

Weighted and Unweighted Subspace Fitting without Eigendecomposition

Luc Deneire* Jaouhar Ayadi Dirk T.M. Slock

Institut EURECOM, 2229 route des Crêtes, B.P. 193,
F-06904 Sophia Antipolis Cedex, FRANCE
{deneire, ayadi, slock}@eurecom.fr

Abstract

Subspace fitting has become a well known method to identify FIR Single Input Multiple Output (SIMO) systems, only resorting to second-order statistics. The main drawback of this method is its computational cost, due to the eigendecomposition of the sample covariance matrix. We propose a scheme that solves the subspace fitting problem without using the eigendecomposition of the cited matrix. The approach is based on the observation that the signal subspace is also the column space of the noise-free covariance matrix. We suggest a two-step procedure. In the first step, the column space is generated by arbitrary combinations of the columns. In the second step, this column space estimate is refined by optimally combining the columns using the channel estimate resulting from the first step. Using recent results on Weighted Subspace Fitting, we are able to incorporate the optimal weighting in the second step. For the unweighted signal subspace fitting, our method only requires computation of two eigenvectors of a small matrix and of two projection matrices, although yielding the same performance as the usual subspace fitting. Furthermore, a detailed analysis of the displacement structure of the various matrices involved leads to a fast algorithm. Furthermore, detailing the expressions of the criteria leads us to draw equivalences between various blind identifications schemes, namely SRM (Subchannel Response Matching), DML (Deterministic Maximum Likelihood) and linear prediction approaches.

1 Introduction

Subspace fitting algorithms have been applied to the multi-channel identification problem. In [11], it was shown that oversampled and/or multiple antenna received signals may be modeled as low rank processes and thus lend themselves to subspace methods, exploiting the or-

thogonality property between the noise subspace of the covariance matrix and the convolution matrix of the channel. Recent papers [1, 9, 7, 13] provide performance analysis of these methods. The huge majority of algorithms recently proposed to perform subspace fitting resort to SVD (singular value decomposition), which make them of little use for real-time implementations.

Recently, using the circularity property of the noise in a real symbol constellation based communication system, Kristensson, Ottersten and Slock proposed in [8] an alternative subspace fitting algorithm. In this paper, we show that this method can be used in the general case, leading to a consistent estimate, and that the performance is similar to that of the usual subspace fitting algorithms. Our method does not require the eigendecomposition of the covariance matrix. Nevertheless, as in the usual subspace fitting method, the computation of a projection matrix is required which may remain computationally demanding. A study of the displacement rank of the matrices involved, and the use of fast convolution algorithms, leads us to develop a fast algorithm.

In this paper, we consider the channel identification problem, but the ideas presented here apply to any subspace fitting problem.

2 Data Model

We consider a communication system with one emitter and a receiver consisting of an array of M antennas. The received signals are oversampled by a factor m w.r.t. the symbol rate. We furthermore consider linear digital modulation over a linear channel with additive noise, so that the received signal $\mathbf{y}(t) = [y_1(t) \dots y_M(t)]^T$ has the following form

$$\mathbf{y}(t) = \sum_k \mathbf{h}(t - kT) a(k) + \mathbf{v}(t)$$

*The work of Luc Deneire is supported by the EC by a Marie-Curie Fellowship (TMR program) under contract No ERBFMBICT950155

where $a(k)$ are the transmitted symbols, T is the symbol period and $\mathbf{h}(t) = [h_1(t) \dots h_M(t)]^T$ is the channel impulse response. The channel is assumed to be FIR with duration NT . If the received signals are oversampled at the rate $\frac{m}{T}$, the discrete input-output relationship can be written as :

$$\mathbf{y}(k) = \sum_{i=0}^{N-1} \mathbf{h}(i)a(k-i) + \mathbf{v}(k) = \mathbf{h}\mathbf{A}_N(k) + \mathbf{v}(k)$$

where

$$\begin{aligned} \mathbf{y}(k) &= [\mathbf{y}(kT)^H \mathbf{y}(kT + \frac{1}{m}T)^H \dots \mathbf{y}(kT + \frac{m-1}{m}T)^H]^H, \\ \mathbf{h}(k) &= [\mathbf{h}(kT)^H \mathbf{h}(kT + \frac{1}{m}T)^H \dots \mathbf{h}(kT + \frac{m-1}{m}T)^H]^H, \\ \mathbf{v}(k) &= [\mathbf{v}(kT)^H \mathbf{v}(kT + \frac{1}{m}T)^H \dots \mathbf{v}(kT + \frac{m-1}{m}T)^H]^H, \\ \mathbf{h} &= [\mathbf{h}(0), \dots, \mathbf{h}(N-1)] \text{ and} \end{aligned}$$

$\mathbf{A}_N(k) = [a(kT) \dots a((k-N+1)T)]$, superscript H denoting conjugate transpose. So we get a SIMO system with Mm channels. We consider additive temporally and spatially white Gaussian circular noise $\mathbf{v}(k)$ with $r_{vv}(k-i) = \mathbb{E}\{\mathbf{v}(k)\mathbf{v}(i)^H\} = \sigma_v^2 I_{Mm} \delta_{ki}$. Assume we receive L samples :

$$\mathbf{Y}_L(k) = \mathcal{T}_L(\mathbf{h})\mathbf{A}_{L+N-1}(k) + \mathbf{V}_L(k)$$

where $\mathcal{T}_L(\mathbf{h})$ is the convolution matrix of \mathbf{h} , $\mathbf{Y}_L(k) = [\mathbf{y}^H(k) \dots \mathbf{y}^H(k-L+1)]^H$ and similarly for $\mathbf{V}_L(k)$. In an obvious shorthand notation, we will use the following expression :

$$\mathbf{Y} = \mathcal{H}\mathbf{A} + \mathbf{V}.$$

We assume that $mML > L + N - 1$, so that the convolution matrix \mathcal{H} is ‘‘tall’’ and we assume \mathcal{H} to have full column rank (which leads to the usual identifiability conditions [10, 12]). The covariance matrix of \mathbf{Y} is

$$R_{YY} = \mathbb{E}\{\mathbf{Y}\mathbf{Y}^H\} = \mathcal{H}R_{AA}\mathcal{H}^H + \sigma_v^2 I.$$

