Reduced-Complexity Linear and Nonlinear Precoding for Frequency-Selective MIMO Channels

Johannes Brehmer, Guido Dietl, Michael Joham, Wolfgang Utschick
Institute for Circuit Theory and Signal Processing
Munich University of Technology, 80333 Munich, Germany
Email: jobr@nws.ei.tum.de

***Abstract*—Focusing on systems with simple non-cooperating receivers, we present an approach to reduce the complexity of multi-user precoding schemes for frequency-selective MIMO channels. The *transmit Wiener filter* (TxWF) and Wiener *Tomlinson-Harashima precoding* (THP) both provide attractive performance, but the complexity involved in their computation may be prohibitive. We develop the reduced-complexity multi-user *multi-stage TxWF* (MSTxWF) by applying the multi-stage decomposition known from receive processing to the TxWF. We show that the block-Lanczos algorithm can be used for efficiently computing the reduced rank MSTxWF. Moreover, we extend the MSTxWF approach to Wiener temporal THP, resulting in a reduced-complexity nonlinear precoding scheme. Simulation results demonstrate that both schemes can provide close to optimum performance at significantly reduced complexity.**

I. Introduction

We consider a point-to-multipoint wireless communication system with a multi-antenna transmitter and low-complexity single-antenna receivers. In such a system, *precoding* (or *pre-equalization*) represents the adequate transmission strategy if (at least partial) *channel state information* (CSI) is available to the transmitter. The general case of a frequency-selective channel is investigated. Compared to frequency-flat scenarios, the frequency-selectivity of the channel leads to a substantial increase in the dimension of the precoding problem. However, in practical scenarios, the computational resources available at the transmitter are limited. Thus, not only the performance but also the complexity of the chosen precoding scheme deserves careful examination.

Among linear precoding schemes for frequency-selective channels, the FIR *transmit Wiener filter* (TxWF, [1]) provides an attractive trade-off between performance and complexity. Its performance is optimum in a sum of *mean squared error* (MSE) sense. Concerning complexity, its main advantage lies in the existence of a closed-form solution, resulting in a considerably lower complexity than linear precoding schemes based on SINR criteria, which usually only provide an iterative solution (e.g., [2], [3]). Still, even the complexity involved in the computation of the TxWF may be too high in practical applications, motivating the work presented in this paper.

In receive Wiener filtering, reduced rank processing is a well-known approach to complexity reduction. The basic idea is to reduce complexity by approximating the optimum Wiener solution in a lower dimensional subspace. Clearly, we desire a subspace basis that can be computed efficiently while providing good performance at low rank. Among the reduced rank techniques known from single-user receive processing, the *multi-stage Wiener filter* (MSWF) introduced by Goldstein et. al. [4] is one of the most promising. In [5], we developed the single-user (or vector) MSTxWF by applying the multi-stage concept to the single-user TxWF. Although the derivation in [5] is based on the single-user case, it can be applied to multi-user scenarios by employing multiple vector MSTxWFs in parallel. However, this setting is clearly *sub-optimum*. The first contribution of this work is the generalization of the concepts from [5] to the multi-user case, resulting in the development of the multi-user (or matrix) MSTxWF. First, we apply the multi-stage concept to find an algorithm for iteratively computing a basis that has good properties. It turns out that this basis can be computed with the block-Lanczos algorithm, showing the relationship between the MSTxWF and block-Krylov methods. In a second step, the reduced rank solution is computed in the subspace spanned by this basis. Finally, we derive an algorithm which efficiently combines the computation of subspace and reduced rank solution.

Nonlinear *Tomlinson-Harashima precoding* (THP, [6], [7], [8], [9]) can be considered as an extension to linear precoding by adding a feedback filter at the transmitter and modulo operators at both the transmitter and the receivers. In frequency-selective multi-user scenarios, three variants of THP can be distinguished [10]: THP can be employed to mitigate interference from symbols sent at an earlier time instant (temporal THP), to combat multi-user interference (spatial THP), or in its most general form to mitigate both types of interference (spatio-temporal THP). We show that the MSTxWF approach can also be applied to MSE-optimum temporal THP (Wiener T-THP). In highly loaded scenarios, the resulting nonlinear precoding scheme provides optimum performance at low complexity.

