Analysis of smoothing techniques for subspace estimation with application to channel estimation

Pedersen, Niels Lovmand; Jakobsen, Morten Lomholt; Rom, Christian; Fleury, Bernard Henri

Published in:
I E E E International Conference on Communications

DOI (link to publication from Publisher):
10.1109/icc.2011.5962560

Publication date:
2011

Document Version
Accepted author manuscript, peer reviewed version

Analysis of Smoothing Techniques for Subspace Estimation with Application to Channel Estimation

Niels Lovmand Pedersen*, Morten Lomholt Jakobsen* Christian Rom† and Bernard Henri Fleury*
*Department of Electronic Systems, Aalborg University, Denmark
†Infineon Technologies Denmark A/S, Aalborg, Denmark

***Abstract*—In this paper, we present an investigation on the impact of spatial smoothing and forward-backward averaging techniques for subspace-based channel estimation. The spatial smoothing technique requires the selection of a window size, which, if not chosen properly, leads to dramatic performance breakdown of subspace-based methods. We provide an explanation of the performance drop for certain window sizes and subsequently an understanding of a proper window size selection. In particular, we describe the behavior of the magnitude of the least signal eigenvalue as a function of the used window size. Through simulations we show that the magnitude of this eigenvalue is of particular importance for estimating the signal subspace and the entailing performance of the channel estimator.**

I. Introduction

Subspace-based methods such as MUSIC [1] and ESPRIT [2] are commonly employed for the purpose of extracting unknown parameters from structured observation models. The unknown parameters of interest are estimated by exploiting properties of certain subspaces created via matrix factorization techniques, e.g. eigenvalue-decomposition of a sample covariance matrix. Accordingly, the estimation accuracy associated with the unknown parameters relies upon the "quality" of the subspaces involved. Preprocessing techniques, such as spatial smoothing (SS) and forward-backward averaging (FB) can be applied prior to the matrix factorization [3], [4]. This may trigger extraction of the unknown parameters with greater precision due to an improved representation of the parameter-revealing subspace. In practice, only a limited number of observations are available to compute a sample covariance matrix. By application of SS one can artificially generate additional observations at the cost of a reduction of the matrix dimensions. This trade-off is dictated by the window size which needs to be specified by the designer. The change in the original matrix dimensions is to some extend harmless as long as the parameter-revealing properties are sustained.

Preprocessing techniques have been used in various applications, e.g. in *direction-of-arrival* (DOA) estimation [5] and in enhanced propagation delay estimation [6], [7] for decorrelation of coherent sources. The above mentioned trade-off on the selected window size is commonly determined based on simulations, see e.g. [7] and the references therein. Subspace-based methods and preprocessing techniques have also been applied for *orthogonal frequency-division multiplexing* (OFDM) pilot-aided channel estimation [8], [9]. As shown by [9], selecting a too large (or small) window size relative to the available observation window leads to a severe drop in performance of the subspaced-based channel estimator.

A well-known observation from the DOA literature is that one should select the window size to be approximately half the available observation window. Furthermore, in [10] the performance breakdown of subspace-based methods is investigated when the signal-to-noise ratio (SNR) falls below a threshold SNR. However, to the authors' best knowledge, no comprehensive explanation for this choice of window size is available nor a proper understanding of the performance drop for some selected window sizes. In this paper, we aim at providing such an understanding. To do so, we decouple the compound impact of SS and FB into three distinct effects. From this decoupling, we indirectly explore the impact of SS and FB on the performance of channel estimation by investigating how these techniques affect the underlying subspace estimation. This approach has the advantage of being general in the sense that it can be conducted without any particular channel estimator in mind. In a next step, we infer how SS and FB affects the performance of a particular channel estimator. We consider an OFDM system with the channel estimation performed as in [8] and [9] operating in a multipath environment.

II. System Description

A. OFDM Signal Model

We consider a single-input single-output OFDM system with N subcarriers, where only $N_u \leq N$ of these are used for transmission. A cyclic prefix is added to preserve orthogonality between subcarriers and to eliminate inter-symbol interference between consecutive OFDM symbols. The channel is assumed static during the transmission of each OFDM symbol. In baseband representation the received OFDM signal in matrix-vector notation reads

$$\mathbf{r} = [r_1, r_2, \ldots, r_{N_u}]^T = \mathbf{X}\mathbf{h} + \mathbf{w}, \tag{1}$$

where $(\cdot)^T$ denotes the transpose operation. The diagonal matrix $\mathbf{X} = \operatorname{diag}\{x_1, x_2, \ldots, x_{N_u}\}$ is built from the transmitted symbols. The vector $\mathbf{h} = [h_1, h_2, \ldots, h_{N_u}]^T$ contains as components samples of the channel frequency response at the N_u active subcarriers. Samples of additive complex white Gaussian noise with variance σ^2 are contained in the vector $\mathbf{w} = [w_1, w_2, \ldots, w_{N_u}]^T$.

