Block Krylov Methods in Time-Dispersive MIMO Systems

Guido Dietl, Peter Breun, and Wolfgang Utschick

Institute for Circuit Theory and Signal Processing
Munich University of Technology, 80333 Munich, Germany
E-mail: gudi@nws.ei.tum.de

***Abstract*— Compared to the conventional full-rank Wiener Filter (WF), reduced-rank processing in the minimum mean square error sense is a well-known strategy in order to reduce computational complexity and enhance performance in case of low sample support.**

In this paper, we reveal the relationship between block Krylov methods and the Multi-Stage Matrix WF (MSMWF) as a reduced-rank matrix WF which estimates a signal vector instead of a scalar. The new insights lead to an implementation of the MSMWF based on Ruhe's variant of the block Lanczos algorithm which is more flexible with respect to rank selection compared to existing algorithms.

Finally, the application to a time-dispersive Multiple-Input Multiple-Output (MIMO) system demonstrates the ability of the new algorithm to lessen receiver complexity while maintaining the same level of system performance or even improve it if second order statistics are not perfectly known. Moreover, the MSMWF outperforms the parallel implementation of multi-stage vector WFs with a comparable computational complexity.

I. Introduction

The *Wiener Filter* (WF) [1], [2] estimates an unknown signal from an observation signal by minimizing the *Mean Square Error* (MSE). The derivation ends up in solving the *Wiener-Hopf equation* which is computational intense for observations of high dimensionality. Since reduced-rank methods approximate the WF in a lower dimensional subspace, they reduce computational complexity and enhance the robustness against estimation errors of second order statistics due to low sample support.

The *Principal Component* (PC) *algorithm* [3] was the first reduced-rank approach where the eigenvectors corresponding to the largest eigenvalues of the auto-correlation matrix of the observation vector span the subspace for the approximation of the WF. An alternative method exploits information about the cross-correlation between the observation and the desired signal by using the *Cross-Spectral* (CS) *metric* [4] instead of the eigenvalue magnitude to choose the eigenvectors spanning the subspace. Contrary to the eigenspace based PC or CS method, the *Multi-Stage Vector WF* (MSVWF) [5], [6] approximates the *vector WF* in a Krylov subspace composed of the auto-correlation matrix of the observation and the cross-correlation vector between the observation and the desired signal. More recently, Goldstein et al. [7], [8] derived the *Multi-Stage Matrix WF* (MSMWF) by applying the multi-stage principle to a *Matrix WF* (MWF) which estimates a signal vector instead of a scalar. Due to the block structure, the dimension of the observation vector and the rank of the MSMWF is restricted to be an integer multiple of the dimension of the desired signal vector.

Our contribution is to show the relationship between the MSMWF and *block Krylov methods* [9] like the *block Arnoldi* [10] or the *block Lanczos algorithm* [11], [12], [13] if we exploit additionally that auto-correlation matrices are Hermitian. Moreover, we consider *Ruhe's version* [14] of the Lanczos algorithm in order to get a new implementation of the MSMWF which is more flexible concerning rank selection, i. e. the rank can be any integer between the dimension of the signal vector and the dimension of the observation vector. Furthermore, the dimension of the observation is no longer restricted to be an integer multiple of the dimension of the desired signal.

In the next section, we briefly review the MSMWF. Before we derive a MSMWF based on the Lanczos-Ruhe algorithm in Section IV, we show the relationship between the MSMWF and block Krylov methods in Section III. Finally, we apply the considered algorithms to a time-dispersive MIMO System in Section V. Throughout the paper, $\boldsymbol{R}_{\mathbf{u}} = \mathrm{E}\left\{\mathbf{u}\mathbf{u}^{\mathrm{H}}\right\}$ denotes the auto-correlation matrix of the random vector $\mathbf{u}$, $\boldsymbol{R}_{\mathbf{u},\mathbf{v}} = \mathrm{E}\left\{\mathbf{u}\mathbf{v}^{\mathrm{H}}\right\}$ the cross-correlation matrix between the vectors $\mathbf{u}$ and $\mathbf{v}$, $\mathbf{1}_n$ a $n \times n$ identity, and $\mathbf{0}_{m\times n}$ a $m \times n$ zero matrix.

II. Multi-Stage Matrix Wiener Filter (MSMWF)

Applying the linear matrix filter $\boldsymbol{\mathcal{W}} \in \mathbb{C}^{M\times N}$, $M \in \mathbb{N}$, $N \in \mathbb{N}$, to the observation vector $\mathbf{y}[n] \in \mathbb{C}^N$, $N \geq M$, leads to the estimate $\hat{\mathbf{x}}[n] = \boldsymbol{\mathcal{W}}\mathbf{y}[n]$ of the desired signal vector $\mathbf{x}[n] \in \mathbb{C}^M$. The power of the Euclidian norm of the estimation error $\mathbf{e}[n] = \mathbf{x}[n] - \hat{\mathbf{x}}[n]$ is the *Mean Square Error* (MSE)

$$\xi\left(\boldsymbol{\mathcal{W}}\right) = \mathrm{tr}\left\{\boldsymbol{R}_{\mathbf{x}} - 2\,\mathrm{Re}\left\{\boldsymbol{\mathcal{W}}\boldsymbol{R}_{\mathbf{y},\mathbf{x}}\right\} + \boldsymbol{\mathcal{W}}\boldsymbol{R}_{\mathbf{y}}\boldsymbol{\mathcal{W}}^{\mathrm{H}}\right\}. \quad (1)$$

The *Matrix Wiener Filter* (MWF) $\boldsymbol{W}$ minimizes $\xi\left(\boldsymbol{\mathcal{W}}\right)$ and consequently solves the *Wiener-Hopf equation*

$$\boldsymbol{R}_{\mathbf{y}}\boldsymbol{W}^{\mathrm{H}} = \boldsymbol{R}_{\mathbf{y},\mathbf{x}} \quad \Leftrightarrow \quad \boldsymbol{W} = \boldsymbol{R}_{\mathbf{y},\mathbf{x}}^{\mathrm{H}}\boldsymbol{R}_{\mathbf{y}}^{-1}. \quad (2)$$

Thus, the *Minimum MSE* (MMSE) may be written as $\xi\left(\boldsymbol{W}\right) = \mathrm{tr}\left\{\boldsymbol{R}_{\mathbf{x}} - \boldsymbol{R}_{\mathbf{y},\mathbf{x}}^{\mathrm{H}}\boldsymbol{R}_{\mathbf{y}}\boldsymbol{R}_{\mathbf{y},\mathbf{x}}\right\}$.