Throughout this work, we assume the transmitter to have full CSI, a valid assumption in time division duplex systems if the coherence time of the channel is large enough.

The remainder of this paper is organized as follows: In Section II we present our system model. The TxWF and Wiener T-THP are briefly reviewed in Section III. The reduced rank MSTxWF is developed in Section IV. In Section V, we discuss the application of the MSTxWF approach to Wiener T-THP. Simulation results are presented in Section VI, conclusions are provided in Section VII.

0-7803-8521-7/04/$20.00 © 2004 IEEE

A. Notation

Vectors and matrices are denoted by lower case bold and capital bold letters, respectively. We use $\mathrm{E}[\bullet]$, $\|\bullet\|_{\mathrm{F}}$, '$\otimes$', $(\bullet)^*$, $(\bullet)^{\mathrm{T}}$, and $(\bullet)^{\mathrm{H}}$ for expectation, Frobenius norm, Kronecker product, complex conjugation, transposition, and conjugate transposition, respectively. All random processes are assumed to be zero-mean and stationary. Let $\sigma_x^2 = \mathrm{E}\left[|x[n]|^2\right]$ and $\sigma_{\boldsymbol{x}}^2 = \mathrm{E}\left[\|\boldsymbol{x}[n]\|_2^2\right]$ denote the variance of a scalar process $x[n]$ and a vector process $\boldsymbol{x}[n]$, respectively. The $N \times M$ zero matrix is $\boldsymbol{0}_{N\times M}$, the $M\times 1$ zero vector is $\boldsymbol{0}_M$, and the $N\times N$ identity matrix is $\boldsymbol{1}_N$, whose n-th column is $\boldsymbol{e}_n$.

II. System Model

The system under consideration consists of a N_{a}-antenna transmitter, a FIR MIMO channel

$$\boldsymbol{H}[n] = \sum_{q=0}^{Q} \tilde{\boldsymbol{H}}_q \delta[n-q], \quad \tilde{\boldsymbol{H}}_q \in \mathbb{C}^{K\times N_{\mathrm{a}}},$$

and K non-cooperating single-antenna receivers. The general model depicted in Fig. 1 is applicable to both linear precoding and THP. The feedforward filter $\boldsymbol{P}[n]$ is constrained to be FIR:

$$\boldsymbol{P}[n] = \sum_{\ell=0}^{L} \boldsymbol{P}_\ell \delta[n-\ell], \quad \boldsymbol{P}_\ell \in \mathbb{C}^{N_{\mathrm{a}}\times K}.$$

For linear precoding, the feedback filter is inactive, $\boldsymbol{F}[n] = \boldsymbol{0}_{K\times K}$. Moreover, $\boldsymbol{a}[n] = \tilde{\boldsymbol{a}}[n] = \boldsymbol{0}_K$. For temporal THP,

$$\boldsymbol{F}[n] = \sum_{r=1}^{R} \boldsymbol{F}_r \delta[n-r], \quad \boldsymbol{F}_r \in \mathbb{C}^{K\times K},$$

while $\boldsymbol{a}[n]$ and $\tilde{\boldsymbol{a}}[n]$ are chosen such that $\boldsymbol{v}[n]$ and $\hat{\boldsymbol{s}}[n]$ lie within the feasible region of the THP modulo operation (see, e.g., [9], [10]).

The received signal is scaled by a factor β^{-1} at each receiver (gain control). Note that all receivers employ the same scaling factor. This simplification ensures the existence of a closed form-solution for the TxWF.

