3^{-2} \otimes \mathbf{I})(\mathbf{U}_3^H \otimes \mathbf{U}_2^H)$$
$$= (\mathbf{U}_3 \boldsymbol{\Sigma}_3^{-2} \mathbf{U}_3^H) \otimes (\mathbf{U}_2 \mathbf{U}_2^H).$$

Then we have $\mathbf{c}^H(\boldsymbol{\theta})\mathbf{R}^\dagger \mathbf{c}(\boldsymbol{\theta})$

$$= \|(\boldsymbol{\Sigma}_3^{-1} \mathbf{U}_3^H \otimes \mathbf{U}_2^H)\mathbf{c}(\boldsymbol{\theta})\|^2$$
$$= \|(\boldsymbol{\Sigma}_3^{-1} \mathbf{U}_3^H \otimes \mathbf{U}_2^H) \, \mathrm{vec}(\tilde{\mathbf{U}}_2 \tilde{\mathbf{U}}_2^H \mathbf{A}(\boldsymbol{\theta}))\|^2 \qquad (10)$$
$$= \|\mathbf{U}_2^H \tilde{\mathbf{U}}_2 \tilde{\mathbf{U}}_2^H \mathbf{A}(\boldsymbol{\theta})\mathbf{U}_3^* \boldsymbol{\Sigma}_3^{-1}\|_F^2.$$

We can simplify this expression by using (9) and noting that

$$\mathbf{U}_3^* \boldsymbol{\Sigma}_3^2 \mathbf{U}_3^H = \mathbf{A}^H(\boldsymbol{\theta}_0)\mathbf{U}_1 \boldsymbol{\Sigma}_1^{-2} \mathbf{U}_1^H \mathbf{A}(\boldsymbol{\theta}_0)$$

and so

$$\mathbf{U}_3^* \boldsymbol{\Sigma}_3^{-2} \mathbf{U}_3^H = (\mathbf{U}_1^H \mathbf{A}(\boldsymbol{\theta}_0))^{-1} \boldsymbol{\Sigma}_1^2 \mathbf{A}(\boldsymbol{\theta}_0)\mathbf{U}_1.$$

Using this equation and the fact that

$$\|\mathbf{M}\|_F^2 = \mathrm{trace}(\mathbf{M}\mathbf{M}^H)$$

we can rewrite (10) as

$$\|\mathbf{U}_2^H \tilde{\mathbf{U}}_2 \tilde{\mathbf{U}}_2^H \mathbf{A}(\boldsymbol{\theta})(\mathbf{U}_1^H \mathbf{A}(\boldsymbol{\theta}_0))^{-1} \boldsymbol{\Sigma}_1\|_F^2. \qquad (11)$$

This is the optimally weighted criterion that we want to minimize. However, it depends on the true parameter vector $\boldsymbol{\theta}_0$ as well as the SVD of the noise-free signal matrix. In order to get a computable cost function we use $\tilde{\mathbf{U}}_1$ and $\tilde{\mathbf{U}}_2$ in place of $\mathbf{U}_1$ and $\mathbf{U}_2$. In place of $\boldsymbol{\Sigma}_1$ we use

$$\hat{\boldsymbol{\Sigma}}_1 = (\tilde{\boldsymbol{\Sigma}}_1^2 - n\hat{\sigma}^2 \mathbf{I})^{0.5}$$

where $\hat{\sigma}^2$ is the average of the squared singular values in $\tilde{\boldsymbol{\Sigma}}_2$. In place of $\boldsymbol{\theta}_0$ we simply use the $\boldsymbol{\theta}$. Thus, the DOAs are estimated by solving the following optimization problem:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \|\tilde{\mathbf{U}}_2^H \mathbf{A}(\boldsymbol{\theta})(\tilde{\mathbf{U}}_1^H \mathbf{A}(\boldsymbol{\theta}))^{-1} \hat{\boldsymbol{\Sigma}}_1\|_F^2. \qquad (12)$$

The gradient for this cost function is derived in [7].