To estimate the vector $\mathbf{h}$ in (1), a total of M pilot symbols

are transmitted systematically across selected subcarriers with indices in the subset

$$\mathcal{P} := \{p(1), p(2), \ldots, p(M)\} \subset \{1, 2, \ldots, N_u\}. \quad (2)$$

The received symbols observed at the pilot positions are divided by the corresponding pilot symbols to produce the observations used to estimate the channel vector $\mathbf{h}$:

$$\mathbf{y} := (\mathbf{X}_{\mathcal{P}})^{-1} \mathbf{r}_{\mathcal{P}} = \mathbf{h}_{\mathcal{P}} + (\mathbf{X}_{\mathcal{P}})^{-1} \mathbf{w}_{\mathcal{P}}. \quad (3)$$

We assume that all pilot symbols hold unit power such that the statistics of the noise term $(\mathbf{X}_P)^{-1}\mathbf{w}_P$ remain unchanged compared to $\mathbf{w}$, i.e. $\mathbf{y}$ yields the samples of the true channel frequency response (at the pilot subcarriers) corrupted by additive complex white Gaussian noise with variance σ^2.

B. Multipath Channel Model

To estimate $\mathbf{h}$ we invoke a parametric model of the wireless channel. The task is thereby altered to the estimation of the parameters of the model instead of the samples of the channel frequency response at the $N_u - M$ subcarriers. The time-varying impulse response of the channel is modeled as a sum of multipath components:

$$g(t, \tau) = \sum_{l=1}^{L} \alpha_l(t) \delta(\tau - \tau_l). \quad (4)$$

In this expression, $\alpha_l(t)$ and τ_l are respectively the complex weight and the delay of the lth multipath component, while $\delta(\cdot)$ is the Dirac delta. The total number of multipath components L is assumed fixed. The delay parameters $\{\tau_l\}$ are also assumed persistently static. The weights $\{\alpha_l(t)\}$ are mutually uncorrelated wide-sense stationary, zero-mean proper complex Gaussian processes with their power normalized such that $\sum_{l=1}^{L} \mathbb{E}\left[|\alpha_l(t)|^2\right] = 1$. Thus, the channel described by (4) is a *wide-sense stationary and uncorrelated scattering* [11] (WSSUS) Rayleigh fading channel. Additional details regarding the assumptions on the channel model are provided in Section VI.

III. Subspace Decomposition

Taking the parametric model (4) of the channel into account, we reformulate (3) as

$$\mathbf{y} = \mathbf{T}\boldsymbol{\alpha} + \mathbf{n}, \quad (5)$$

where $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_L]^T$ and $\mathbf{T}$ is an $M \times L$ matrix depending on the known pilot positions $\mathcal{P}$ as well as the unknown delay parameters $\{\tau_l\}$. Specifically, the (m, l)th entry of $\mathbf{T}$ reads

$$\mathbf{T}_{m,l} := \exp\left(-j2\pi \frac{p(m)}{N} \tau_l\right) \quad , \begin{matrix} m = 1, 2, \ldots, M \\ l = 1, 2, \ldots, L. \end{matrix} \quad (6)$$

The vector $\mathbf{y}$ in (5) is proper complex Gaussian distributed with zero-mean and covariance matrix

$$\mathbf{R} := \mathbb{E}\left[\mathbf{y}\mathbf{y}^H\right] = \mathbf{T}\mathbf{A}\mathbf{T}^H + \sigma^2 \mathbf{I}_M, \quad (7)$$

where $(\cdot)^H$ denotes the conjugate transpose operation and $\mathbf{I}_M$ is the $M \times M$ identity matrix. In writing (7) we have assumed that $\boldsymbol{\alpha}$ and $\mathbf{n}$ are statistically independent, and due to the uncorrelated scattering assumption, $\mathbf{A} := \mathbb{E}\left[\boldsymbol{\alpha}\boldsymbol{\alpha}^H\right]$ is an $L \times L$ diagonal matrix. It is crucial to realize that both matrices $\mathbf{A}$ and $\mathbf{R}$ are theoretical quantities which are not available in practice. These matrices can be estimated only if certain ergodic properties are satisfied and still it would require an observation window of extensive duration. In practice we are limited to work with finite sample sizes and observations are usually collected during short periods of time. Accordingly, we have to be careful when applying algorithms which are based on a theoretical quantity such as $\mathbf{R}$ or its associated eigen-decomposition.

The M eigenvalues of $\mathbf{R}$ can be arranged in decreasing order as [12, sec. 4.5]

$$\lambda_m = \begin{cases} \mu_m + \sigma^2 & , \quad m = 1, 2, \ldots, L \\ \sigma^2 & , \quad m = L+1, \ldots, M, \end{cases} \quad (8)$$

where $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_L$ are the L strictly positive eigenvalues of the matrix $\mathbf{B} := \mathbf{T}\mathbf{A}\mathbf{T}^H$. The subspace spanned by the L eigenvectors of $\mathbf{R}$ associated with $\lambda_1, \ldots, \lambda_L$ is identical to the column space of $\mathbf{T}$ [12, sec. 4.5]. We refer to this L-dimensional subspace as the signal subspace, and its orthogonal complement as the noise subspace. That is, (8) allows for the orthonormal eigenvector basis of $\mathbf{R}$ to be split into two bases, one spanning the signal subspace and the other one spanning the noise subspace. Practical algorithms such as MUSIC and ESPRIT exploit the partly known structure of $\mathbf{T}$ to extract the unknown delay parameters $\{\tau_l\}$ from estimates of these two distinct subspaces. As will be argued this is possible since (8) may apply to matrices obtained from finite sample sizes.