3 Signal Subspace Fitting (SSF)

3.1 Classical SSF

One can write the eigendecomposition of the covariance matrix $R_{YY} = \mathbb{E}\{\mathbf{Y}\mathbf{Y}^H\} = V_S \Lambda_S V_S^H + V_N \Lambda_N V_N^H$ in which V_S has the same dimensions as \mathcal{H} and $\Lambda_N = \sigma_v^2 I$. The signal subspace can be expressed as:

$$\text{range}\{V_S\} = \text{range}\{\mathcal{H}\}$$

We can then formulate the classical subspace fitting problem :

$$\min_{\mathbf{h}, Q} \|\mathcal{H} - V_S Q\|_F^2$$

Since V_N spans the noise subspace, this leads to

$$\min_{\mathbf{h}} \mathbf{h}^t \left[\sum_{i=D^\perp}^{LMm} \mathcal{T}_N(V_i^{Ht}) \mathcal{T}_N^H(V_i^{Ht}) \right] \mathbf{h}^{Ht}$$

where V_i is column i of $V = [V_S V_N]$, $D^\perp = N + L$ and superscript t denotes the transposition of the blocks of a

block matrix. Under the constraint $\|\mathbf{h}\| = 1$, $\hat{\mathbf{h}}^{Ht}$ is then the eigenvector corresponding to the minimum eigenvalue of the matrix between the brackets. One can lower the computational burden by using $D^\perp > N + L$, loosing some performance (see a.o. [10],[12]).

Obviously, the projection on the noise subspace satisfies :

$$\mathbf{P}_{V_N} = \mathbf{P}_{V_S}^\perp = I - \mathbf{P}_{V_S} = I - V_S(V_S^H V_S)^{-1} V_S^H$$

which leads to the equivalent maximization :

$$\max_{\mathbf{h}} \mathbf{h}^t \left[\sum_{i=1}^{D^\perp - 1} \mathcal{T}_N(V_i^{Ht}) \mathcal{T}_N^H(V_i^{Ht}) \right] \mathbf{h}^{Ht}$$

3.2 Alternative Signal Subspace Fitting

The method In the absence of noise, we have :

$$R_{YY} = R = \mathcal{H}R_{AA}\mathcal{H}^H = V_S \Lambda_S' V_S^H + V_N \Lambda_N' V_N^H$$

where $\Lambda_S' = \Lambda_S - \sigma_v^2 I$ and $\Lambda_N' = 0$. From this expression, we observe that the column spaces of \mathcal{H} and R are the same, leading us to introduce the following subspace fitting criterion :

$$\min_{\mathbf{h}, Q} \|\mathcal{H} - \hat{R}BQ\|_F^2 \quad (1)$$

where $\|\cdot\|_F$ denotes Frobenius norm and \hat{R} is a consistent estimate of R . The matrix B has the same dimensions as \mathcal{H} and is fixed; we will see later how its choice influences the performance. Note that the range of $F = \hat{R}B$ provides an estimate for the signal subspace. We can take \hat{R} as $\hat{R} = \hat{R}_{YY} - \hat{\sigma}_v^2 I = \hat{R}_{YY} - \lambda_{\min}(\hat{R}_{YY})I$ where $\lambda_{\min}(\cdot)$ denotes the minimum eigenvalue (a rank revealing decomposition of $\hat{R}_{YY} - \lambda_{\min}(\hat{R}_{YY})I$ would lead to a better estimate of R). We note that the simulations below show that even simply $\hat{R} = \hat{R}_{YY}$ can work well also. The criterion (1) is separable in \mathbf{h} and Q . Minimizing w.r.t. Q first yields

$$Q = (F^H F)^{-1} F^H \mathcal{H}.$$

Substitution in (1) yields :

$$\min_{\mathbf{h}} \|\mathbf{P}_F^\perp \mathcal{H}\|_F^2 = \min_{\mathbf{h}} \text{trace} \{ \mathcal{H}^H \mathbf{P}_F^\perp \mathcal{H} \}.$$

With the constraint $\|\mathbf{h}\| = 1$, we get:

$$\begin{aligned} \hat{\mathbf{h}} &= \arg \min_{\|\mathbf{h}\|=1} \text{trace} \{ \mathcal{H}^H \mathbf{P}_F^\perp \mathcal{H} \} \\ &= \arg \min_{\|\mathbf{h}\|=1} \mathbf{h}^t \underbrace{\mathcal{B}_N^H(\mathcal{P}_F^\perp) \mathcal{B}_N(\mathcal{P}_F^\perp)}_{=\mathcal{F}} \mathbf{h}^{Ht} \end{aligned}$$

where $\mathcal{P}_F^\perp = [\mathbf{P}_F^\perp(0), \dots, \mathbf{P}_F^\perp(L-1)]$, $\mathbf{P}_F^\perp(\tau)$ are $MmL \times Mm$ matrices forming \mathbf{P}_F^\perp , and $\mathcal{B}_N(\mathbf{P})$ is a block Toeplitz

matrix with first block column $\begin{bmatrix} \mathbf{P} \\ 0 \end{bmatrix}$. The solution is thus $V_{max}(\mathcal{F})$, the eigenvector of \mathcal{F} corresponding to $\lambda_{max}(\mathcal{F})$. Given that $R = \mathcal{H}R_{AA}\mathcal{H}^H$, not every choice for B is acceptable. For instance, if the columns of B are in the noise subspace, then $F = 0$ for $\hat{R} = R$. Intuitively, the best choice for B should be $B = \mathcal{H}$, which corresponds to matched filtering \mathcal{H}^H with \mathcal{H} (post-multiplication of B with a square non-singular matrix does not change anything since that matrix can be absorbed in Q). These considerations lead to the following *two-step procedure*:

step 1 at first, B is chosen to be a fairly arbitrary selection matrix. The first step yields a consistent channel estimate (if $\mathcal{H}^H B$ is non-singular).

step 2 in this step, the consistent channel estimate of the first step is used to form $\hat{\mathcal{H}}$ and we solve (1) again, but now with $B = \hat{\mathcal{H}}$.

For the first step, for instance the choice $B^H = [I \ 0]$ leads to something that is quite closely related to the ‘‘rectangular Pisarenko’’ method of Fuchs [5]. We found however that a B of the same block Toeplitz form as \mathcal{H} but filled with a randomly generated channel works fairly well (this choice will be the one used in the simulations).