0-7803-8533-0/04/$20.00 (c) 2004 IEEE

The first step of the *Multi-Stage Matrix Wiener Filter* (MSMWF) introduced by Goldstein et al. [7], [8] is to prefilter the observation vector $\mathbf{y}[n]$ with the full-rank matrix $\tilde{\boldsymbol{T}}_1 = \left[\boldsymbol{M}_1^{\mathrm{T}}, \boldsymbol{B}_1^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{C}^{N \times N}$ to get the transformed observation vector

$$\mathbf{z}_1[n] = \tilde{\boldsymbol{T}}_1 \mathbf{y}[n] = \begin{bmatrix} \boldsymbol{M}_1 \mathbf{y}[n] \\ \boldsymbol{B}_1 \mathbf{y}[n] \end{bmatrix} =: \begin{bmatrix} \mathbf{x}_1[n] \\ \mathbf{y}_1[n] \end{bmatrix}, \tag{3}$$

where

$$\boldsymbol{M}_1 = \left(\boldsymbol{R}_{\mathbf{y},\mathbf{x}}^{\mathrm{H}} \boldsymbol{R}_{\mathbf{y},\mathbf{x}}\right)^{-\frac{1}{2}} \boldsymbol{R}_{\mathbf{y},\mathbf{x}}^{\mathrm{H}} \in \mathbb{C}^{M \times N} \tag{4}$$

is a matched filter with orthonormal rows, i. e. the argument of the optimization

$$\max_{\boldsymbol{\mathcal{M}}} \mathrm{E}\left\{\mathrm{Re}\left\{\mathbf{x}_1^{\mathrm{H}}[n]\mathbf{x}[n]\right\}\right\} \quad \text{s.t.} \quad \boldsymbol{\mathcal{M}}\boldsymbol{\mathcal{M}}^{\mathrm{H}} = \mathbf{1}_M. \tag{5}$$

Note that $\boldsymbol{M}_1$ maximizes the sum of cross-correlations between each element of the signal vector $\mathbf{x}[n]$ and the corresponding element of the output vector $\mathbf{x}_1[n] \in \mathbb{C}^M$. The rows of the so-called blocking matrix $\boldsymbol{B}_1 \in \mathbb{C}^{(N-M) \times N}$ are orthonormal to the rows of $\boldsymbol{M}_1$, i. e.

$$\mathrm{span}\left\{\boldsymbol{B}_1^{\mathrm{H}}\right\} = \mathrm{null}\left\{\boldsymbol{M}_1\right\} \Leftrightarrow \boldsymbol{B}_1 \boldsymbol{M}_1^{\mathrm{H}} = \mathbf{0}_{(N-M)\times M}. \tag{6}$$

Thus, the transformed observation vector $\mathbf{y}_1[n] \in \mathbb{C}^{N-M}$ is uncorrelated to the signal vector $\mathbf{x}[n]$ but still bears information about the interference in $\mathbf{x}_1[n]$ which was not suppressed by the matched filter $\boldsymbol{M}_1$ with orthonormal rows.

With the inversion lemma for partitioned matrices [1], the MWF estimating the signal vector $\mathbf{x}[n]$ from the transformed observation vector $\mathbf{z}_1[n]$ may be written as

$$\boldsymbol{W}_{\mathbf{z}_1} = \boldsymbol{\Delta}_1 \left[\mathbf{1}_M, -\boldsymbol{W}_1\right] \in \mathbb{C}^{M \times N}, \tag{7}$$

where

$$\boldsymbol{\Delta}_1 = \boldsymbol{R}_{\mathbf{e}_1,\mathbf{x}}^{\mathrm{H}} \boldsymbol{R}_{\mathbf{e}_1}^{-1} \quad \text{with} \tag{8}$$

$$\boldsymbol{R}_{\mathbf{e}_1,\mathbf{x}} = \boldsymbol{M}_1 \boldsymbol{R}_{\mathbf{y},\mathbf{x}} = \left(\boldsymbol{R}_{\mathbf{y},\mathbf{x}}^{\mathrm{H}} \boldsymbol{R}_{\mathbf{y},\mathbf{x}}\right)^{\frac{1}{2}} \quad \text{and} \tag{9}$$

$$\boldsymbol{R}_{\mathbf{e}_1} = \boldsymbol{R}_{\mathbf{x}_1} - \boldsymbol{R}_{\mathbf{y}_1,\mathbf{x}_1}^{\mathrm{H}} \boldsymbol{R}_{\mathbf{y}_1}^{-1} \boldsymbol{R}_{\mathbf{y}_1,\mathbf{x}_1} \tag{10}$$

is a quadratic MWF which estimates the desired signal vector $\mathbf{x}[n]$ from the error vector $\mathbf{e}_1[n] = \mathbf{x}_1[n] - \hat{\mathbf{x}}_1[n] \in \mathbb{C}^M$ with $\hat{\mathbf{x}}_1[n] = \boldsymbol{W}_1 \mathbf{y}_1[n] \in \mathbb{C}^M$ (cf. Equations 4, 6, and 11). The *reduced-dimension MWF*[1]

$$\boldsymbol{W}_1 = \boldsymbol{R}_{\mathbf{y}_1,\mathbf{x}_1}^{\mathrm{H}} \boldsymbol{R}_{\mathbf{y}_1}^{-1} \in \mathbb{C}^{M \times (N-M)} \tag{11}$$

estimates $\mathbf{x}_1[n]$ from the prefiltered observation vector $\mathbf{y}_1[n]$. The resulting structure which produces the same output $\hat{\mathbf{x}}[n]$ as the MWF is summarized in Figure 1.

Note that this structure is very related to the *Generalized Sidelobe Canceller* (GSC) [15], [16], especially if we assume the vector case, i. e. $M = 1$. The MWF $\boldsymbol{W}_1$ estimates the interference remained in $\mathbf{x}_1[n]$, i. e. $\hat{\mathbf{x}}_1[n]$, in order to subtract it from $\mathbf{x}_1[n]$. The quadratic MWF $\boldsymbol{\Delta}_1$ is needed to reconstruct the desired signal vector in the MMSE sense. Nevertheless, the classical GSC has been designed to be the solution of the *minimum variance distortionless response* constrained optimization problem and not the MMSE solution.

[1]The notation *reduced-dimension MWF* denotes a matrix whose rows are reduced-dimension vector WFs.

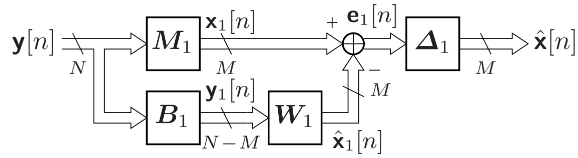

Fig. 1. Multi-Stage Matrix Wiener Filter (MSMWF) after First Step

Analogous to the original MWF $\boldsymbol{W}$, the reduced-dimension MWF $\boldsymbol{W}_1$ can be replaced by the prefilter matrix $\tilde{\boldsymbol{T}}_2 = \left[\boldsymbol{M}_2^{\mathrm{T}}, \boldsymbol{B}_2^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{C}^{(N-M)\times(N-M)}$ and the reduced-dimension MWF $\boldsymbol{W}_{\mathbf{z}_2} = \boldsymbol{\Delta}_2 \left[\mathbf{1}_M, -\boldsymbol{W}_2\right] \in \mathbb{C}^{M \times (N-M)}$. Continuing this replacement method for every $\boldsymbol{W}_\ell \in \mathbb{C}^{M \times (N-\ell M)}$, $\ell \in \{2, 3, \ldots, L-2\}$, and finally substituting $\boldsymbol{W}_{L-1} \in \mathbb{C}^{M \times M}$ by $\boldsymbol{\Delta}_L \boldsymbol{M}_L$, leads to the filter bank depicted in Figure 2 if N is an integer multiple of M, i. e. $N = LM$, $L \in \mathbb{N}$.