IV. Preprocessing Techniques

Since the theoretical covariance matrix $\mathbf{R}$ is unobtainable in practice, we are compelled to acquire an estimate of it. Using the word "estimate" is in fact rather misleading in this case. We merely seek a matrix $\hat{\mathbf{R}} = \hat{\mathbf{U}}\tilde{\mathbf{\Lambda}}\hat{\mathbf{U}}^H$ such that the L eigenvectors in $\hat{\mathbf{U}}$ associated with the L largest eigenvalues form a basis for the column space of $\mathbf{T}$. To acquire such a matrix $\hat{\mathbf{R}}$, we collect K temporal observations $\mathbf{y}_1, \ldots, \mathbf{y}_K$ from (5) and store them in the $M \times K$ matrix

$$\mathbf{Y} := \begin{bmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \ldots & \mathbf{y}_K \end{bmatrix}. \quad (9)$$

From (9) we compute the sample covariance matrix

$$\hat{\mathbf{R}} := \frac{1}{K}\mathbf{Y}\mathbf{Y}^H = \mathbf{T}\tilde{\mathbf{A}}\mathbf{T}^H + \mathbf{E} \quad (10)$$

with

$$\tilde{\mathbf{A}} := \frac{1}{K}\sum_{k=1}^{K} \boldsymbol{\alpha}_k \boldsymbol{\alpha}_k^H \quad (11)$$

and with noise and cross-term contributions collected in

$$\mathbf{E} := \frac{1}{K}\sum_{k=1}^{K} \mathbf{n}_k \mathbf{n}_k^H + \frac{1}{K}\sum_{k=1}^{K} \left(\mathbf{T}\boldsymbol{\alpha}_k \mathbf{n}_k^H + \mathbf{n}_k \boldsymbol{\alpha}_k^H \mathbf{T}^H\right). \quad (12)$$

Throughout the paper our main focus is aimed at the matrix $\tilde{\mathbf{B}} := \mathbf{T}\tilde{\mathbf{A}}\mathbf{T}^H$. It is again crucial to realize that we do not consider the matrix $\tilde{\mathbf{A}}$ as a proper or direct estimate of $\mathbf{A}$, neither as $\tilde{\mathbf{B}}$ as an estimate of $\mathbf{B}$. The important thing is that $\tilde{\mathbf{A}}$ holds similar properties as $\mathbf{A}$, e.g. that it is nonsingular. The decomposition of $\hat{\mathbf{R}}$ into a signal and noise subspace as in (8) makes sense only when $\tilde{\mathbf{B}}$ has rank L, i.e. when $\tilde{\mathbf{A}}$ is nonsingular. However, $\tilde{\mathbf{A}}$ may easily happen to be singular, because the samples $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_K$ are usually correlated. When $K < L$ the matrix is indeed singular, e.g. with $K = 1$ the matrix $\tilde{\mathbf{A}} = \boldsymbol{\alpha}_1\boldsymbol{\alpha}_1^H$ has rank one (in fact, $\hat{\mathbf{R}} = \mathbf{y}_1\mathbf{y}_1^H$ only holds a single non-zero eigenvalue). The integer K should be chosen as small as possible to mitigate the effect of large-scale fluctuations of the channel response. So we cannot increase K arbitrarily to build up rank in $\tilde{\mathbf{A}}$. Another issue is the matrix term $\mathbf{E}$ in (10) which desirably (but loosely speaking) should be $\mathbf{E} \approx \sigma^2\mathbf{I}_M$. Preprocessing techniques are the key to achieve these goals. Any technique for doing so is of course only meaningful in this context if it leaves the properties of the parameter-revealing subspace unaltered. In the following we describe the SS and FB techniques, and discuss why they preserve the subspace properties.

A. Spatial Smoothing

The SS technique [3] applies a sliding window to the matrix $\mathbf{Y}$ in (9). We select the subset $\mathcal{P}$ in (2) such that the pilots are equally spaced[1] by a fixed amount Δp. Then $\mathcal{P}$ is divided into overlapping windows of size $M_1 \leq M$. The set of positions indexed by $\{p(1), p(2), \ldots, p(M_1)\}$ forms the first window, the set $\{p(2), p(3), \ldots, p(M_1+1)\}$ forms the second window and so on to a total of $\bar{M} := M - M_1 + 1$ windows. This procedure artificially builds up additional observations at the expense of lowering the observation bandwidth, i.e. the resolution in the delay domain. Let $\mathbf{y}_k^{(m)}$ denote the M_1 components of $\mathbf{y}_k$ corresponding to the mth window. Then, by exploiting the particular shift structure of $\mathbf{T}$, we can write

$$\mathbf{y}_k^{(m)} = \mathbf{T}_{M_1}\mathbf{D}^m\boldsymbol{\alpha}_k + \mathbf{n}_k^{(m)}, \quad m = 0, 1, \ldots, \bar{M}-1, \quad (13)$$

where $\mathbf{D} = \operatorname{diag}\{\exp(-j2\pi\Delta f\tau_1), \ldots, \exp(-j2\pi\Delta f\tau_L)\}$ with $\Delta f := \Delta p/N$. The matrix $\mathbf{T}_{M_1}$ is made of the first M_1 rows of $\mathbf{T}$, while $\mathbf{n}_k^{(m)}$ denotes the M_1 components of $\mathbf{n}_k$ corresponding to the mth window. The spatially smoothed sample covariance matrix is then defined as

$$\hat{\mathbf{R}}^{\text{ss}} := \frac{1}{K}\sum_{k=1}^{K}\frac{1}{\bar{M}}\sum_{m=0}^{\bar{M}-1}\mathbf{y}_k^{(m)}\left(\mathbf{y}_k^{(m)}\right)^H \in \mathbb{C}^{M_1\times M_1}. \quad (14)$$