3.3 Asymptotics: exact estimation

Asymptotically, $\hat{R} = R$. We get $F = RB = \mathcal{H}R_{AA}\mathcal{H}^H B$. Assuming $R_{AA} > 0$, then if $\mathcal{H}^H B$ is non-singular, we get

$$\mathbf{P}_F = \mathbf{P}_{\mathcal{H}} = \mathbf{P}_{V_S}. \quad (2)$$

If furthermore we have a consistent channel estimate, then we can take asymptotically $B = \mathcal{H}$. In that case, the use of $R = R_{YY}$ and hence $F = R_{YY}\mathcal{H}$ also leads to (2). Pursuing this issue further, and applying a perturbation analysis similar to the one hereunder, we will have a consistent estimate of the channel with $\hat{R} = \hat{R}_{YY}$ as $\text{SNR} \rightarrow \infty$.

3.4 Perturbation analysis

For the first step of the algorithm, we get a consistent channel estimate if $\mathcal{H}^H B$ is non-singular. We can furthermore pursue the following asymptotic (first order perturbation) analysis. This analysis is based on the perturbation analysis of a projection matrix. If $\hat{F} = F + \Delta F$, then up to first order in ΔF , $\mathbf{P}_{\hat{F}} = \mathbf{P}_F + \Delta \mathbf{P}_F$ where

$$\Delta \mathbf{P}_F = 2\text{Sym}(\mathbf{P}_F^\perp \Delta F (F^H F)^{-1} F^H) \quad (3)$$

where $2\text{Sym}(\mathbf{X}) = \mathbf{X} + \mathbf{X}^H$.

Optimality of $B = \mathcal{H}$ Let $\hat{R} = R + \Delta R$, then using the eigendecomposition of R , we get up to first order

$$\Delta R = \Delta V_S \Lambda_S' V_S^H + V_S \Delta \Lambda_S' V_S^H + V_S \Lambda_S' \Delta V_S^H + \Delta V_N \Lambda_N' V_N^H + V_N \Delta \Lambda_N' V_N^H + V_N \Lambda_N' \Delta V_N^H$$

Let $B = V_S B_S + V_N B_N$ where we assume B_S non-singular. Then using $\hat{F} = \hat{R}B$ leads to

$$\Delta \mathbf{P}_F = 2\text{Sym}\left(\mathbf{P}_{V_N} \left[\Delta V_S \Lambda_S' + V_N \Delta \Lambda_N' B_N B_S^{-1} \right] (\mathcal{H}^H V_S)^{-1} R_{AA}^{-1} (\mathcal{H}^H \mathcal{H})^{-1} \mathcal{H}^H\right)$$

This shows that $B_N = 0$ is optimal.

Asymptotic equivalence of the two SSF Using $\hat{F} = (R_{YY} - \lambda_{min}(\hat{R}_{YY})I)\hat{\mathcal{H}}$ with \hat{R}_{YY} and $\hat{\mathcal{H}}$ consistent estimates, one can show that $\Delta \mathbf{P}_F$ is the same as with $\hat{F} = \hat{V}_S$. Hence we get up to first order

$$\mathbf{P}_{(\hat{R}_{YY} - \lambda_{min}(\hat{R}_{YY})I)\hat{\mathcal{H}}} = \mathbf{P}_{(\hat{R}_{YY} - \sigma_v^2 I)\mathcal{H}} = \mathbf{P}_{\hat{V}_S}.$$

This shows that the alternative signal subspace fitting method gives asymptotically exactly the same performance as the original SSF method. Furthermore, as long as consistent estimates are used for σ_v^2 and \mathcal{H} , the corresponding estimation errors have no influence up to first order.

Simplified method When we use simply $\hat{F} = \hat{R}_{YY}\hat{\mathcal{H}}$, then we get

$$\Delta \mathbf{P}_F = 2\text{Sym}(\mathbf{P}_{V_N} (\Delta V_S + \Delta \mathcal{H} (V_S^H \mathcal{H})^{-1} \Lambda_N \Lambda_S^{-1}) V_S^H)$$

The use of \hat{R}_{YY} instead of \hat{R} leads to the appearance of the second term, the relative importance of which is proportional to $\Lambda_N \Lambda_S^{-1}$. Hence this term is negligible at high SNR.

3.5 Weighted SSF

Gorokhov has shown [1] that the optimal weighting matrix for the weighted signal subspace fitting, corresponding to:

$$\hat{\mathbf{h}} = \arg \min_{\|\mathbf{h}\|=1} \mathbf{h}^t \mathcal{B}_N^H (\mathbf{P}_F^\perp) W \mathcal{B}_N (\mathbf{P}_F^\perp) \mathbf{h}^{Ht}$$

is $W_o = \Sigma^\#$ where $\#$ denotes the pseudo-inverse and $\sigma_v^2 \Sigma$ is the covariance matrix of $\mathcal{B}_N (\mathbf{P}_{V_N}) \mathbf{h}^{Ht}$, which can be expressed in terms of \mathbf{P}_{V_N} , R_{AA} and \mathbf{h} . Moreover, Σ can be replaced by its consistent estimate $\hat{\Sigma}$ without any influence on the asymptotic accuracy. Following these facts, a straightforward manner of doing Weighted SSF is to compute the weighting matrix after the first step (as we have consistent estimates for \mathbf{h} and \mathbf{P}_{V_N} , we can do it) and take it into account in the second step.

3.6 Fast Algorithm

The computational efficiency of the new eigendecomposition free subspace fitting method still largely depends on how the Projection matrix $I - F(F^H F)^{-1}F$ is computed, other computation optimizations (e.g. fast convolutions) being identical to the classical algorithms. This can be done using fast algorithms for near-Toeplitz matrices (see [6] and related papers).

Lemma 1 *The displacement rank of $F = RB$ is at most equal to $3Mm$, i.e., $\text{rank}(\Delta_{\{Z^{Mm}, Z^1\}}F) = 3Mm$ where*

$$\Delta_{\{Z^{Mm}, Z^1\}}F = F - Z^{Mm}FZ^{1T}$$

where Z^i is a shift matrix of appropriate size, zero everywhere except for 1's on the i^{th} subdiagonal.