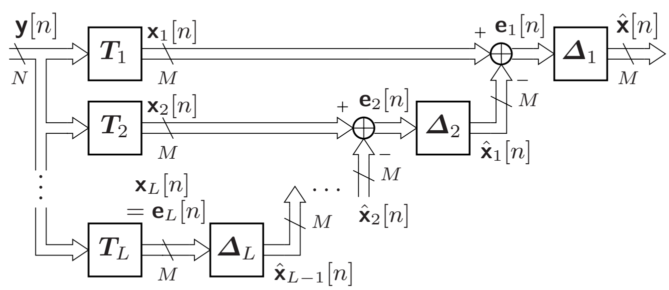

Fig. 2. MSMWF as a Filter Bank

The prefilter matrices

$$\boldsymbol{T}_\ell = \boldsymbol{M}_\ell \prod_{\substack{k=\ell-1 \\ k>0}}^{1} \boldsymbol{B}_k \in \mathbb{C}^{M \times N}, \quad \ell \in \{1, 2, \ldots, L\}, \tag{12}$$

are composed of the matched filters $\boldsymbol{M}_\ell \in \mathbb{C}^{M \times (N-(\ell-1)M)}$ with orthonormal rows, i. e.

$$\boldsymbol{M}_\ell = \left(\boldsymbol{R}_{\mathbf{y}_{\ell-1},\mathbf{x}_{\ell-1}}^{\mathrm{H}} \boldsymbol{R}_{\mathbf{y}_{\ell-1},\mathbf{x}_{\ell-1}}\right)^{-\frac{1}{2}} \boldsymbol{R}_{\mathbf{y}_{\ell-1},\mathbf{x}_{\ell-1}}^{\mathrm{H}}, \tag{13}$$

maximizing the sum of cross-correlations between each element of the M-dimensional transformed observation vectors $\mathbf{x}_\ell[n]$ and the corresponding element of $\mathbf{x}_{\ell-1}[n]$, where $\mathbf{x}_0[n] := \mathbf{x}[n]$ and $\mathbf{y}_0[n] := \mathbf{y}[n]$. The complex conjugate rows of the blocking matrices $\boldsymbol{B}_\ell \in \mathbb{C}^{(N-\ell M)\times(N-(\ell-1)M)}$ are chosen to span the null-space of $\boldsymbol{M}_\ell$, i. e.

$$\mathrm{span}\left\{\boldsymbol{B}_\ell^{\mathrm{H}}\right\} = \mathrm{null}\left\{\boldsymbol{M}_\ell\right\} \Leftrightarrow \boldsymbol{B}_\ell \boldsymbol{M}_\ell^{\mathrm{H}} = \mathbf{0}_{(N-\ell M)\times M}. \tag{14}$$

Thus, $\mathbf{y}_\ell[n]$ is uncorrelated to $\mathbf{x}_{\ell-1}[n]$. Finally, the matrices

$$\boldsymbol{\Delta}_\ell = \boldsymbol{R}_{\mathbf{e}_\ell,\mathbf{x}_{\ell-1}}^{\mathrm{H}} \boldsymbol{R}_{\mathbf{e}_\ell}^{-1} \quad \text{with} \tag{15}$$

$$\boldsymbol{R}_{\mathbf{e}_\ell,\mathbf{x}_{\ell-1}} = \left(\boldsymbol{R}_{\mathbf{y}_{\ell-1},\mathbf{x}_{\ell-1}}^{\mathrm{H}} \boldsymbol{R}_{\mathbf{y}_{\ell-1},\mathbf{x}_{\ell-1}}\right)^{\frac{1}{2}} \quad \text{and} \tag{16}$$

$$\boldsymbol{R}_{\mathbf{e}_\ell} = \begin{cases} \boldsymbol{R}_{\mathbf{x}_\ell} - \boldsymbol{R}_{\mathbf{y}_\ell,\mathbf{x}_\ell}^{\mathrm{H}} \boldsymbol{R}_{\mathbf{y}_\ell}^{-1} \boldsymbol{R}_{\mathbf{y}_\ell,\mathbf{x}_\ell}, & \ell < L, \\ \boldsymbol{R}_{\mathbf{x}_L}, & \ell = L, \end{cases} \tag{17}$$

are $M \times M$ MWFs which estimate $\mathbf{x}_{\ell-1}[n]$ from the error vector $\mathbf{e}_\ell[n] = \mathbf{x}_\ell[n] - \hat{\mathbf{x}}_\ell[n]$.

Up to now, the MSMWF is equivalent to the optimal MWF, i. e. it produces the same output $\hat{\mathbf{x}}[n]$. If we neglect $\hat{\mathbf{x}}_d[n]$, $d \in \{1, 2, \ldots, L-1\}$, such that $\mathbf{e}_d[n] = \mathbf{x}_d[n]$ (cf. Figure 2), we obtain the rank D MSMWF $\boldsymbol{W}^{(D)} \in \mathbb{C}^{M \times N}$, $D = dM \in \mathbb{N}$, which is composed of the multi-stage prefilter matrix

$$\boldsymbol{T}^{(D)} = \left[\boldsymbol{T}_1^{\mathrm{T}}, \boldsymbol{T}_2^{\mathrm{T}}, \ldots, \boldsymbol{T}_d^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{C}^{D \times N}, \tag{18}$$

and the reduced-dimension MWF

$$\begin{aligned} \boldsymbol{W}_{\mathrm{rd}}^{(D)} &= \underset{\boldsymbol{\mathcal{W}}_{\mathrm{rd}}^{(D)}}{\operatorname{argmin}}\, \xi\left(\boldsymbol{\mathcal{W}}_{\mathrm{rd}}^{(D)} \boldsymbol{T}^{(D)}\right) \\ &= \boldsymbol{R}_{\mathbf{y},\mathbf{x}}^{\mathrm{H}} \boldsymbol{T}^{(D),\mathrm{H}} \left(\boldsymbol{T}^{(D)} \boldsymbol{R}_{\mathbf{y}} \boldsymbol{T}^{(D),\mathrm{H}}\right)^{-1} \in \mathbb{C}^{M \times D}, \end{aligned} \tag{19}$$

estimating $\mathbf{x}[n]$ from the prefiltered observation vector $\boldsymbol{T}^{(D)}\mathbf{y}[n] \in \mathbb{C}^D$.

Note that the rank D as well as the dimension N of the observation vector is restricted to be an integer multiple of the dimension M of the signal vector. In Section IV, we propose an implementation of the MSMWF where D can be any integer between M and N, and N is no longer limited to integer multiples of M.

The *reduced-rank MWF*[2] $\boldsymbol{W}^{(D)} = \boldsymbol{W}_{\mathrm{rd}}^{(D)} \boldsymbol{T}^{(D)}$ is an approximation of the MWF $\boldsymbol{W}$ in the subspace spanned by the complex conjugate rows of the multi-stage prefilter matrix $\boldsymbol{T}^{(D)}$. Due to the characteristics of $\boldsymbol{T}^{(D)}$ which are investigated in more detail in the following section, the reduced-dimension MWF $\boldsymbol{W}_{\mathrm{rd}}^{(D)}$ decomposes into the quadratic MWFs $\boldsymbol{\Delta}_\ell$, $\ell \in \{1, 2, \ldots, d\}$, arranged as shown in Figure 2. The rank D MSMWF achieves the MMSE

$$\xi\left(\boldsymbol{W}^{(D)}\right) = \operatorname{tr}\left\{\boldsymbol{R}_{\mathbf{x}} - \boldsymbol{W}^{(D)} \boldsymbol{R}_{\mathbf{y},\mathbf{x}}\right\}, \tag{20}$$

which is generally greater than $\xi(\boldsymbol{W})$.