4. EXAMPLE

The array used in this example is a uniform linear array of $m = 10$ sensors with half wavelength spacing [4]. For such an array, the matrix $\mathbf{A}(\boldsymbol{\theta})$ has the form

$$\mathbf{A}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{a}(\omega_1) & \mathbf{a}(\omega_2) & \cdots & \mathbf{a}(\omega_r) \end{bmatrix}$$

where

$$\mathbf{a}(\omega_i) = \begin{bmatrix} 1 & e^{j\omega_i} & e^{j2\omega_i} & \cdots & e^{j(m-1)\omega_i} \end{bmatrix}^H,$$
$$\omega_i = \pi \sin\theta_i, \; i = 1, \cdots, r. \qquad (13)$$

The noisy data is generated as

$$\tilde{\mathbf{Y}} = \mathbf{A}(\boldsymbol{\theta}_0)\mathbf{S} + \mathbf{N}$$

where each column of $\mathbf{S}$ consists of complex-valued i.i.d. Gaussian random variables with variance $\sigma_S^2(i)$ (we consider uncorrelated sources). The elements of the observation noise matrix $\mathbf{N}$ are also complex-valued i.i.d. Gaussian random variables, uncorrelated with those in $\mathbf{S}$, with variance σ^2. The signal-to-noise ratio in dB for the ith source is defined to be

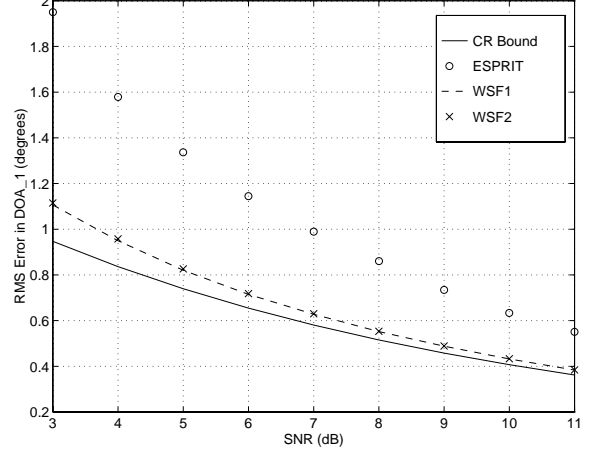$$\mathrm{SNR}_i = 10\log_{10}\frac{\sigma_S^2(i)}{\sigma^2}.$$

Figure 1: RMS error of $\hat{\theta}_1$ vs. SNR

In the first example the data were generated with $r = 2$ sources. The first source was at an angle of $-3°$ relative to the array broadside and the second source was at an angle of $2°$. The number of snapshots (columns of $\mathbf{S}$) was fixed at $n = 10$. The SNR was varied from 3 to 11 dB. Below 3 dB outliers begin to appear. Fig. 1 shows the RMS error for estimates of the DOA at $-3°$ (the other DOA estimate behaves similarly). Initial estimates were obtained using ESPRIT [6, 3]. The estimates labeled 'WSF1' were obtained

using the weighted subspace fitting approach of Viberg and Ottersten [8]. The estimates labeled 'WSF2' were obtained by minimizing (12). The CR bound for this problem is derived in [5]. This bound is asymptotic n, the number of data snapshots. That is, a statistically efficient method will achieve this bound when n is large enough. From Fig. 1 we see that the bound is nearly attained for 10 snapshots.
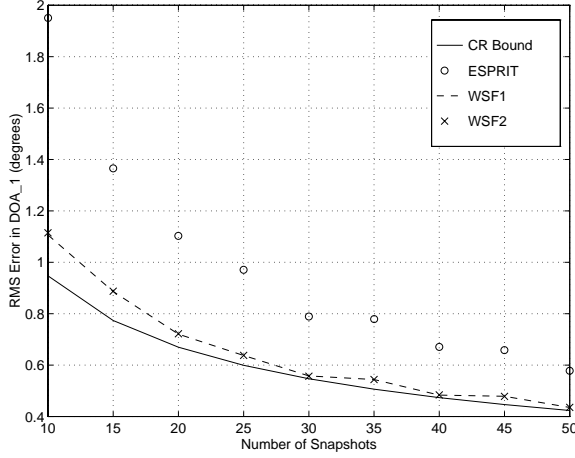