Notice that $\hat{\mathbf{R}}^{\text{ss}}$ can be split in a similar way as the right-hand side of (10):

$$\hat{\mathbf{R}}^{\text{ss}} = \mathbf{T}_{M_1}\mathbf{A}^{\text{ss}}\mathbf{T}_{M_1}^H + \mathbf{E}^{\text{ss}} \quad (15)$$

with

$$\mathbf{A}^{\text{ss}} := \frac{1}{\bar{M}}\sum_{m=0}^{\bar{M}-1}\mathbf{D}^m\tilde{\mathbf{A}}\left(\mathbf{D}^m\right)^H. \quad (16)$$

[1]Meaning that $p(m) - p(m-1) = \Delta p$ for $m = 2, 3, \ldots, M$.

We illustrate the principle of the SS technique for $K = 1$ and $M_1 = M - 1$ by recasting $\mathbf{A}^{\text{ss}}$ as

$$\mathbf{A}^{\text{ss}} = \frac{1}{2}\Big[\boldsymbol{\alpha}_1 \;\; \mathbf{D}\boldsymbol{\alpha}_1\Big]\Big[\boldsymbol{\alpha}_1 \;\; \mathbf{D}\boldsymbol{\alpha}_1\Big]^H. \quad (17)$$

From (17) we observe that $\mathbf{A}^{\text{ss}}$ has rank equal to two whereas $\tilde{\mathbf{A}} = \boldsymbol{\alpha}_1\boldsymbol{\alpha}_1^H$ has rank one. More generally, by means of SS we aim at building up L linearly independent columns and hence, we must have $K\bar{M} \geq L$.

B. Forward-Backward Averaging

The FB technique (see e.g. [4]) is a well-known and simple method for increasing the rank without lowering the dimension of $\hat{\mathbf{R}}$. We perform SS together with FB (denoted FBSS) and define $\hat{\mathbf{R}}^{\text{fbss}}$ as

$$\hat{\mathbf{R}}^{\text{fbss}} := \frac{1}{2}\left(\hat{\mathbf{R}}^{\text{ss}} + \mathbf{J}\left(\hat{\mathbf{R}}^{\text{ss}}\right)^*\mathbf{J}\right) \in \mathbb{C}^{M_1\times M_1}. \quad (18)$$

Here, $(\cdot)^*$ denotes complex conjugation and $\mathbf{J}$ is the reversal matrix with 1's on its entire antidiagonal and 0's elsewhere. The matrix in (18) is persymmetric, i.e. $\mathbf{J}\hat{\mathbf{R}}^{\text{fbss}} = (\mathbf{J}\hat{\mathbf{R}}^{\text{fbss}})^T$.

From (18) and in analogy with (10) and (15) we write $\hat{\mathbf{R}}^{\text{fbss}}$ as

$$\hat{\mathbf{R}}^{\text{fbss}} = \mathbf{T}_{M_1}\mathbf{A}^{\text{fbss}}\mathbf{T}_{M_1}^H + \mathbf{E}^{\text{fbss}} \quad (19)$$

with

$$\mathbf{A}^{\text{fbss}} := \frac{1}{2}\left(\mathbf{A}^{\text{ss}} + \mathbf{Q}(\mathbf{A}^{\text{ss}})^*\mathbf{Q}^H\right). \quad (20)$$

The $L \times L$ diagonal matrix[2] $\mathbf{Q}$ is obtained from the identity $\mathbf{J}\mathbf{T}_{M_1}^* = \mathbf{T}_{M_1}\mathbf{Q}$. By jointly applying FB and SS we build up rank in $\mathbf{A}^{\text{fbss}}$ more rapidly than in $\mathbf{A}^{\text{ss}}$. Performing FB only may not be sufficient, because it can at most double the rank of a matrix. Notice that the two techniques can be applied in any order.

C. Discussion

It is meaningful to apply SS and FB if the properties of the parameter-revealing subspace are sustained. Hence, the subspace-based methods should still be able to extract the desired parameters. For SS a requirement is that $M_1 > L$, otherwise we cannot separate the eigenvalues into signal and noise eigenvalues as in (8). The FB technique does not change the signal subspace. To see this, we let

$$\mathbf{B}^{\text{ss}} := \mathbf{T}_{M_1}\mathbf{A}^{\text{ss}}\mathbf{T}_{M_1}^H \quad (21)$$

and

$$\begin{aligned}\mathbf{B}^{\text{fbss}} &:= \frac{1}{2}\left(\mathbf{B}^{\text{ss}} + \mathbf{J}\left(\mathbf{B}^{\text{ss}}\right)^*\mathbf{J}\right)\\ &= \mathbf{T}_{M_1}\frac{1}{2}\left(\mathbf{A}^{\text{ss}} + \mathbf{Q}(\mathbf{A}^{\text{ss}})^*\mathbf{Q}^H\right)\mathbf{T}_{M_1}^H. \end{aligned} \quad (22)$$

As long as $\mathbf{A}^{\text{ss}}$ is nonsingular, the columns of $\mathbf{B}^{\text{fbss}}$ and $\mathbf{B}^{\text{ss}}$ span the same L-dimensional signal subspace.