After defining

$$\boldsymbol{P} = \left[\boldsymbol{P}_0^{\mathrm{T}}, \ldots, \boldsymbol{P}_L^{\mathrm{T}}\right] \in \mathbb{C}^{K\times N_{\mathrm{a}}(L+1)}, \quad \text{and}$$
$$\boldsymbol{v}_k[n] = \left[\boldsymbol{e}_k^{\mathrm{T}}\boldsymbol{v}[n], \ldots, \boldsymbol{e}_k^{\mathrm{T}}\boldsymbol{v}[n-Q-L]\right]^{\mathrm{T}} \in \mathbb{C}^{Q+L+1},$$

the estimate $\hat{\boldsymbol{d}}[n]$ can be written as

$$\hat{\boldsymbol{d}}[n] = \beta^{-1}\sum\nolimits_{k,i=1}^{K} \boldsymbol{e}_k \boldsymbol{e}_i^{\mathrm{T}} \boldsymbol{P}\boldsymbol{H}_k \boldsymbol{v}_i[n] + \beta^{-1}\boldsymbol{\eta}[n].$$

The block-Toeplitz matrix $\boldsymbol{H}_k$ is constructed from the coefficients of the channel to the k-th receiver:

$$\boldsymbol{H}_k = \left[\bar{\boldsymbol{H}}_{k,1}^{\mathrm{T}}, \ldots, \bar{\boldsymbol{H}}_{k,L+1}^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{C}^{N_{\mathrm{a}}(L+1)\times Q+L+1},$$

where $\bar{\boldsymbol{H}}_{k,b} \in \mathbb{C}^{N_{\mathrm{a}}\times Q+L+1}$ is given by

$$\bar{\boldsymbol{H}}_{k,b} = \left[\boldsymbol{0}_{N_{\mathrm{a}}\times b-1}, \tilde{\boldsymbol{H}}_0^{\mathrm{T}}\boldsymbol{e}_k, \ldots, \tilde{\boldsymbol{H}}_Q^{\mathrm{T}}\boldsymbol{e}_k, \boldsymbol{0}_{N_{\mathrm{a}}\times L-b-1}\right].$$

For compactness, the matrices $\boldsymbol{H}_k$ are collected in a matrix

$$\boldsymbol{H} = [\boldsymbol{H}_1, \ldots, \boldsymbol{H}_K] \in \mathbb{C}^{N_{\mathrm{a}}(L+1)\times K(Q+L+1)}.$$

Assuming that $\mathrm{E}\left[\boldsymbol{v}[n]\boldsymbol{v}^{\mathrm{H}}[n+m]\right] = \sigma_v^2 \boldsymbol{1}_K \delta[m]$ [10], the average transmit power is given by $E_{\mathrm{avg}} = \sigma_v^2 \|\boldsymbol{P}\|_{\mathrm{F}}^2$.

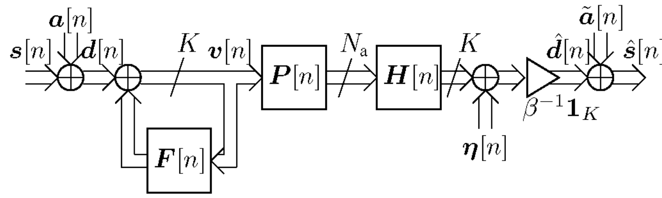

Fig. 1. Frequency-Selective MIMO System

III. TxWF and Wiener T-THP

We define the MSE as $\sigma_{\boldsymbol{\varepsilon}}^2 = \mathrm{E}[\|\boldsymbol{d}[n-\nu] - \hat{\boldsymbol{d}}[n]\|_2^2]$, with a fixed latency time ν that is chosen *a priori*. Recall that for linear precoding, $\boldsymbol{v}[n] = \boldsymbol{d}[n] = \boldsymbol{s}[n]$ and $\hat{\boldsymbol{d}}[n] = \hat{\boldsymbol{s}}[n]$.