III. Relationship between MSMWF and Block Krylov Methods

Recall that the matched filters $\boldsymbol{M}_\ell$ maximize the sum of cross-correlations between each element of the transformed observation vectors of two adjacent stages, viz. $\mathbf{x}_\ell[n]$ and $\mathbf{x}_{\ell-1}[n]$. Hence, if we assume blocking matrices with orthonormal rows, i. e. $\boldsymbol{B}_\ell \boldsymbol{B}_\ell^{\mathrm{H}} = \mathbf{1}_{N-\ell M}$, the prefilter matrices may be obtained by the following optimization [6], [17]:

$$\begin{aligned} \boldsymbol{T}_\ell = \underset{\boldsymbol{\mathcal{T}}}{\operatorname{argmax}}\, &\mathrm{E}\left\{\mathrm{Re}\left\{\mathbf{x}_\ell^{\mathrm{H}}[n] \mathbf{x}_{\ell-1}[n]\right\}\right\} \\ \text{s.t.} \quad &\boldsymbol{\mathcal{T}}\boldsymbol{\mathcal{T}}^{\mathrm{H}} = \mathbf{1}_M \quad \text{and} \\ &\boldsymbol{\mathcal{T}}\boldsymbol{T}_k^{\mathrm{H}} = \mathbf{0}_{M \times M}, \quad k \in \{1, 2, \ldots, \ell-1\}. \end{aligned} \tag{21}$$

It can easily be seen that $\boldsymbol{T}_1 = \boldsymbol{M}_1$ since the second side condition is not active for $\ell = 1$. Henceforth, we consider the case $\ell > 1$. With the Lagrangian function

$$\begin{aligned} L\left(\boldsymbol{\mathcal{T}}, \boldsymbol{\Lambda}_1, \boldsymbol{\Lambda}_2, \ldots, \boldsymbol{\Lambda}_\ell\right) = \operatorname{tr}\Big\{\mathrm{Re}\Big\{&\boldsymbol{T}_{\ell-1} \boldsymbol{R}_{\mathbf{y}} \boldsymbol{\mathcal{T}}^{\mathrm{H}} \\ &- \sum_{k=1}^{\ell-1} \boldsymbol{\Lambda}_k^{\mathrm{H}} \boldsymbol{\mathcal{T}} \boldsymbol{T}_k^{\mathrm{H}} - \boldsymbol{\Lambda}_\ell^{\mathrm{H}}\left(\boldsymbol{\mathcal{T}}\boldsymbol{\mathcal{T}}^{\mathrm{H}} - \mathbf{1}_M\right)\Big\}\Big\} \end{aligned} \tag{22}$$

and the projector matrices $\boldsymbol{P}_{\perp \boldsymbol{T}_k^{\mathrm{H}}} = \mathbf{1}_N - \boldsymbol{T}_k^{\mathrm{H}} \boldsymbol{T}_k$ projecting onto the space orthogonal to the one spanned by the complex conjugate rows of $\boldsymbol{T}_k$, we get after some computation steps

$$\tilde{\boldsymbol{\Lambda}}_\ell \boldsymbol{\mathcal{T}} = \boldsymbol{T}_{\ell-1} \boldsymbol{R}_{\mathbf{y}} \prod_{k=1}^{\ell-1} \boldsymbol{P}_{\perp \boldsymbol{T}_k^{\mathrm{H}}} =: \boldsymbol{S}\left(\boldsymbol{T}_{\ell-1}\right) =: \boldsymbol{S}_{\ell-1}, \tag{23}$$

where $\tilde{\boldsymbol{\Lambda}}_\ell := \boldsymbol{\Lambda}_\ell^{\mathrm{H}} + \boldsymbol{\Lambda}_\ell$ is Hermitian. Thus, the solution of the prefilter matrix

$$\boldsymbol{T}_\ell = \left(\boldsymbol{S}_{\ell-1} \boldsymbol{S}_{\ell-1}^{\mathrm{H}}\right)^{-\frac{1}{2}} \boldsymbol{S}_{\ell-1} \tag{24}$$

is given recursively due to the dependency of $\boldsymbol{S}_{\ell-1} \in \mathbb{C}^{M \times N}$ on $\boldsymbol{T}_{\ell-1}$.

Proposition 1: The auto-correlation matrix $\boldsymbol{T}^{(D)} \boldsymbol{R}_{\mathbf{y}} \boldsymbol{T}^{(D),\mathrm{H}}$ of the prefiltered observation vector is *block tridiagonal*[3].

Proof: If we replace ℓ by $i+1$ in Equation (23), multiply it by $\boldsymbol{T}_\ell^{\mathrm{H}}$ on the right hand side, and use additionally Equation (24), we get for $i \in \{1, 2, \ldots, \ell-1\}$

$$\left(\boldsymbol{S}_i \boldsymbol{S}_i^{\mathrm{H}}\right)^{\frac{1}{2}} \boldsymbol{T}_{i+1} \boldsymbol{T}_\ell^{\mathrm{H}} = \boldsymbol{T}_i \boldsymbol{R}_{\mathbf{y}} \boldsymbol{T}_\ell^{\mathrm{H}}, \tag{25}$$

since the matrices $\boldsymbol{T}_i$ fulfill the constraints of the optimization in Equation (21). Using again this property, it follows that

$$\boldsymbol{T}_i \boldsymbol{R}_{\mathbf{y}} \boldsymbol{T}_\ell^{\mathrm{H}} = \begin{cases} \left(\boldsymbol{S}_{\ell-1} \boldsymbol{S}_{\ell-1}^{\mathrm{H}}\right)^{\frac{1}{2}}, & \text{for } i = \ell - 1, \\ \mathbf{0}_{M \times M}, & \text{for } i < \ell - 1, \end{cases} \tag{26}$$

i. e. $\boldsymbol{T}^{(D)} \boldsymbol{R}_{\mathbf{y}} \boldsymbol{T}^{(D),\mathrm{H}}$ has a *lower block Hessenberg structure*[4]. If we recall that auto-correlation matrices are Hermitian, $\boldsymbol{T}^{(D)} \boldsymbol{R}_{\mathbf{y}} \boldsymbol{T}^{(D),\mathrm{H}}$ is a block tridiagonal matrix, a special case of block Hessenberg matrices. ■

Usually we use *block Krylov methods* (e. g. [9]) to generate block Hessenberg matrices by similarity transformation. If we look at Equation (23), we see that the term $\boldsymbol{S}_{\ell-1}$ is strongly related to the *block Arnoldi algorithm* [10], [9]. The only difference is that the latter one performs a reduced LQ-decomposition of $\boldsymbol{S}\left(\boldsymbol{Q}_{\ell-1}\right) = \boldsymbol{L}_\ell \boldsymbol{Q}_\ell$ for $\ell \in \{2, 3, \ldots, d\}$ and $\boldsymbol{R}_{\mathbf{y},\mathbf{x}}^{\mathrm{H}} = \boldsymbol{L}_1 \boldsymbol{Q}_1$ for $\ell = 1$, instead of the decomposition $\boldsymbol{S}_{\ell-1} = \left(\boldsymbol{S}_{\ell-1} \boldsymbol{S}_{\ell-1}^{\mathrm{H}}\right)^{1/2} \boldsymbol{T}_\ell$ for $\ell \in \{2, 3, \ldots, d\}$ and $\boldsymbol{R}_{\mathbf{y},\mathbf{x}}^{\mathrm{H}} = \left(\boldsymbol{R}_{\mathbf{y},\mathbf{x}}^{\mathrm{H}} \boldsymbol{R}_{\mathbf{y},\mathbf{x}}\right)^{1/2} \boldsymbol{T}_1$ for $\ell = 1$, in order to generate the prefilter matrices with orthonormal rows, i. e. $\boldsymbol{Q}_\ell$ or $\boldsymbol{T}_\ell$, respectively. Due to the reduced LQ-decomposition, the matrix $\boldsymbol{T}_{\mathrm{Krylov}}^{(D)} \boldsymbol{R}_{\mathbf{y}} \boldsymbol{T}_{\mathrm{Krylov}}^{(D),\mathrm{H}}$ with the Krylov prefilter matrix

$$\boldsymbol{T}_{\mathrm{Krylov}}^{(D)} = \left[\boldsymbol{Q}_1^{\mathrm{T}}, \boldsymbol{Q}_2^{\mathrm{T}}, \ldots, \boldsymbol{Q}_d^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{C}^{D \times N}, \tag{27}$$

[2]Again, the *reduced-rank MWF* is a matrix whose rows are reduced-rank vector WFs.