[2]More specifically, $\mathbf{Q}$ has diagonal entries $\mathbf{Q}_{l,l} = \exp\left(j2\pi\big(2p(1) + (M_1-1)\Delta p\big)\tau_l/N\right)$, $l = 1, \ldots, L$.

It can be shown analytically that the elements of the matrices in (11), (16) and (20) fulfill the following relations:

$$\left|\mathbf{A}^{\text{fbss}}_{l,l'}\right| \leq \left|\mathbf{A}^{\text{ss}}_{l,l'}\right| \leq \left|\tilde{\mathbf{A}}_{l,l'}\right| \quad , \quad l,l' = 1,\ldots,L \tag{23}$$

with equality in both relations if $l = l'$. In analogy with [5]–[7], we refer to the relationship (23) as the decorrelation effect inherited from the preprocessing.

Let μ^{ss}_L and μ^{fbss}_L denote the Lth eigenvalue of $\mathbf{B}^{\text{ss}}$ and $\mathbf{B}^{\text{fbss}}$ respectively. Then, the inequality

$$\mu^{\text{fbss}}_L \geq \mu^{\text{ss}}_L, \tag{24}$$

holds with equality if and only if $\mu^{\text{ss}}_1 = \mu^{\text{ss}}_2 = \ldots = \mu^{\text{ss}}_L$. This result follows from the proof in [13, Appendix A]. In [13] the inequality is proven for ensemble average matrices, but the proof can be extended to the matrices (21) and (22) as all the necessary assumptions still hold. Hence, by performing FB we may increase the Lth eigenvalue of $\mathbf{B}^{\text{fbss}}$ compared to that of $\mathbf{B}^{\text{ss}}$, while $\mathbf{B}^{\text{ss}}$ and $\mathbf{B}^{\text{fbss}}$ have the same matrix dimensions. This result plays an important role for the previously mentioned preprocessing trade-off on the selection of the window size M_1.

Notice that due to the selection of equally spaced pilots, the Hermitian matrix $\mathbf{R}$ is in fact Toeplitz. This entails that $\mathbf{R}$ is invariant under FB and (8) still holds. Usually FB is applied to $\hat{\mathbf{R}}$ with the justification that $\mathbf{R}$ is Toeplitz and thereby persymmetric. It is therefore important to stress that we only apply the technique due to its property (24) together with its ability to increase the matrix rank without lowering the matrix dimensions.

V. Investigation of the Window Size M_1

Since $\hat{\mathbf{R}}^{\text{ss}}$ and $\hat{\mathbf{R}}^{\text{fbss}}$ are computed from noise corrupted samples and finite sample sizes, their associated noise eigenvectors may be mistaken for signal eigenvectors and vice versa. According to (8), it is therefore desirable that the least signal sample eigenvalue $\tilde{\lambda}_L$ is large relative to $\tilde{\lambda}_{L+1}$, i.e. $\tilde{\mu}_L$ should be as large as possible. In this section, we investigate the behavior of the eigenvalues μ^{ss}_L and μ^{fbss}_L as a function of M_1. Our approach relies on decoupling the effects of the preprocessing techniques, i.e. (i) the reduction of the sample matrix dimensions when applying SS (14), (ii) the decorrelation inherited from the preprocessing (23) and (iii) that the Lth eigenvalue of $\mathbf{B}^{\text{fbss}}$ cannot be smaller than that of $\mathbf{B}^{\text{ss}}$ (24). Additionally, we employ a performance metric to assess the accuracy of the subspace estimates obtained with the preprocessing techniques.

A. Decoupling the Preprocessing Effects

As in Section IV, the following investigations solely address the terms in the sample covariance matrices arising from the signal subspace, i.e. we disregard any term depending on noise, e.g. $\mathbf{E}$ in (10). We can conveniently decouple the effects of the preprocessing techniques into the three separated effects (i)-(iii). Notice that when SS and FB are applied, these effects appear jointly in a convoluted and compound fashion. To describe them separately, we define the matrices

$$\mathbf{F}^{\text{ss}}(M_1) := \mathbf{T}\mathbf{A}^{\text{ss}}\mathbf{T}^H \in \mathbb{C}^{M\times M} \tag{25}$$

$$\mathbf{F}^{\text{fbss}}(M_1) := \mathbf{T}\mathbf{A}^{\text{fbss}}\mathbf{T}^H \in \mathbb{C}^{M\times M} \tag{26}$$

$$\mathbf{F}^{\text{dim}}(M_1) := \mathbf{T}_{M_1}\text{Diag}\{\tilde{\mathbf{A}}\}\mathbf{T}^H_{M_1} \in \mathbb{C}^{M_1\times M_1}, \tag{27}$$

where $\text{Diag}\{\tilde{\mathbf{A}}\}$ is the diagonal matrix built with the diagonal entries of (11). Notice that all three matrices in (25) to (27) are functions of M_1. In (25) and (26) we decrease the magnitudes of the off-diagonal entries of $\tilde{\mathbf{A}}$ but without reducing the matrix dimensions. Hence, (25) and (26) mimic the decorrelation effect inherited from the preprocessing. In (27) however, we reduce the matrix dimensions while pretending that $\tilde{\mathbf{A}}$ has diagonal form. Hence, the matrix in (27) imitates the effect of reduced matrix dimensions.