Figure 2: RMS error of $\hat{\theta}_1$ vs. number of snapshots

The second example is the same as the first except that the SNR is fixed at 3 dB and the number of snapshots is varied from 10 to 50. As seen in Fig. 2, when the number of snapshots exceeds 30 both WSF1 and WSF2 essentially achieve the CR bound. The third example is the same as the first except that the SNR is fixed at 5 dB and the source separation is varied by changing θ_2 from $1°$ to $7°$. The results are shown in Fig. 3.
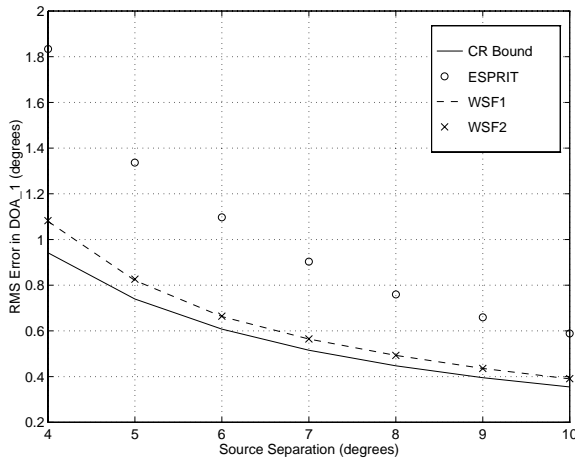

Figure 3: RMS error of $\hat{\theta}_1$ vs. source separation

5. CONCLUSIONS

A subspace perturbation expansion was presented as a new approach to deriving optimal weights for a weighted subspace fitting algorithm. When applied to the DOA estimation problem, the performance of this new weighted subspace fitting (WSF) algorithm achieves the CR bound and is essentially identical to the WSF algorithm of Viberg and Ottersten which is based on the asymptotic distribution of sample eigenvectors. In this paper, only the first-order perturbation expansion was used. The second-order expansion from [7] could be used to derive a WSF algorithm that could have better performance than the existing WSF algorithms. The precise relationship between the asymptotic approach and the subspace perturbation expansion approach is a topic for future investigation.

6. REFERENCES

[1] J.-F Cardoso, P. Loubaton, and E. Moulines. On weighted subspace estimates in system identification. In *Proc. IEEE SSAP Workshop*, Corfou, 1996.

[2] A. Graham. *Kronecker Products and Matrix Calculus*. Ellis Horwood, Ltd., 1981.

[3] H. Krim and M. Viberg. Two decades of array signal processing research. *IEEE Signal Processing Magazine*, pages 67–94, July 1996.

[4] F. Li and R.J. Vaccaro. Unified analysis for doa estimation algorithms in array signal processing. *Signal Processing*, 25:147–169, 1991.

[5] B. Ottersten, M. Viberg, and T. Kailath. Analysis of subspace fitting and ML techniques for parameter estimation from sensor array data. *IEEE Transactions on Signal Processing*, 40:590–600, March 1992.

[6] R. Roy and T. Kailath. ESPRIT - estimation of signal parameters via rotational invariance techniques. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 35:984–995, July 1989.

[7] R.J. Vaccaro. Weighted subspace fitting using subspace perturbation expansions. Technical Report ESAT-SISTA/TR 1997-45, Katholieke Universiteit Leuven, Dept. of Elec. Eng., *ftp.esat.kuleuven.ac.be* in pub/SISTA/guests/vaccaro.ps.Z, June 1997.

[8] M. Viberg and B. Ottersten. Sensor array processing based on subspace fitting. *IEEE Transactions on Signal Processing*, 39:1110–1121, May 1991.