[3]A block tridiagonal matrix consists of block matrices in the main block diagonal and the two adjacent block subdiagonals. In our case, the block matrices are elements of $\mathbb{C}^{M \times M}$.

[4]A lower block Hessenberg matrix is a lower block triangular matrix with additional non-zero block entries in the first block subdiagonal.

has no longer lower block Hessenberg structure but *lower band Hessenberg structure*[5] because in this case

$$\boldsymbol{Q}_i \boldsymbol{R}_{\mathbf{y}} \boldsymbol{Q}_\ell^{\mathrm{H}} = \begin{cases} \boldsymbol{L}_\ell, & \text{for } i = \ell - 1, \\ \mathbf{0}_{M\times M}, & \text{for } i < \ell - 1. \end{cases} \tag{28}$$

Proposition 2: The complex conjugate rows of $\boldsymbol{T}^{(D)}$ and $\boldsymbol{T}^{(D)}_{\text{Krylov}}$ span both the subspace

$$\mathcal{K}^{(D)}_{\text{block}} = \operatorname{span}\left\{\left[\boldsymbol{R}_{\mathbf{y},\mathbf{x}}, \boldsymbol{R}_{\mathbf{y}}\boldsymbol{R}_{\mathbf{y},\mathbf{x}}, \ldots, \boldsymbol{R}_{\mathbf{y}}^{d-1}\boldsymbol{R}_{\mathbf{y},\mathbf{x}}\right]\right\}, \tag{29}$$

which is the Krylov subspace [9].

Proof: First, the columns of $\boldsymbol{T}^{(D)}_{\text{Krylov}}$ span the Krylov subspace if each $\boldsymbol{Q}_\ell^{\mathrm{H}}$ can be expressed by the matrix polynomial of degree $\ell - 1$, i. e.

$$\boldsymbol{Q}_\ell^{\mathrm{H}} = \boldsymbol{\Psi}^{(\ell-1)}(\boldsymbol{R}_{\mathbf{y}}) := \sum_{k=0}^{\ell-1} \boldsymbol{R}_{\mathbf{y}}^k \boldsymbol{R}_{\mathbf{y},\mathbf{x}} \boldsymbol{\Psi}_k, \tag{30}$$

where the matrix weights $\boldsymbol{\Psi}_k \in \mathbb{C}^{M\times M}$, $\boldsymbol{\Psi}_{\ell-1} \neq \mathbf{0}_{M\times M}$. This can be shown by induction on ℓ as follows. The result is clearly true for $\ell = 1$ since $\boldsymbol{Q}_1^{\mathrm{H}} = \boldsymbol{R}_{\mathbf{y},\mathbf{x}}\boldsymbol{\Psi}_0$ with $\boldsymbol{\Psi}_0 := \boldsymbol{L}_1^{\mathrm{H},-1}$. We have to show that $\boldsymbol{Q}_\ell^{\mathrm{H}} = \boldsymbol{\Psi}^{(\ell-1)}(\boldsymbol{R}_{\mathbf{y}})$ if we assume that $\boldsymbol{Q}_k^{\mathrm{H}} = \boldsymbol{\Psi}^{(k-1)}(\boldsymbol{R}_{\mathbf{y}})$ for $k < \ell$. With Equation (23) and the definition $\boldsymbol{H}_{k,\ell-1} := \boldsymbol{Q}_k \boldsymbol{R}_{\mathbf{y}} \boldsymbol{Q}_{\ell-1}^{\mathrm{H}} \in \mathbb{C}^{M\times M}$, we get

$$\boldsymbol{Q}_\ell^{\mathrm{H}} \boldsymbol{L}_\ell^{\mathrm{H}} = \boldsymbol{S}^{\mathrm{H}}(\boldsymbol{Q}_{\ell-1}) = \boldsymbol{R}_{\mathbf{y}} \boldsymbol{Q}_{\ell-1}^{\mathrm{H}} - \sum_{k=1}^{\ell-1} \boldsymbol{Q}_k^{\mathrm{H}} \boldsymbol{H}_{k,\ell-1}. \tag{31}$$

Using the induction assumption yields $\boldsymbol{Q}_\ell^{\mathrm{H}} = \boldsymbol{\Psi}^{(\ell-1)}(\boldsymbol{R}_{\mathbf{y}})$ since the degree of $\boldsymbol{Q}_{\ell-1}^{\mathrm{H}} = \boldsymbol{\Psi}^{(\ell-2)}(\boldsymbol{R}_{\mathbf{y}})$ is increased by one due to the multiplication by $\boldsymbol{R}_{\mathbf{y}}$ on the left hand side.

It remains to show that the columns of $\boldsymbol{T}^{(D)}$ span the same subspace as the columns of $\boldsymbol{T}^{(D)}_{\text{Krylov}}$. This is true if $\operatorname{span}\{\boldsymbol{T}_\ell^{\mathrm{H}}\} = \operatorname{span}\{\boldsymbol{Q}_\ell^{\mathrm{H}}\}$ for $\ell \in \{1, 2, \ldots, d\}$ which is again proven by induction on ℓ. For $\ell = 1$, it can easily be seen that $\operatorname{span}\{\boldsymbol{T}_1^{\mathrm{H}}\} = \operatorname{span}\{\boldsymbol{R}_{\mathbf{y},\mathbf{x}}\} = \operatorname{span}\{\boldsymbol{Q}_1^{\mathrm{H}}\}$. If we assume $\operatorname{span}\{\boldsymbol{T}_k^{\mathrm{H}}\} = \operatorname{span}\{\boldsymbol{Q}_k^{\mathrm{H}}\}$ for $k < \ell$ and use Equation (23) to get

$$\boldsymbol{S}^{\mathrm{H}}(\boldsymbol{T}_{\ell-1}) = \boldsymbol{R}_{\mathbf{y}} \boldsymbol{T}_{\ell-1}^{\mathrm{H}} - \sum_{k=1}^{\ell-1} \boldsymbol{T}_k^{\mathrm{H}} \left(\boldsymbol{T}_k \boldsymbol{R}_{\mathbf{y}} \boldsymbol{T}_{\ell-1}^{\mathrm{H}}\right), \tag{32}$$

we see that $\operatorname{span}\{\boldsymbol{S}^{\mathrm{H}}(\boldsymbol{Q}_{\ell-1})\} = \operatorname{span}\{\boldsymbol{S}^{\mathrm{H}}(\boldsymbol{T}_{\ell-1})\}$ since the columns of the product of $\boldsymbol{R}_{\mathbf{y}}$ and $\boldsymbol{T}_{\ell-1}^{\mathrm{H}}$ or $\boldsymbol{Q}_{\ell-1}^{\mathrm{H}}$, respectively, span the same subspace. This concludes the proof of Proposition 2. ■