Notice that $\mathbf{F}^{\text{dim}}$ in (27) is Hermitian and Toeplitz. From this, we are able to show the following result by application of Weyl's inequality [14]. For any choice of window sizes M_1 and M_2, with $M_1 < M_2$, we have for each $n = 1,2,\ldots,M_1$

$$\lambda_n\left(\mathbf{F}^{\text{dim}}(M_1)\right) \leq \lambda_n\left(\mathbf{F}^{\text{dim}}(M_2)\right), \tag{28}$$

where $\lambda_n\left(\mathbf{F}^{\text{dim}}\right)$ denotes the nth eigenvalue of $\mathbf{F}^{\text{dim}}$. Now, as the window size decreases the matrices $\mathbf{B}^{\text{ss}}$ and $\mathbf{B}^{\text{fbss}}$ are forced towards being Toeplitz due to (23). In the limiting case when $\mathbf{A}^{\text{ss}}$ and $\mathbf{A}^{\text{fbss}}$ become diagonal, then $\mathbf{B}^{\text{ss}}$ and $\mathbf{B}^{\text{fbss}}$ are identical and they equal (27). Hence, according to (28) every eigenvalue of (27) are decreasing (or constant) as a function of decreasing window size.

In Section VI, we analyse the compound decorrelation and dimension reduction effects on the eigenvalues μ^{ss}_L and μ^{fbss}_L by tracking individually the Lth eigenvalues of (25), (26) and (27). We denote the Lth eigenvalue of $\mathbf{F}^{\text{ss}}(M_1)$, $\mathbf{F}^{\text{fbss}}(M_1)$ and $\mathbf{F}^{\text{dim}}(M_1)$ by γ^{ss}_L, γ^{fbss}_L and γ^{dim}_L respectively. Specifically, μ^{ss}_L is analysed from γ^{ss}_L and γ^{dim}_L, while μ^{fbss}_L from γ^{fbss}_L and γ^{dim}_L.

B. Performance Metric for Subspace Estimation Accuracy

As mentioned in Section III, the column space of $\mathbf{T}$ coincides with the span of the signal eigenvectors. These eigenvectors are mutually orthogonal whereas, in general, the columns of $\mathbf{T}$ are not (they are only linearly independent). Therefore, to assess the "quality" of an estimated signal subspace, we employ the performance metric

$$\mathcal{N}(M_1) := \frac{1}{M_1}\left\|\mathbf{\Pi}_{\mathbf{T}} - \mathbf{\Pi}_{\hat{\mathbf{U}}_s}\right\|^2_F. \tag{29}$$

In (29), $\mathbf{\Pi}_{\mathbf{T}}$ and $\mathbf{\Pi}_{\hat{\mathbf{U}}_s}$ denote the operators projecting orthogonally onto respectively the true and the estimated signal subspaces, while $\|\cdot\|_F$ is the Frobenius norm. The projection operator $\mathbf{\Pi}_{\mathbf{T}}$ is defined as

$$\mathbf{\Pi}_{\mathbf{T}} := \mathbf{T}_{M_1}\mathbf{T}^{\dagger}_{M_1} \in \mathbb{C}^{M_1\times M_1}, \tag{30}$$

where $(\cdot)^{\dagger}$ denotes the Moore-Penrose generalized matrix inverse. Our choice of the performance metric (29) is based on the fact that the projection operator is invariant to the selected

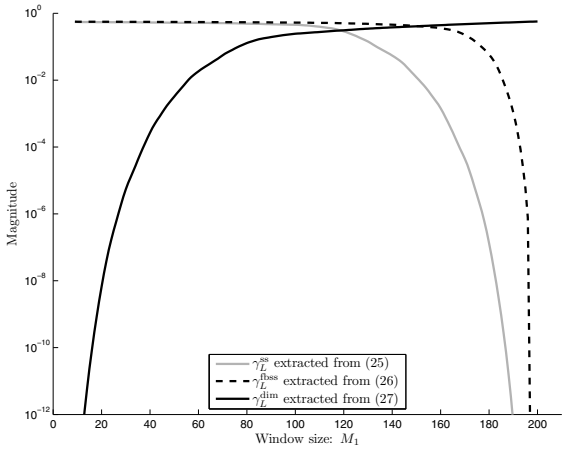

Fig. 1. Eigenvalues $\gamma_L^{\rm ss}$, $\gamma_L^{\rm fbss}$ and $\gamma_L^{\rm dim}$ as a function of M_1.

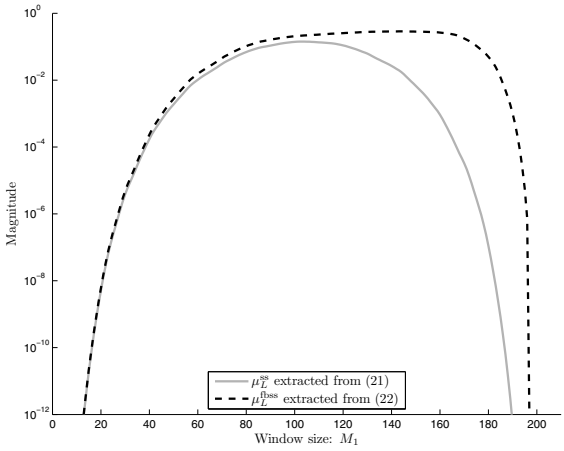

Fig. 2. Eigenvalues $\mu_L^{\rm ss}$ and $\mu_L^{\rm fbss}$ as a function of M_1.

basis used to span the subspace onto which the operator projects. The $M_1 \times L$ matrix $\hat{\mathbf{U}}_s$ contains the L estimated signal eigenvectors. Hence, to compute $\mathbf{\Pi}_{\hat{\mathbf{U}}_s}$ we simply insert $\hat{\mathbf{U}}_s$ instead of $\mathbf{T}_{M_1}$ in (30). The metric in (29) is related to the principal angles between the subspaces, see e.g. [15]. Notice that the squared Frobenius norm in (29) is weighted with $1/M_1$ for the purpose of allowing a performance comparison across different matrix dimensions.