Proof: From the properties of Toeplitz and near-Toeplitz matrices, we know that the displacement rank of $F = RB$ is the displacement rank of $\mathcal{M} = \begin{bmatrix} I & B \\ R & 0 \end{bmatrix}$. From $\text{rank}(\Delta_{\{Z^{Mm}, Z^{Mm}\}}R) = 2Mm$ and $\text{rank}(\Delta_{\{Z^{Mm}, Z^1\}}B) = Mm$, it is easily seen that $\text{rank}(\Delta_{\{Z^{Mm} \oplus Z^{Mm}, Z^{Mm} \oplus Z^1\}}\mathcal{M}) \leq 3Mm$ where $A \oplus B = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$, which completes the proof. \diamond

Lemma 2 *The displacement rank of \mathbf{P}_F^\perp is at most equal to $6Mm$.*

Constructing the matrix $\mathcal{M} = \begin{bmatrix} F^H F & F^H \\ F & I \end{bmatrix}$, it is easily seen that $\text{rank}(\Delta_{\{Z^1 \oplus Z^{Mm}, Z^1 \oplus Z^{Mm}\}}\mathcal{M}) \leq 6Mm$. \mathbf{P}_F^\perp is the Schur complement of $F^H F$ in \mathcal{M} and thus has the same displacement rank with respect to $\{Z^{Mm}, Z^{Mm}\}$. \diamond

Following these lemmas, one can construct a fast algorithm [6] requiring $O((6Mm)^3 K \log^2(K))$ flops, where K is the size of \tilde{R} .

4 Noise Subspace Fitting

4.1 Noise Subspace Linear Parameterizations

In this section, we focus our attention on different Noise Subspace Linear Parameterizations (NSLP) in term of sub-channel blocking equalizers. Consider the case of two channels: $Mm = 2$. One can observe that for noise-free signals, we have $\mathbf{h}_2(z)y_1(k) - \mathbf{h}_1(z)y_2(k) = 0$, which can be written in a matrix form as $[\mathbf{h}_2(z) \quad -\mathbf{h}_1(z)] \mathbf{y}(k) = \mathbf{h}^{\perp\top}(z) \mathbf{y}(k) = 0$. The matrix $\mathbf{h}^{\perp\top}(z)$ is parameterized by the channel impulse response and satisfies $\mathbf{h}^{\perp\top}(z) \mathbf{h}(z) = 0$. The counterpart of this parameterization in the time

domain is $\mathcal{T}^H(\mathbf{h}^\perp)$ which spans the orthogonal complement of $\mathcal{T}(\mathbf{h})$ and satisfies $\mathcal{T}(\mathbf{h}^\perp)\mathcal{T}(\mathbf{h}) = 0$. For $m > 2$, blocking equalizers $\mathbf{h}^{\perp\top}(z)$ can be constructed by considering the channels in pairs. The choice of $\mathbf{h}^{\perp\top}(z)$ is far from unique. To begin with, the number of pairs to be considered, which is the number of rows in $\mathbf{h}^{\perp\top}(z)$, is not unique. The minimum number is $m-1$ whereas the maximum number is $\frac{m(m-1)}{2}$. We shall call $\mathbf{h}^{\perp\top}(z)$ balanced if $\text{trace}\{\mathbf{h}^{\perp\top}(z)\mathbf{h}^\perp(z)\} = \alpha \mathbf{h}^\top(z)\mathbf{h}(z)$ for some real scalar α and $\mathbf{h}^\top(z) = \mathbf{h}^H(1/z^*)$. People usually take the maximum number of rows, which corresponds to a balanced $\mathbf{h}^{\perp\top}(z)$. The minimum number of rows in $\mathbf{h}^{\perp\top}(z)$ to be balanced is m . We get for instance

$$\begin{aligned} \mathbf{h}_{min}^{\perp\top}(z) &= \begin{bmatrix} -\mathbf{h}_2(z) & \mathbf{h}_1(z) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -\mathbf{h}_m(z) & 0 & \cdots & \mathbf{h}_1(z) \end{bmatrix} \\ \mathbf{h}_{bal}^{\perp\top}(z) &= \begin{bmatrix} -\mathbf{h}_2(z) & \mathbf{h}_1(z) & 0 & \cdots & 0 \\ 0 & -\mathbf{h}_3(z) & \mathbf{h}_2(z) & \cdots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ \mathbf{h}_m(z) & 0 & \cdots & 0 & -\mathbf{h}_1(z) \end{bmatrix} \end{aligned} \quad (4)$$

Continuing with this $\mathbf{h}_{bal}^{\perp\top}(z)$, its i^{th} row can be written as

$$\begin{aligned} \mathbf{h}_{bal,i}^{\perp\top}(z) &= \mathbf{h}^T(z) \mathcal{P}_i, \mathcal{P}_i = \mathcal{C} \mathcal{P}_{i-1} \mathcal{C}^H \\ \mathcal{P}_1 &= \begin{bmatrix} 0 & 1 & 0 & \cdots \\ -1 & 0 & \cdots \\ 0 & \vdots & \ddots \\ \vdots & & & \end{bmatrix} \\ \mathcal{C} &= \begin{bmatrix} 0 & \cdots & 0 & 1 \\ 1 & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & 0 & 1 & 0 \end{bmatrix} \end{aligned} \quad (5)$$

4.2 Two Blind Channel Estimation Methods Using NSLP

Subchannel Response Matching (SRM) Since for the noise-free signal we get $\mathcal{T}(\mathbf{h}^\perp)\mathbf{Y} = 0$, the SRM method minimizes the criterion $\|\mathcal{T}(\mathbf{h}^\perp)\mathbf{Y}\|_2^2$. The counterpart formulation of this criterion in the z domain uses $\mathbf{h}^\perp(z)$. The SRM method is always proposed using $\mathbf{h}_{max}^\perp(z)$. For $\mathbf{h}_{bal}^{\perp\top}(z)$, the SRM criterion $\|\mathcal{T}(\mathbf{h}^\perp)\mathbf{Y}\|_2^2$ can be written as the minimization w.r.t. \mathbf{h} of

$$\begin{aligned}
& \text{trace} \{ \mathcal{T}(\mathbf{h}^\perp) \mathbf{Y} \mathbf{Y}^H \mathcal{T}^H(\mathbf{h}^\perp) \} \\
&= \text{trace} \left\{ \mathbf{h}^\perp \left(\sum_{k=N-1}^{M-1} \mathbf{Y}_N(k) \mathbf{Y}_N^H(k) \right) \mathbf{h}^{\perp H} \right\} \quad (6) \\
&= (M-N+1) \text{trace} \{ \mathbf{h}^\perp \hat{R}_{YY} \mathbf{h}^{\perp H} \}
\end{aligned}$$

where the i^{th} row of \mathbf{h}^\perp is $\mathbf{h}_i^\perp = \mathbf{h}^T \mathcal{S}_i$, $\mathcal{S}_i = I_N \otimes \mathcal{P}_i$ and \otimes denotes Kronecker product. Hence the SRM criterion in (6) becomes

$$\min_{\mathbf{h}} \mathbf{h}^H B \mathbf{h}, \text{ where } B = \sum_{i=1}^m \mathcal{S}_i \hat{R}_{YY}^* \mathcal{S}_i^H. \quad (7)$$

It is expected that the use of a $\mathbf{h}_{bal}^{\perp \dagger}(z)$ with more rows leads to improved performance.