The TxWF minimizes the MSE under an average transmit power constraint [1]:

$$\{\boldsymbol{P}_{\mathrm{WF}}, \beta_{\mathrm{WF}}\} = \underset{\{\boldsymbol{P},\beta\}}{\operatorname{argmin}}\, \sigma_{\boldsymbol{\varepsilon}}^2 \quad \text{s.t.} \quad \sigma_{\boldsymbol{s}}^2 \left\|\boldsymbol{P}\right\|_{\mathrm{F}}^2 \leq E_{\mathrm{tr}}. \tag{1}$$

The TxWF solution is given by $\boldsymbol{P}_{\mathrm{WF}} = \beta_{\mathrm{WF}}\boldsymbol{P}_0$, with

$$\begin{aligned}\boldsymbol{P}_0 &= \boldsymbol{V}_0^{\mathrm{H}}\boldsymbol{R}_0^{-1}, \quad \beta_{\mathrm{WF}} = \sqrt{E_{\mathrm{tr}}/\sigma_{\boldsymbol{s}}^2}\left\|\boldsymbol{P}_0\right\|_{\mathrm{F}}^{-1},\\ \boldsymbol{R}_0 &= \boldsymbol{H}\boldsymbol{H}^{\mathrm{H}} + \xi_{\mathrm{WF}}\boldsymbol{1}_{N_{\mathrm{a}}(L+1)}, \quad \text{and}\\ \boldsymbol{V}_0 &= \sum\nolimits_{k=1}^{K} \boldsymbol{H}_k \boldsymbol{e}_{\nu+1}\boldsymbol{e}_k^{\mathrm{T}},\end{aligned} \tag{2}$$

where $\xi_{\mathrm{WF}} = \sigma_{\boldsymbol{\eta}}^2/E_{\mathrm{tr}}$. The computation of $\boldsymbol{P}_{\mathrm{WF}}$ requires the inversion of an $N\times N$ matrix, with $N = N_{\mathrm{a}}(L+1)$. Accordingly, the full-rank solution has complexity $\mathrm{O}(N^3)$. For large N, the complexity involved in the computation of the TxWF may be prohibitive. Thus, it is desirable to develop reduced rank methods that reduce complexity without sacrificing performance.

Note that in the limit $\sigma_{\boldsymbol{\eta}}^2 \to \infty$, the TxWF converges to the *transmit matched filter* (TxMF) $\boldsymbol{P}_{\mathrm{MF}} = \beta_{\mathrm{MF}}\boldsymbol{V}_0^{\mathrm{H}}$ [1]. The TxMF is simple to compute, but does not take into account interference.

In Wiener T-THP, the feedback filter is active and of order $R = Q + L - \nu$. Accordingly,

$$\{\boldsymbol{P}_{\mathrm{WTHP}}, \boldsymbol{F}_{\mathrm{WTHP}}, \beta_{\mathrm{WTHP}}\} = \underset{\{\boldsymbol{P},\boldsymbol{F},\beta\}}{\operatorname{argmin}}\, \sigma_{\boldsymbol{\varepsilon}}^2 \quad \text{s.t.} \quad \sigma_v^2\left\|\boldsymbol{P}\right\|_{\mathrm{F}}^2 \leq E_{\mathrm{tr}}.$$

After introducing the matrix

$$\check{\boldsymbol{H}} = \left[\check{\boldsymbol{H}}_1, \ldots, \check{\boldsymbol{H}}_K\right] \in \mathbb{C}^{N_{\mathrm{a}}(L+1)\times K(\nu+1)},$$

where $\check{\boldsymbol{H}}_k$ contains the first $\nu+1$ columns of $\boldsymbol{H}_k$, the optimum feedforward filter $\boldsymbol{P}_{\mathrm{WTHP}}$ is given by $\boldsymbol{P}_{\mathrm{WTHP}} = \beta_{\mathrm{WTHP}}\check{\boldsymbol{P}}_0$, with

$$\begin{aligned}\check{\boldsymbol{P}}_0 &= \boldsymbol{V}_0^{\mathrm{H}}\check{\boldsymbol{R}}_0^{-1}, \quad \text{and}\\ \check{\boldsymbol{R}}_0 &= \check{\boldsymbol{H}}\check{\boldsymbol{H}}^{\mathrm{H}} + \xi_{\mathrm{WF}}\boldsymbol{1}_{N_{\mathrm{a}}(L+1)}.\end{aligned} \tag{3}$$