VI. Experimental Results

We consider a 3GPP long term evolution alike scenario [16], using the parameters

$$N = 2048, \quad N_u = 1200, \quad M = 200, \quad \Delta p = 6.$$

The multipath channel in (4) is based on the 3GPP Extended Vehicular A Model [16, Annex B.2]. More specifically, the channel constantly holds $L = 9$ multipath components, where L is assumed known to the receiver. Relative multipath delays, power delay profile and maximum excess delay of the channel are specified in [16, Annex B.2].

We let $K = 1$ for all simulations. Hence, the rank of $\hat{\mathbf{R}}$ is one and we can only obtain the eigenvalue ordering in (8) through smoothing. Uncoded QPSK modulation is used with Gray mapping both for data and pilot symbols. All curves are computed based on a total of 1000 Monte Carlo runs.

A. Subspace Estimation Performance

We plot the Lth eigenvalue of (25), (26) and (27) in Fig. 1 and the Lth eigenvalue of (21) and (22) in Fig. 2. Notice that all reported eigenvalues do not depend on the SNR level.

By comparing Fig. 1 and Fig. 2 we observe that the behavior of $\mu_L^{\rm ss}$ can be explained from $\gamma_L^{\rm ss}$ and $\gamma_L^{\rm dim}$, while $\mu_L^{\rm fbss}$ is explained from $\gamma_L^{\rm fbss}$ and $\gamma_L^{\rm dim}$, as described in Section V. From Fig. 2 we see that $\mu_L^{\rm fbss}$ and $\mu_L^{\rm ss}$ are related as given in (24), i.e. $\mu_L^{\rm fbss} \geq \mu_L^{\rm ss}$. Moreover, $\mu_L^{\rm fbss}$ is near its maximum for a wider M_1-region compared to $\mu_L^{\rm ss}$. Finally, in Fig. 1, we see how $\gamma_L^{\rm dim}$ decreases with M_1 according to (28).

In Fig. 3 we depict the metric (29) versus M_1 for three

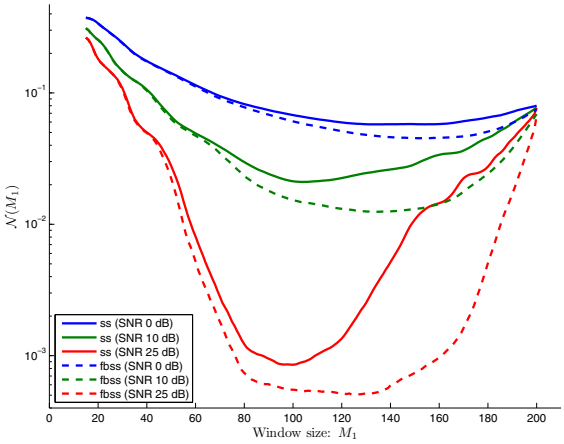

Fig. 3. Performance metric (29) versus M_1 with the SNR as a parameter.

selected levels of SNR. As a consequence of (24) (see Fig. 2), we see how FBSS achieves wider M_1-regions with better subspace estimation performance as compared to SS. The gain from the preprocessing increases with the SNR, which emphasizes that the subspace estimation performance depends on the actual SNR level. However, the near optimum window size almost remain unchanged regardless of the SNR level.

B. Channel Estimation Performance

We now use the ESPRIT algorithm for the estimation of the channel multipath delays, see [8]. Prior to ESPRIT, SS with and without FB are applied. OFDM channel estimation is performed using the LMMSE estimator from [8]. Simulations have also been conducted using Unitary ESPRIT [17] instead of (standard) ESPRIT. However, both algorithms perform similarly because FB is a built-in feature of Unitary ESPRIT.

In Fig. 4 we report the uncoded *bit-error-rate* (BER) performance of the OFDM system versus the window size M_1. By jointly employing both preprocessing schemes we achieve wide M_1-regions with BER performance close to the performance obtained when the channel is known. We observe that for high SNR the drop in BER performance for large and

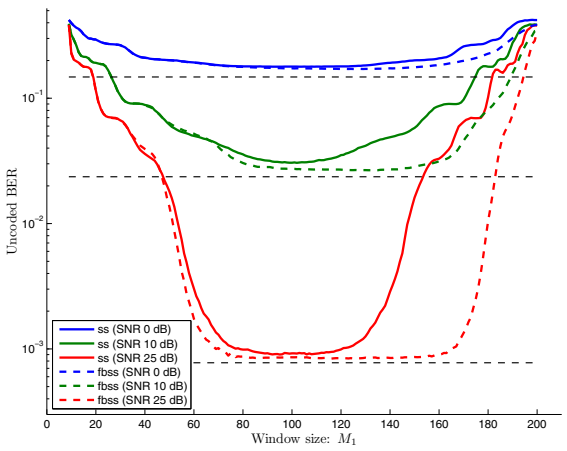

Fig. 4. Uncoded BER performance as a function of M_1 with the SNR as a parameter. The black-dashed lines indicate the BER performance at the three considered SNR values when the channel response, i.e. $\mathbf{h}$ in (1), is known.