If the exact R_{YY} is used, then the noise contribution to the criterion (7) is $2\sigma_v^2 \|\mathbf{h}\|^2$ (and here the motivation for choosing a balanced $\mathbf{h}^\perp(z)$ becomes apparent). Hence the minimization of (7) subject to $\|\mathbf{h}\| = 1$ leads to the consistent SRM estimate $\mathbf{h} = V_{min}(B)$, at least if the channel length is chosen correctly. Since $\sigma_v^2 = \lambda_{min}(R_{YY})$, the minimum eigenvalue of R_{YY} , the noise contribution can be eliminated by replacing \hat{R}_{YY} by $\hat{R}_{YY} - \lambda_{min}(\hat{R}_{YY})I$ or, even better, by replacing B by $A = B - \lambda_{min}(B)I$ (the former choice doesn't make B singular with a finite amount of data). With this modification, the criterion in (7) becomes (asymptotically) insensitive to the noise contribution and any normalization of \mathbf{h} will lead to a consistent estimate.

Determination of $\mathbf{h}(z)$ from $\bar{\mathbf{P}}(z) = \mathbf{h}^{\perp H}(0)\mathbf{P}(z)$
This method uses a NLSP that is not linear in terms of the channel response ($\bar{\mathbf{P}}(z)$ proposed in [12]). We shall review the derivation of this technique. Let $\mathbf{P}(z) = \sum_{i=0}^L \mathbf{p}(i) z^{-i}$ with $\mathbf{p}(0) = I_m$ be the MMSE multivariate prediction error filter of order L for the noise-free received signal $\mathbf{y}(k)$. If $L \geq \underline{L} = \lceil \frac{N-1}{m-1} \rceil$, then it can be shown [12] that

$$\mathbf{P}(z)\mathbf{h}(z) = \mathbf{h}(0). \quad (8)$$

From (8) it is clear that $\mathbf{h}(z)$ and $\mathbf{P}(z)$, $\mathbf{h}(0)$ are equivalent parameterizations. Consider the full rank $m \times (m-1)$ matrix $\mathbf{h}^\perp(0)$ defined such that $\mathbf{h}^{\perp H}(0)\mathbf{h}(0) = 0$, then (8) implies that $\bar{\mathbf{P}}(z) = \mathbf{h}^{\perp H}(0)\mathbf{P}(z)$ is a $(m-1) \times m$ polynomial that satisfies

$$\bar{\mathbf{P}}(z)\mathbf{h}(z) = 0.$$

$\bar{\mathbf{P}}(z)$ or equivalently $\mathbf{P}(z)$ and $\mathbf{h}(0)$ can be estimated using linear prediction or Iterative Quadratic DML (IQDML). If $\bar{\mathbf{P}}(z)$ is estimated in a way that is robust to order overestimation, then the order of $\mathbf{h}(z)$ is known and $\mathbf{h}(z)$ can be estimated straightforwardly from $\bar{\mathbf{P}}(z)$. If not, then we can consider the following problem

$$\min_{\mathbf{h}} \frac{1}{2\pi j} \oint \mathbf{h}^\dagger(z) \bar{\mathbf{P}}^\dagger(z) \bar{\mathbf{P}}(z) \mathbf{h}(z) \frac{dz}{z} \quad (9)$$

Since $\mathbf{h}(z) = \mathbf{Q}(z)\mathbf{h} = \begin{bmatrix} I_m & \dots & z^{-(N-2)}I_m & z^{-(N-1)}I_m \end{bmatrix} \mathbf{h}$, the minimization problem given in (9) can be written as

$$\begin{aligned}
& \min_{\mathbf{h}} \frac{1}{2\pi j} \oint \mathbf{h}^H \mathbf{Q}^\dagger(z) \bar{\mathbf{P}}^\dagger(z) \bar{\mathbf{P}}(z) \mathbf{Q}(z) \mathbf{h} \frac{dz}{z} \\
&= \min_{\mathbf{h}} \mathbf{h}^H \left(\frac{1}{2\pi j} \oint \mathbf{Q}^\dagger(z) \bar{\mathbf{P}}^\dagger(z) \bar{\mathbf{P}}(z) \mathbf{Q}(z) \frac{dz}{z} \right) \mathbf{h} \quad (10)
\end{aligned}$$

which is again of the form $\min_{\mathbf{h}} \mathbf{h}^H A \mathbf{h}$ and hence, the solution is $V_{min}(A)$.

4.3 NSF Without Eigendecomposition

The classical NSF Approach A natural NSF problem is based on the eigendecomposition of the covariance matrix of the received signal which leads to signal and noise subspace contributions:

$$R_{YY} = E\mathbf{Y}\mathbf{Y}^H = V_S \Lambda_S V_S^H + V_N \Lambda_N V_N^H \quad (11)$$

Similarly to V_S which spans the signal subspace, V_N spans the noise subspace and $\mathcal{T}^H(\mathbf{h}^\perp)$ spans most of it. Hence, the following noise subspace fitting can be introduced:

$$\min_{\mathbf{h}, T} \left\| \mathcal{T}^H(\mathbf{h}^\perp) - V_N T \right\|_F \quad (12)$$

After optimization w.r.t. T , we obtain $\min_{\|\mathbf{h}\|_2=1}$ of

$$\text{trace} \{ \mathcal{T}(\mathbf{h}^\perp) \mathbf{P}_{V_N}^\perp \mathcal{T}^H(\mathbf{h}^\perp) \} = \mathbf{h}^H A \mathbf{h} \quad (13)$$

for some matrix A .