The computation of the optimum feedback filter is simple as soon as the feedforward filter is known [10]. Thus, the main complexity of Wiener T-THP lies in the computation of the optimum feedforward filter $\boldsymbol{P}_{\mathrm{WTHP}}$. In terms of Eq. (3), the feedforward filter solution is again of complexity $\mathrm{O}(N^3)$. After applying the matrix inversion lemma (see, e.g., [11]), we can also write

$$\check{\boldsymbol{P}}_0 = \check{\boldsymbol{E}}_{\nu+1}^{\mathrm{T}}\left(\check{\boldsymbol{H}}^{\mathrm{H}}\check{\boldsymbol{H}} + \xi_{\mathrm{WF}}\boldsymbol{1}_{K(\nu+1)}\right)^{-1}\check{\boldsymbol{H}}^{\mathrm{H}}, \tag{4}$$

with $\breve{\boldsymbol{E}}_{\nu+1} = \mathbf{1}_K \otimes \boldsymbol{e}_{\nu+1} \in \mathbb{C}^{K(\nu+1)\times K}$. In this case, a $M \times M$ matrix has to be inverted, with $M = K(\nu+1)$. In the following, it is assumed that

$$N_\mathrm{a}(L+1) \leq K(\nu+1). \tag{5}$$

If Eq. (5) does not hold, complexity can already be reduced by solving Eq. (4) instead of Eq. (3).

IV. Reduced rank MSTxWF

The complexity involved in the computation of the TxWF can be reduced by approximating each row of the full-rank solution $\boldsymbol{P}_\mathrm{WF}$ in a lower-dimensional subspace of $\mathbb{C}^N$. First, we aim at finding a subspace basis that provides a good trade-off between complexity of computation and quality of approximation.

A. Krylov Subspace Basis

Motivated by the receive MSWF [4], we apply a generalized stage-wise decomposition to $\boldsymbol{P}_0$: Given $\boldsymbol{P}_i \in \mathbb{C}^{K\times N-iK}$, expand $\boldsymbol{P}_i$ into

$$\boldsymbol{P}_i = \boldsymbol{A}_{i+1}\left(\boldsymbol{Q}_{i+1} - \boldsymbol{P}_{i+1}\boldsymbol{B}_{i+1}\right), \tag{6}$$

with two linearly independent bases $\boldsymbol{Q}_{i+1} \in \mathbb{C}^{K\times N-iK}$ and $\boldsymbol{B}_{i+1} \in \mathbb{C}^{N-(i+1)K\times N-iK}$, i.e.,

$$\mathrm{span}(\boldsymbol{B}_{i+1}^\mathrm{T}) = \mathrm{null}(\boldsymbol{Q}_{i+1}^*). \tag{7}$$

Eq. (6) can be understood as the expansion of $\boldsymbol{P}_i$ in terms of the bases $\boldsymbol{Q}_{i+1}$ and $\boldsymbol{B}_{i+1}$. In the next stage, $\boldsymbol{P}_{i+1}$ is expanded in the same manner. Assuming that N is an integer multiple of K, the full decomposition is obtained after the N/K-th stage:

$$\boldsymbol{P}_0 = \sum_{i=1}^{N/K} (-1)^{i+1}\left(\prod_{k=1}^{i}\boldsymbol{A}_k\right)\boldsymbol{T}_i$$

with $\boldsymbol{T}_1 = \boldsymbol{Q}_1$ and for $i > 1$,

$$\boldsymbol{T}_i = \boldsymbol{Q}_i \prod_{k=i-1}^{1}\boldsymbol{B}_k \in \mathbb{C}^{K\times N}. \tag{8}$$

A basis of a DK-dimensional subspace of $\mathbb{C}^N$ can be found by stopping the decomposition after D stages and stacking $\boldsymbol{T}_1,\ldots,\boldsymbol{T}_D$ in a matrix

$$\boldsymbol{T}^{(D)} = \left[\boldsymbol{T}_1^\mathrm{T},\ldots,\boldsymbol{T}_D^\mathrm{T}\right]^\mathrm{T} \in \mathbb{C}^{DK\times N}.$$