small values of M_1 may be explained by the decrease of μ_L^{ss} and μ_L^{fbss} in these M_1-regions (see Fig. 2). As in Fig. 3 we observe that the general behavior of the curves (and thereby the choice of window size) remain similar across SNR levels. However, we do not observe a performance gain in Fig. 3 for small M_1 as in Fig. 4. Therefore, the metric (29) does not encompass all aspects determining for the system assessment. A more adequate metric for comparing the subspace estimation performance across different window sizes is still an open issue.

VII. Conclusion

In this paper, we have provided an analysis of spatial smoothing and forward-backward averaging for subspace-based methods. We have decoupled the compound impacts of these techniques into separate effects, more specifically, a decorrelation effect and a dimension reduction effect. From this we have been able to describe the overall behavior of the least signal eigenvalue as a function of the size of the used window. Through Monte Carlo simulations we have demonstrated that this behavior critically affects a proper separation of signal and noise subspaces.

We have applied the insight gained from the above investigations to the problem of channel estimation in an OFDM system with the pilot positions appropriately selected, so that the preprocessing techniques can be applied. The results show that the selection of the appropriate window size is dictated by the behavior of the least signal eigenvalue. Furthermore, jointly applying forward-backward averaging and spatial smoothing yields near optimum performance for a broad range of windows sizes. This allows to select the window size with greater flexibility, as compared to using spatial smoothing alone.

The dramatic performance drop of the subspace-based methods for certain window sizes underlines the importance of the analysis conducted in this paper.

Acknowledgment

This work was supported in part by the 4GMCT cooperative research project funded by Infineon Technologies Denmark A/S, Agilent Technologies, Aalborg University and the Danish National Advanced Technology Foundation and by the European Commission within the ICT-216715 FP7 Network of Excellence in Wireless Communications (NEWCOM++) and the two projects ICT-217033 and ICT-248894 Wireless Hybrid Enhanced Mobile Radio Estimators (WHERE and WHERE2).

References

[1] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, 1986.

[2] R. Roy and T. Kailath, "ESPRIT - Estimation of signal parameters via rotational invariance techniques", *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, no. 7, 1989.

[3] T. J. Shan, M. Wax and T. Kailath, "On spatial smoothing for direction-of-arrival estimation of coherent signals", *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 33, no. 4, 1985.

[4] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach", *IEEE Signal Processing Magazine*, vol. 13, no. 4, 1996.

[5] V. Reddy, A. Paulraj and T. Kailath, "Performance analysis of the optimum beamformer in the presence of correlated sources and its behavior under spatial smoothing", *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 35, no. 7, 1987.

[6] H. Yamada, M. Ohmiya, Y. Ogawa, and K. Itoh, "Superresolution techniques for time-domain measurements with a network analyzer", *IEEE Transactions on Antennas and Propagation*, vol. 39, no. 2, 1991.

[7] X. Li and K. Pahlavan "Super-resolution TOA estimation with diversity for indoor geolocation", *IEEE Transactions on Wireless Communications*, vol. 3, no. 1, 2004.

[8] B. Yang, K. B. Letaief, R. S. Cheng and Z. Cao, "Channel estimation for OFDM transmission in multipath fading channels based on parametric channel modeling", *IEEE Transactions on Communications*, vol. 49, no. 3, 2001.

[9] M. L. Jakobsen, K. Laugesen, C. Navarro, G. E. Kirkelund, C. Rom and B. Fleury, "Parametric modeling and pilot-aided estimation of the wireless multipath channel in OFDM systems", *IEEE International Conference on Communications*, 2010.

[10] J. K. Thomas, L. L. Scharf and D. W. Tufts, "The probability of a subspace swap in the SVD," *IEEE Transactions on Signal Processing*, vol. 43, no. 3, 1995.

[11] P. A. Bello, "Characterization of randomly time-variant linear channels", *IEEE Transactions on Communications Systems*, 1963.

[12] P. Stoica and R. Moses, "Spectral Analysis of Signals", *Pearson Prentice Hall*, 2005.

[13] B. D. Rao and K. V. S. Hari, "Weighted subspace methods and spatial smoothing: analysis and comparison", *IEEE Transactions on Signal Processing*, vol. 41, no. 2, 1993.

[14] Roger A. Horn and Charles R. Johnson, "Matrix Analysis", *Cambridge University Press*, 1990.

[15] A. Edelman, T. A. Arias and S. T. Smith, "The geometry of algorithms with orthogonality constraints", *SIAM J. Matrix Anal. Appl.*, 20, 1998.

[16] "Evolved Universal Terrestrial Radio Access (E-UTRA); Base Station (BS) radio transmission and reception", *3rd Generation Partnership Project (3GPP) Technical Specification*, TS 36.104 V8.4.0, Dec. 2008.

[17] M. Haardt and J. A. Nossek, "Unitary ESPRIT: How to obtain increased estimation accuracy with a reduced computational burden," *IEEE Transactions on Signal Processing*, vol. 43, no. 5, 1995.