If we recall once again that auto-correlation matrices are Hermitian, the block Arnoldi algorithm can be replaced by the computational cheaper *block Lanczos algorithm* [11], [12], [9], i. e.

$$\boldsymbol{L}_\ell \boldsymbol{Q}_\ell = \boldsymbol{Q}_{\ell-1} \boldsymbol{R}_{\mathbf{y}} \left(\mathbf{1}_N - \boldsymbol{Q}_{\ell-2}^{\mathrm{H}} \boldsymbol{Q}_{\ell-2} - \boldsymbol{Q}_{\ell-1}^{\mathrm{H}} \boldsymbol{Q}_{\ell-1}\right). \tag{33}$$

Note that contrary to $\boldsymbol{T}_\ell$, $\ell \in \{1, 2, \ldots, d\}$, the prefilter matrices $\boldsymbol{Q}_\ell$ are no solution of the optimization problem given in Equation (21) because compared to $\left(\boldsymbol{S}_{\ell-1}\boldsymbol{S}_{\ell-1}^{\mathrm{H}}\right)^{1/2}$ for $\ell \in \{2, 3, \ldots, d\}$ and $\left(\boldsymbol{R}_{\mathbf{y},\mathbf{x}}^{\mathrm{H}}\boldsymbol{R}_{\mathbf{y},\mathbf{x}}\right)^{1/2}$ for $\ell = 1$, the lower triangular matrix $\boldsymbol{L}_\ell$ of the LQ-decomposition is not Hermitian. Nevertheless, the Krylov prefilter matrix $\boldsymbol{T}^{(D)}_{\text{Krylov}}$ followed by a reduced-dimension MWF consisting of d quadratic MWFs, produces the same output as the MSMWF derived in Section II since they are MMSE approximations of the MWF in the same subspace, i. e. the Krylov subspace $\mathcal{K}^{(D)}_{\text{block}}$.

[5] A lower band Hessenberg matrix is a lower block Hessenberg matrix where the non-zero block entries in the first block subdiagonal are lower triangular matrices.

IV. Flexible Rank Selection of Lanczos-Ruhe based MSMWF

Remember that the dimension N of the observation vector and the rank D of the reduced-rank MSMWF are restricted to be integer multiples of M. The restriction on N leads to either a very small subset of systems where the MSMWF can be applied or to an unnecessary increase in computational complexity if the observation vector is filled with zeros such that $N = LM$. Moreover, the restriction on D bears the problem that the optimum rank cannot be achieved if it is no integer multiple of M.

Ruhe [14], [9] derived a version of the block Lanczos algorithm where the complex conjugate rows of the Krylov prefilter matrix $\boldsymbol{T}^{(D)}_{\text{Krylov}}$ are computed step by step. Therefore, the rank D can be any integer between M and N. Algorithm I summarizes the Lanczos-Ruhe computation of the Krylov prefilter matrix which is used as the prefilter of the MSMWF in the simulations of Section V. The function $\operatorname{QR}\{\boldsymbol{R}_{\mathbf{y},\mathbf{x}}\}$ denotes the QR-decomposition of the cross-correlation matrix $\boldsymbol{R}_{\mathbf{y},\mathbf{x}}$. Note that the resulting upper triangular matrix $\boldsymbol{R}$ can be discarded since it is not needed for the remaining steps. Consequently, the dimension N of the observation vector is no longer limited to integer multiples of M which offers the application of the proposed method to more general systems.

ALGORITHM I

Lanczos-Ruhe Computation of the Multi-Stage Prefilter

```
{[q_1, q_2, ..., q_M], R} ← QR{R_y,x}
for i ∈ {M, M+1, ..., D−1} do
    k ← i − M + 1
    v ← R_y q_k
    for ℓ ∈ {i−2M+1, i−2M+2, ..., i} ∩ ℕ do
        h_ℓ,k ← v^H q_ℓ
        v ← v − h_ℓ,k q_ℓ
    end for
    h_i+1,k ← ||v||_2
    q_i+1 ← v / h_i+1,k
end for
T_Ruhe^(D),H ← [q_1, q_2, ..., q_D] ∈ ℂ^(N×D)
```

V. Application to a Time-Dispersive MIMO System

A. Channel Model

The time-dispersive MIMO system with S inputs, R outputs and Q propagation paths is described by the channel matrix impulse response

$$\boldsymbol{H}[n] = \sum_{q=0}^{Q-1} \boldsymbol{H}_q \delta[n-q] \in \mathbb{C}^{R\times S}, \quad (34)$$

with the unit impulse function $\delta[n]$. The weighting matrices of the propagation paths, $\boldsymbol{H}_q$, $q \in \{0,1,\ldots,Q-1\}$, are realizations of the random variable [18]

$$\mathsf{H} = \boldsymbol{U}_{\mathrm{Rx}} \boldsymbol{\Sigma}_{\mathrm{Rx}} \mathsf{Z} \boldsymbol{\Sigma}_{\mathrm{Tx}} \boldsymbol{V}_{\mathrm{Tx}}^{\mathrm{H}} \in \mathbb{C}^{R\times S}. \quad (35)$$

The matrix $\mathsf{Z} \in \mathbb{C}^{R\times S}$ has i. i. d. random entries with $\mathcal{N}_c\left(0, \sigma_{\mathsf{z}}^2\right)$. The columns of the unitary matrix $\boldsymbol{U}_{\mathrm{Rx}} \in \mathbb{C}^{R\times R}$ are eigenmodes of the correlations at the receiver, i. e. $\mathrm{E}\left\{\mathsf{H}\mathsf{H}^{\mathrm{H}}\right\} = \sigma_{\mathsf{z}}^2 \operatorname{tr}\left\{\boldsymbol{\Sigma}_{\mathrm{Tx}}^2\right\} \boldsymbol{U}_{\mathrm{Rx}} \boldsymbol{\Sigma}_{\mathrm{Rx}}^2 \boldsymbol{U}_{\mathrm{Rx}}^{\mathrm{H}}$ with the power distribution described by the diagonal matrix of eigenvalues $\boldsymbol{\Sigma}_{\mathrm{Rx}}^2 \in \mathbb{R}_{0,+}^{R\times R}$. Analogous, the unitary matrix $\boldsymbol{V}_{\mathrm{Tx}} \in \mathbb{C}^{S\times S}$ and the diagonal matrix $\boldsymbol{\Sigma}_{\mathrm{Tx}}^2 \in \mathbb{R}_{0,+}^{S\times S}$ from the eigenvalue decomposition of $\mathrm{E}\left\{\mathsf{H}^{\mathrm{H}}\mathsf{H}\right\} = \sigma_{\mathsf{z}}^2 \operatorname{tr}\left\{\boldsymbol{\Sigma}_{\mathrm{Rx}}^2\right\} \boldsymbol{V}_{\mathrm{Tx}} \boldsymbol{\Sigma}_{\mathrm{Tx}}^2 \boldsymbol{V}_{\mathrm{Tx}}^{\mathrm{H}}$ depict the correlations at the transmitter.