The Alternative NSF Approach An alternative NSF formulation that does not use the eigendecomposition of the covariance matrix is the following one:

$$\min_{\mathbf{h}} \|\mathcal{T}(\mathbf{h}^\perp) R^{\frac{1}{2}}\|_F^2 \quad (14)$$

where the matrix square root is of the form $R^{\frac{1}{2}} = V_S \Lambda_S^{\frac{1}{2}} Q$ for some unitary matrix Q ($\mathcal{T}(\mathbf{h}) R_{AA} \mathcal{T}^H(\mathbf{h}) = V_S \Lambda_S' V_S^H + V_N \Lambda_N' V_N^H$ with $\Lambda_S' = \Lambda_S - \sigma_v^2 I$ and $\Lambda_N' = 0$). Equivalence between classical NSF and eigendecomposition free NSF can be proved by a similar scheme as for the SSF. We established in [2] that the NSF problem without eigendecomposition given by (14) is nearly equivalent to the SRM problem apart from a weighting matrix.

Another equivalence is the following one

Theorem 1 Assume that $\mathbf{h}^{\perp \dagger}(z) = \bar{\mathbf{P}}(z)$, then the Eigendecomposition Free NSF (EFNSF) problem given by (14) is equivalent to the linear prediction criterion.

Proof:

$$\begin{aligned}
& \min_{\mathbf{h}} \|\mathcal{T}(\mathbf{h}^\perp) R^{\frac{1}{2}}\|^2 = \min_{\bar{P}} \|\mathcal{T}(\bar{P}) R^{\frac{1}{2}}\|^2 \\
&= \min_{\bar{P}} \text{trace} \left\{ \mathcal{T}(\bar{P}) R^{\frac{1}{2}} R^{\frac{H}{2}} \mathcal{T}^H(\bar{P}) \right\} \\
&= \min_{\bar{P}} \text{trace} \left\{ \mathcal{T}(\bar{P}) R \mathcal{T}^H(\bar{P}) \right\} \\
&= \min_{\bar{P}} \text{trace} \left\{ \bar{P} \left(\sum_{k=N-1}^{M-1} R(k) \right) \bar{P}^H \right\} \\
&= \min_{\bar{P}} (M - N + 1) \text{trace} \left\{ \bar{P} \hat{R}_{YY} \bar{P}^H \right\} \\
&= \min_{\bar{P}} (M - N + 1) \\
&\quad \text{trace} \left\{ \mathbf{h}^{\perp H}(0) P \hat{R}_{YY} P^H \mathbf{h}^\perp(0) \right\}
\end{aligned}$$

This coincides with the linear prediction normal equations given by

$$\begin{aligned}
& \min_{P(1:L)} [P(0:L)] \hat{R}_{YY} [P(0:L)]^H \Rightarrow \\
& \left\{ \begin{array}{l} [I \quad P(1:L)] \hat{R}_{YY} = \begin{bmatrix} \sigma_y^2 & 0 & \cdots & 0 \end{bmatrix} \\ \min_{\mathbf{h}(0)} \mathbf{h}^{\perp H}(0) \sigma_y^2 \mathbf{h}^\perp(0) \Rightarrow \mathbf{h}(0) = V_{max}(\sigma_y^2) \end{array} \right.
\end{aligned}$$

where $V_{max}(A)$ denotes the eigenvector corresponding to the maximum eigenvalue of A .

Theorem 2 For any choice of $\mathbf{h}^{\perp\dagger}(z)$, the optimal weighted EFNSF problem is equivalent to the Deterministic Maximum Likelihood (DML) parameterized by the same $\mathbf{h}^{\perp\dagger}(z)$.

Proof:

In the DML case, maximizing the likelihood reduces to $\min_{\mathbf{h}} \|\mathbf{Y} - \mathcal{T}(\mathbf{h})A\|^2$. This minimization problem is separable: for a fixed \mathbf{h} , the optimal transmitted symbol estimates are

$$A = (\mathcal{T}^H(\mathbf{h})\mathcal{T}(\mathbf{h}))^{-1} \mathcal{T}^H(\mathbf{h})\mathbf{Y}, \quad (15)$$

eliminating A in terms of \mathbf{h} , we get

$$\min_{\mathbf{h}} \mathbf{Y}^H P_{\mathcal{T}(\mathbf{h})}^\perp \mathbf{Y}. \quad (16)$$

Since $P_{\mathcal{T}(\mathbf{h})}^\perp \approx P_{\mathcal{T}^H(\mathbf{h}^\perp)}$. Hence, we get $\min_{\|\mathbf{h}\|=1}$ of

$$\begin{aligned}
& \mathbf{Y}^H P_{\mathcal{T}^H(\mathbf{h}^\perp)} \mathbf{Y} \\
&= \mathbf{Y}^H \mathcal{T}^H(\mathbf{h}^\perp) [\mathcal{T}(\mathbf{h}^\perp) \mathcal{T}^H(\mathbf{h}^\perp)]^\# \mathcal{T}(\mathbf{h}^\perp) \mathbf{Y} \\
&= \mathbf{h}^H (\mathcal{Y}^H [\mathcal{T}(\mathbf{h}^\perp) \mathcal{T}^H(\mathbf{h}^\perp)]^\# \mathcal{Y}) \mathbf{h} \\
&= \mathbf{h}^H A \mathbf{h}
\end{aligned} \quad (17)$$

where $\mathcal{T}(\mathbf{h}^\perp) \mathbf{Y} = \mathcal{Y} \mathbf{h}$ for some \mathcal{Y} .

Consider now the NSF problem given by (14)

$$\begin{aligned}
& \min_{\mathbf{h}} \|\mathcal{T}(\mathbf{h}^\perp) \hat{R}^{\frac{1}{2}}\|^2 \\
&= \min_{\mathbf{h}} \text{trace} \left\{ \mathcal{T}(\mathbf{h}^\perp) \hat{R} \mathcal{T}^H(\mathbf{h}^\perp) \right\} \\
&= \min_{\mathbf{h}} \text{trace} \left\{ \mathcal{T}(\mathbf{h}^\perp) (\hat{R}_{YY} - \lambda_{\min}(\hat{R}_{YY})) \mathcal{T}^H(\mathbf{h}^\perp) \right\} \\
&= \min_{\mathbf{h}} \text{trace} \left\{ \mathbf{h}^\perp \left(\sum_{k=N-1}^{M-1} w_k \mathbf{Y}_N(k) \mathbf{Y}_N^H(k) \right) \mathbf{h}^{\perp H} \right. \\
&\quad \left. - \alpha \lambda_{\min}(\hat{R}_{YY}) \|\mathbf{h}\|^2 \right\} \\
&= \min_{\mathbf{h}} \text{trace} \left\{ \mathcal{T}(\mathbf{h}^\perp) W \mathbf{Y} \mathbf{Y}^H \mathcal{T}^H(\mathbf{h}^\perp) \right\} \\
&= \min_{\mathbf{h}} \|\mathcal{T}(\mathbf{h}^\perp) \mathbf{Y}\|_W^2
\end{aligned}$$

Since the noise in $\mathcal{T}(\mathbf{h}^\perp) \mathbf{Y}$ is $\mathcal{T}(\mathbf{h}^\perp) \mathbf{V}$ with covariance matrix $\sigma_v^2 \mathcal{T}(\mathbf{h}^\perp) \mathcal{T}^H(\mathbf{h}^\perp)$; the optimally weighted LS problem corresponding to the criterion $\min_{\mathbf{h}} \|\mathcal{T}(\mathbf{h}^\perp) \mathbf{Y}\|_W^2$ is given by

$$\min_{\mathbf{h}} \mathbf{Y}^H \mathcal{T}^H(\mathbf{h}^\perp) \left[\mathcal{T}(\mathbf{h}^\perp) \mathcal{T}^H(\mathbf{h}^\perp) \right]^\# \mathcal{T}(\mathbf{h}^\perp) \mathbf{Y}$$

This minimization problem coincides with the DML criterion given by (17).

Corollary 1 The DML criterion parameterized by $\bar{\mathbf{P}}(z)$ is equivalent to an optimally weighted linear prediction problem.

Corollary 2 The optimal weighted NSF problem given by (14) is asymptotically insensitive to the used noise subspace parameterization.

5 Equivalences

Following the previous section, we can draw the following equivalences where $=$ denotes algebraic equivalence and \sim denotes asymptotical equivalence. Note that equivalences for WLP depend on the specific parameterization.

	WSSF	WEFSSF	WNSF	WEFNSF	DML	WSRM	WLP
WSSF	=	\sim	\sim	\sim	\sim	\sim	
WEFSSF		=	\sim	\sim	\sim	\sim	
WNSF			=	\sim	\sim	\sim	\sim
WEFNSF				=	=	=	\sim
DML					=	=	\sim
WSRM						=	\sim
WLP							=

- WSSF : Weighted Signal Subspace Fitting
- WEFSSF: Weighted Eigendecomposition Free SSF
- WNSF : Weighted Noise Subspace Fitting
- WEFNSF: Weighted Eigendecomposition Free NSF
- DML : Deterministic Maximum Likelihood
- WSRM : Weighted Subchannel Response Matching
- WLP : Weighted Linear Prediction

6 Simulation Results

6.1 Noise Subspace Fitting

In order to compare the performance of the NSF without eigendecomposition to the one of the natural approach given by (13), the considered performance measure is the Normalized MSE (NMSE) which is computed over 300 Monte Carlo runs as

$$\text{NMSE} = \frac{1}{300} \sum_{i=1}^{300} \mathbf{h}^H P_{\hat{\mathbf{h}}^{(i)}} \mathbf{h} / \|\mathbf{h}\|^2$$

where $\mathbf{h}^H P_{\hat{\mathbf{h}}} \mathbf{h} = \min_{\alpha} \|\alpha \hat{\mathbf{h}} - \mathbf{h}\|^2$. We use a randomly generated complex channel \mathbf{h} with $N = 3$, $M = 3$ and $m = 1$. The symbols are i.i.d. BPSK, and the data length is $L = 210$. The SNR is defined as $(\|\mathbf{h}\|^2 \sigma_a^2) / (mM \sigma_v^2)$. In Figure 1, we compare the SRM performance (since SRM and NSF without eigendecomposition are equivalent criteria) to the NSF one. The two corresponding curves are close which confirms that NSF and NSF without eigendecomposition lead to the same asymptotic performance.

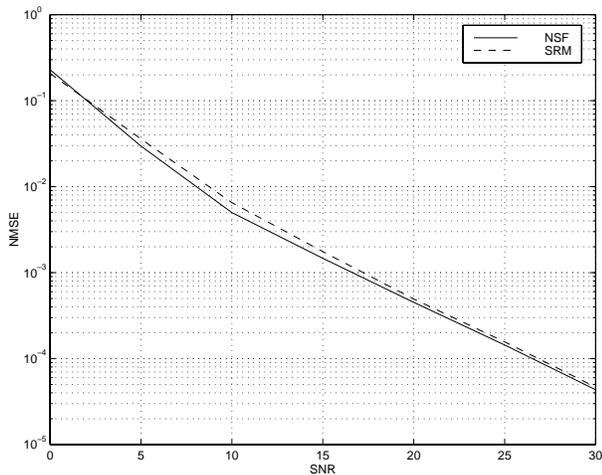


Figure 1: Performance of NSF and NSF without eigendecomposition (SRM)

6.2 Signal Subspace Fitting

In these simulations, we use a randomly generated real channel of length $6T$, an oversampling factor of $m = 1$ and $M = 3$ antennas. We draw the NRMSE of the channel estimate.

The correlation matrix is calculated from a burst of 100 QAM-4 symbols. For these simulations, we used 100 Monte-Carlo runs.

We draw the NRMSE for the first step of the algorithm, the second step and the subspace fitting with eigendecomposition.

These curves show that the proposed algorithm yields the same performance as the subspace fitting algorithm

with eigendecomposition (even slightly better at low SNR, but this is not relevant). It is to note that we made the simulation using \hat{R}_{YY} and $\hat{R}_{YY} - \lambda_{\min}(\hat{R}_{YY})I$, which gives the same performance.