The received signal vector

$$\mathbf{r}[n] = \boldsymbol{H}[n] * \mathbf{s}[n] + \mathbf{n}[n] \in \mathbb{C}^{R}, \quad (36)$$

where '$*$' denotes convolution, is perturbed by additive white Gaussian noise $\mathbf{n}[n] \in \mathbb{C}^R$ with the complex normal distribution $\mathcal{N}_c\left(\mathbf{0}_{R\times 1}, \sigma_{\mathsf{n}}^2 \mathbf{1}_R\right)$. The transmit signal vector $\mathbf{s}[n]$ at time index n is composed of S zero-mean i. i. d. symbols with variance σ_{s}^2. Note that the total transmit power is $P_{\mathrm{Tx}} = \operatorname{tr}\left\{\boldsymbol{R}_{\mathsf{s}}\right\} = S\sigma_{\mathsf{s}}^2$.

In order to compute the linear equalizer filter of length K, we derive an alternative matrix-vector model of the time-dispersive MIMO channel. The vector $\tilde{\mathbf{r}}[n] = \left[\mathbf{r}^{\mathrm{T}}[n], \mathbf{r}^{\mathrm{T}}[n-1], \ldots, \mathbf{r}^{\mathrm{T}}[n-K+1]\right]^{\mathrm{T}} \in \mathbb{C}^{KR}$ is composed of K adjacent received signal vectors $\mathbf{r}[n]$. Using the block Toeplitz matrix

$$\tilde{\boldsymbol{H}} = \sum_{q=0}^{Q-1} \boldsymbol{S}_{(q,K,Q-1)} \otimes \boldsymbol{H}_q \in \mathbb{C}^{KR\times(K+Q-1)S}, \quad (37)$$

where '$\otimes$' denotes the Kronecker product and $\boldsymbol{S}_{(q,K,Q-1)} = \left[\mathbf{0}_{K\times q}, \mathbf{1}_K, \mathbf{0}_{K\times(Q-q-1)}\right] \in \{0,1\}^{K\times(K+Q-1)}$ the selection matrix, Equation (36) may be rewritten as

$$\tilde{\mathbf{r}}[n] = \tilde{\boldsymbol{H}} \tilde{\mathbf{s}}[n] + \tilde{\mathbf{n}}[n] \in \mathbb{C}^{KR}. \quad (38)$$

Analogous to $\tilde{\mathbf{r}}[n]$, the vector $\tilde{\mathbf{s}}[n] \in \mathbb{C}^{(K+Q-1)S}$ is composed of $K+Q-1$ adjacent transmit signal vectors $\mathbf{s}[n]$ and $\tilde{\mathbf{n}}[n] \in \mathbb{C}^{KR}$ of K adjacent noise vectors $\mathbf{n}[n]$. The filter algorithms of the previous sections may be applied to the given MIMO scenario, if we set $\mathbf{y}[n] = \tilde{\mathbf{r}}[n]$ and $\mathbf{x}[n] = \mathbf{s}[n-\nu]$ where ν is the latency time introduced by the equalizer. Consequently, $N = KR$ and $M = S$.

B. Simulation Results

In this section, we apply the algorithms of Section III and IV to a frequency-selective MIMO system with $R = 8$ outputs and $S = 4$ inputs. A transmission of QPSK symbols with variance $\sigma_{\mathsf{x}}^2 = 1$ is considered. The channel impulse response has $Q = 3$ i. i. d. taps with uncorrelated normal distributed entries, i. e. they are realizations of H where $\boldsymbol{\Sigma}_{\mathrm{Tx}}^2 = \mathbf{1}_S/S$ and $\boldsymbol{\Sigma}_{\mathrm{Rx}}^2 = \mathbf{1}_R/R$. The unitary matrices $\boldsymbol{U}_{\mathrm{Rx}}$ and $\boldsymbol{V}_{\mathrm{Tx}}$ are chosen arbitrary for every realization of H. In order to keep the channel amplification comparable to a frequency-flat MIMO channel with zero-mean i. i. d. complex normal distributed random entries of variance one, we set $\sigma_{\mathsf{z}}^2 = RS/Q = 32/3$ because $\operatorname{tr}\left\{\boldsymbol{\Sigma}_{\mathrm{Rx}}^2\right\} = \operatorname{tr}\left\{\boldsymbol{\Sigma}_{\mathrm{Tx}}^2\right\} = 1$. The filter length is $K = 4$ with latency $\nu = 2$, thus, the observation vector has the dimension $N = KR = 32$.

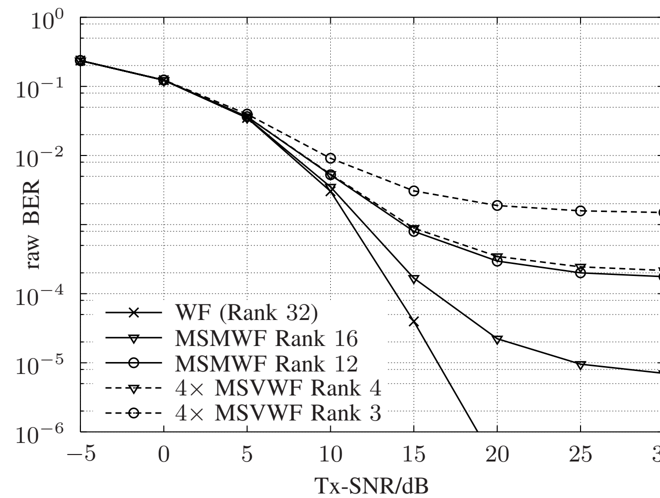

Fig. 3. BER comparison between the MSMWF and the MSVWF with a comparable computational complexity for known statistics

Figure 3 shows the uncoded *Bit Error Rate* (BER) comparison between the MSMWF and a filter structure where each element of the transmit vector $\mathbf{x}[n] = [\mathsf{x}_1[n], \mathsf{x}_2[n], \ldots, \mathsf{x}_M[n]]^{\mathrm{T}}$ is estimated by a *Multi-Stage Vector WF* (MSVWF) [5] over the *Transmit Signal-to-Noise Ratio* (Tx-SNR) $10\lg(P_{\mathrm{Tx}}/\sigma_{\mathsf{n}}^2)$ if statistics are perfectly known. The results are averaged over several thousands channel realizations. Since we choose the rank D of the MSMWF to be $D = dM$, where d is the reduced rank of each MSVWF, the determination of the prefilter matrices for both methods has the same order of computational complexity. Note that for the derivation of the MSVWFs a $d\times d$ tridiagonal matrix has to be inverted M times instead of one inversion of a $D\times D$ band matrix of width $2M+1$ in the case of the MSMWF. Compared to the computation of the prefilter matrices, even the inversion of the band matrix has a negligible computational complexity. It can be seen that the MSMWF with rank $D = 12$ outperforms the implementation of $M = 4$ parallel MSVWFs with rank $d = 3$ as well as with rank $d = 4$.

In Figure 4, we compare the conventional MSMWF to its Lanczos-Ruhe implementation in the previously defined MIMO scenario. Again, statistics are perfectly known. It can be seen that the new algorithm has more flexibility concerning rank selection and that both algorithms are good approximations of the MWF despite the tremendous reduction in computational complexity.