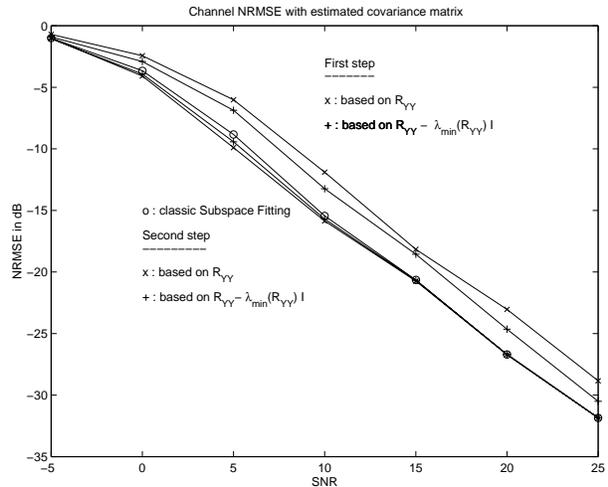


Figure 2: Subspace Fitting performance

Furthermore, we also include the NRMSE of channel estimate when using a perfect estimated covariance for the two steps. Comparison of the two graphs illustrates the preponderance of the covariance estimation error on the channel estimation error.

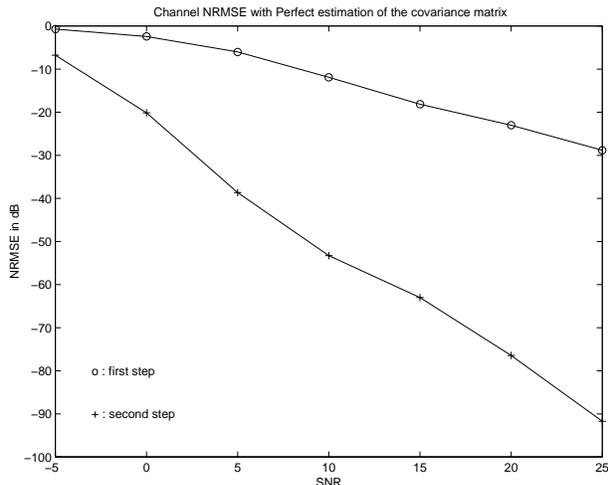


Figure 3: Subspace Fitting performance

7 Conclusions

We have proposed a new two-step algorithm for solving the signal subspace fitting problem, in the channel identification context, which is computationally less demanding than the usual algorithms. Perturbation analysis shows the asymptotic equivalence of the eigendecomposition-free approach to the original method. This equivalence was confirmed by simulation results. Fast algorithms for the calculation of the projection matrix have been explored.

The NSF without eigendecomposition introduced in [3] was analyzed in this paper. Two main results were established. The first concerns the case when we parameterize the noise subspace with $\bar{\mathbf{P}}(z)$, we proved that the minimization problem formulated is identical to the linear prediction criterion. The second result elaborates a weighted least squares approach to the NSF without eigendecomposition. In this context, we proved that the optimally weighted NSF problem parameterized by a specific $\mathbf{h}^{\perp\dagger}(z)$ coincides with the DML criterion formulated with this same parameterization. An immediate consequence of these two results is the following: since the NSF criterion parameterized with $\bar{\mathbf{P}}(z)$ is the linear prediction approach and since the optimally weighted NSF is DML, the DML criterion parameterized by $\bar{\mathbf{P}}(z)$ is a weighted version of the linear prediction approach. It was established in a previous work [2], that the DML optimization problem is asymptotically insensitive to the noise subspace parameterization used. According to Theorem 2, this result holds also to the optimally weighted NSF without eigendecomposition.

References

- [1] K. Abed-Meraim, J.-F. Cardoso, A. Y. Gorokhov, P. Loubaton, and E. Moulines. "On Subspace Methods for Blind Identification of Single-Input Multiple-Output FIR Systems". *IEEE Transactions on Signal Processing*, 45(1):42–55, January 1997.
- [2] Jaouhar Ayadi and Dirk T.M. Slock. "Cramer-Rao Bounds and Methods for Knowledge Based Estimation of Multiple FIR Channels.". In *Proc. 1st IEEE SP Workshop on Signal Processing Advances for Wireless Communications*. IEEE, April 1997.
- [3] Luc Deneire, Jaouhar Ayadi, and Dirk T.M. Slock. "Subspace Fitting Without Eigendecomposition". In "*13th international Conference on Digital Signal Processing*", Santorini, Greece, July 1997.
- [4] Luc Deneire and Dirk T.M. Slock. "Linear Prediction And Subspace Fitting Blind Channel Identification Based On Cyclic Statistics.". In "*13th international Conference on Digital Signal Processing*", Santorini, Greece, July 1997.
- [5] Jean-Jacques Fuchs. "Rectangular Pisarenko Method Applied to Source Localization". *IEEE Transactions on Signal Processing*, 44(10):2377–2383, October 1996.
- [6] J. Chun. "*Fast Array Algorithms for Structured Matrices*". PhD thesis, Information Systems Laboratory, Stanford University, June 1989.
- [7] Martin Kristensson and Björn Ottersten. "Statistical analysis of a subspace method for blind channel identification". In *Proc. ICASSP 96, Atlanta, Georgia, May 1996*, May 1996.
- [8] Martin Kristensson, Björn Ottersten, and Dirk Slock. "Blind Subspace Identification of a BPSK Communication Channel". In *Proc. Asilomar 96, Asilomar, USA*, November 1996.
- [9] B. Ottersten M. Kristensson. "Asymptotic comparison of two blind channel identification methods". In *Proc. 1st IEEE SP Workshop on Signal Processing Advances for Wireless Communications*, april 1997.
- [10] E. Moulines, P. Duhamel, J.-F. Cardoso, and S. Mayrargue. "Subspace Methods for the Blind Identification of Multichannel FIR filters". *IEEE Transactions on Signal Processing*, 43(2):516–525, February 1995.
- [11] D.T.M. Slock. "Blind Fractionally-Spaced Equalization, Perfect-Reconstruction Filter Banks and Multichannel Linear Prediction". In *Proc ICASSP 94 Conf.*, Adelaide, Australia, April 1994.
- [12] D.T.M. Slock and C.B. Papadias. "Blind Fractionally-Spaced Equalization Based on Cyclostationarity". In *Proc. Vehicular Technology Conf.*, pages 1286–1290, Stockholm, Sweden, June 1994.
- [13] Hanks H. Zeng and Lang Tong. "Blind Channel Estimation Using the Second-Order Statistics: Asymptotic Performance and Limitations". *submitted to IEEE Transactions on Signal Processing*, January 1997.