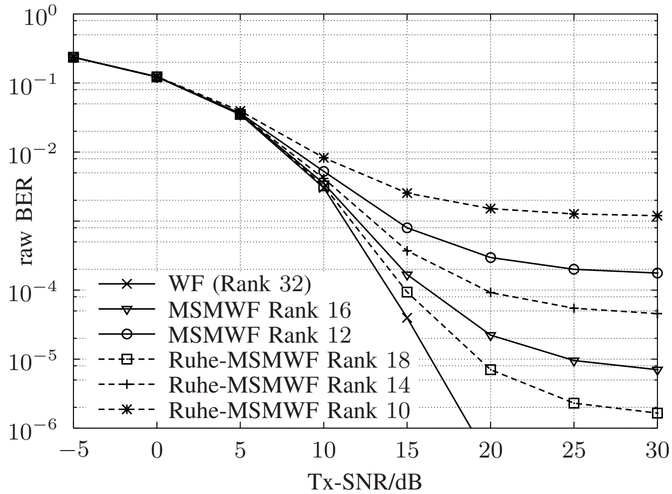

Fig. 4. BER comparison between the MSMWF and its Ruhe implementation for known statistics

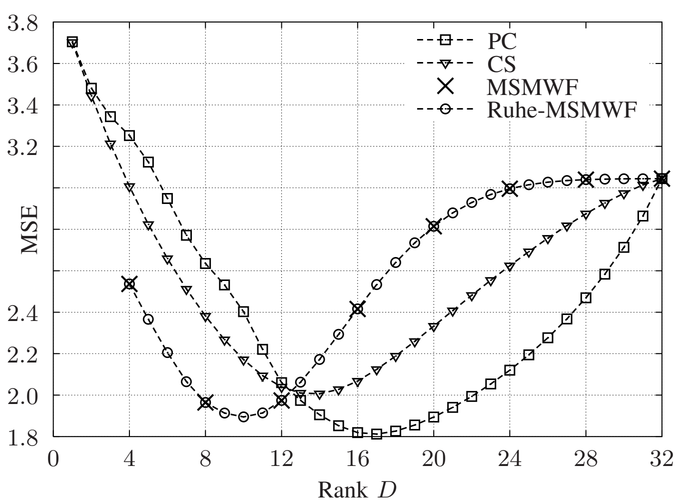

Fig. 5. MSE comparison between the MSMWF, its Ruhe version, and the eigenspace based reduced-rank MWFs for estimated statistics (SNR $= 10\,\text{dB}$)

The MSE is plotted over the rank D in Figure 5 where we estimated the cross-correlation matrix with 50 and the auto-correlation matrix with 100 training symbols. In this case, we choose a strongly correlated MIMO channel where $\boldsymbol{\Sigma}_{\text{Rx}}^2 = \textbf{diag}\{0.7, 0.1, 0.1, 0.05, 0.02, 0.01, 0.01, 0.01\}$ and $\boldsymbol{\Sigma}_{\text{Tx}}^2 = \textbf{diag}\{0.8, 0.1, 0.05, 0.05\}$. Note that due to its flexibility, only the Lanczos-Ruhe implementation of the MSMWF offers access to the optimum rank of $D_{\text{opt}} = 10$ which is not an integer multiple of $M = 4$. Moreover, the MSMWF achieves its optimum with a smaller rank than the eigenspace based PC and CS method. Nevertheless, the latter may also have ranks smaller than $M = 4$ and the PC algorithm achieves the smallest MSE of all at the expense of a higher computational complexity. Since we are interested in computational cheap implementations, the Lanczos-Ruhe based MSMWF has the best performance in case of low sample support. Note that the MSE ripples of the PC method at multiple ranks of M occur due to the eigenvalue profile of the auto-correlation matrix given by $\boldsymbol{\Sigma}_{\text{Rx}}^2$.

VI. Conclusions

In this paper, we derived the relationship between the MSMWF and block Krylov methods, especially the block Lanczos algorithm and used its Ruhe version in order to improve the flexibility in rank selection. Simulation results of an application to a frequency-selective MIMO system showed that only the Lanczos-Ruhe based MSMWF achieves the optimum rank in case of estimated statistics. Moreover, the MSMWF outperforms the parallel implementation of MSVWFs with a comparable computational complexity.

References

[1] L. L. Scharf, *Statistical Signal Processing*, Addison-Wesley, 1991.

[2] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*, Prentice Hall, 1999.

[3] H. Hotelling, "Analysis of a Complex of Statistical Variables into Principal Components," *Journal of Educational Psychology*, vol. 24, no. 6/7, pp. 417–441, 498–520, September/October 1933.

[4] J. S. Goldstein and I. S. Reed, "Reduced-Rank Adaptive Filtering," *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 492–496, February 1997.

[5] J. S. Goldstein, I. S. Reed, and L. L. Scharf, "A Multistage Representation of the Wiener Filter Based on Orthogonal Projections," *IEEE Transactions on Information Theory*, vol. 44, no. 7, pp. 2943–2959, November 1998.

[6] M. Joham, Y. Sun, M. D. Zoltowksi, M. Honig, and J. S. Goldstein, "A New Backward Recursion for the Multi-Stage Nested Wiener Filter Employing Krylov Subspace Methods," in *Proc. Milcom 2001*, October 2001.

[7] J. S. Goldstein, I. S. Reed, D. E. Dudgeon, and J. R. Guerci, "A Multistage Matrix Wiener Filter for Subspace Detection," in *Proc. IT Workshop on Detection, Estimation and Classification and Imaging*, February 1999.

[8] P. Cifuentes, W. L. Myrick, S. Sud, J. S. Goldstein, and M. D. Zoltowski, "Reduced Rank Matrix Multistage Wiener Filter with Applications in MMSE Joint Multiuser Detection for DS-CDMA," in *Proc. ICASSP 2002*, May 2002, vol. 3, pp. 2605–2608.

[9] Y. Saad, *Iterative Methods for Sparse Linear Systems*, SIAM, 2003.

[10] W. E. Arnoldi, "The Principle of Minimized Iterations in the Solution of the Matrix Eigenvalue Problem," *Quarterly of Applied Mathematics*, vol. 9, no. 1, pp. 17–29, January 1951.

[11] C. Lanczos, "An Iteration Method for the Solution of the Eigenvalue Problem of Linear Differential and Integral Operators," *Journal of Research of the National Bureau of Standards*, vol. 45, no. 4, pp. 255–282, October 1950.

[12] G. H. Golub and R. Underwood, "The block Lanczos method for computing eigenvalues," in *Mathematical Software III*, J. R. Rice, Ed., pp. 361–377. Academic Press, 1977.

[13] G. Golub and C. V. Loan, *Matrix Computations*, Johns Hopkins University Press, 1996.

[14] A. Ruhe, "Implementation Aspects of Band Lanczos Algorithms for Computation of Eigenvalues of Large Sparse Symmetric Matrices," *Mathematics of Computation*, vol. 33, no. 146, pp. 680–687, April 1979.

[15] S. P. Applebaum and D. J. Chapman, "Adaptive Arrays with Main Beam Constraints," *IEEE Transactions on Antennas and Propagation*, vol. 24, no. 5, September 1976.

[16] L. J. Griffiths and C. W. Jim, "An Alternative Approach to Linearly Constrained Adaptive Beamforming," *IEEE Transactions on Antennas and Propagation*, vol. 30, no. 1, January 1982.

[17] G. Dietl, M. D. Zoltowski, and M. Joham, "Reduced-Rank Equalization for EDGE Via Conjugate Gradient Implementation of Multi-Stage Nested Wiener Filter," in *Proc. VTC Fall 2001*, October 2001, pp. 1912–1916.

[18] J. P. Kermoal, L. Schumacher, K. I. Pedersen, P. E. Mogensen, and F. Frederiksen, "A Stochastic MIMO Radio Channel Model With Experimental Validation," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 6, pp. 1211–1226, August 2002.