Up to this point, the decomposition is completely generic. The properties of the MSTxWF result from a particular choice of the matrices $\boldsymbol{Q}_i$. At the first stage, we choose the matrix $\boldsymbol{Q}_1$ such that

$$\mathrm{span}(\boldsymbol{Q}_1^\mathrm{T}) = \mathrm{span}(\boldsymbol{V}_0^*).$$

With this choice, it is ensured that the reduced rank MSTxWF performs as least as good as the TxMF. Given $\boldsymbol{P}_0 = \boldsymbol{V}_0^\mathrm{H}\boldsymbol{R}_0$, $\boldsymbol{Q}_1$ and a matrix $\boldsymbol{B}_1$ that satisfies Eq. (7), the matrices $\boldsymbol{A}_1$ and $\boldsymbol{P}_1$ can be found by solving

$$\boldsymbol{V}_0^\mathrm{H}\boldsymbol{R}_0 = \boldsymbol{A}_1\left(\boldsymbol{Q}_1 - \boldsymbol{P}_1\boldsymbol{B}_1\right)$$

for $\boldsymbol{A}_1$ and $\boldsymbol{P}_1$. It turns out that the above choice for $\boldsymbol{Q}_1$ decouples the computation of $\boldsymbol{A}_1$ and $\boldsymbol{P}_1$, yielding a $\boldsymbol{P}_1$ that can be written as $\boldsymbol{P}_1 = \boldsymbol{V}_1^\mathrm{H}\boldsymbol{R}_1^{-1}$, with $\boldsymbol{V}_1 = \boldsymbol{B}_1\boldsymbol{R}_0\boldsymbol{Q}_1^\mathrm{H}$ and $\boldsymbol{R}_1 = \boldsymbol{B}_1\boldsymbol{R}_0\boldsymbol{B}_1^\mathrm{H}$. Based on this observation, in the following stages $\boldsymbol{Q}_{i+1}$ is chosen such that

$$\mathrm{span}(\boldsymbol{Q}_{i+1}^\mathrm{T}) = \mathrm{span}(\boldsymbol{V}_i^*), \tag{9}$$

yielding $\boldsymbol{P}_{i+1} = \boldsymbol{V}_{i+1}^\mathrm{H}\boldsymbol{R}_{i+1}^{-1}$, where

$$\boldsymbol{V}_{i+1} = \boldsymbol{B}_{i+1}\boldsymbol{R}_i\boldsymbol{Q}_{i+1}^\mathrm{H}, \quad \text{and} \tag{10}$$

$$\boldsymbol{R}_{i+1} = \boldsymbol{B}_{i+1}\boldsymbol{R}_i\boldsymbol{B}_{i+1}^\mathrm{H}. \tag{11}$$

From Eq. (11) follows immediately

$$\boldsymbol{R}_i = \left(\prod_{k=i}^{1}\boldsymbol{B}_k\right)\boldsymbol{R}_0\left(\prod_{k=1}^{i}\boldsymbol{B}_k^\mathrm{H}\right). \tag{12}$$

The resulting $\boldsymbol{T}_i$ have the important property

$$\boldsymbol{T}_j\boldsymbol{R}_0\boldsymbol{T}_i^\mathrm{H} = \mathbf{0}_{K\times K}, \quad |j-i| > 1. \tag{13}$$

Eq. (13) can be proved by first plugging Eq. (8) into Eq. (13) and then using Eqs. (12), (10), (9) and (7).

Moreover, plugging Eq. (12) into Eq. (10) yields

$$\boldsymbol{V}_i = \left(\prod_{k=i}^{1}\boldsymbol{B}_k\right)\boldsymbol{R}_0\boldsymbol{T}_i^\mathrm{H}. \tag{14}$$

According to Eq. (9), $\boldsymbol{Q}_i = \boldsymbol{L}_i^{-1}\boldsymbol{V}_{i-1}^\mathrm{H}$ with a non-singular matrix $\boldsymbol{L}_i$. Thus, by combining Eqs. (8) and (14),

$$\boldsymbol{T}_i = \boldsymbol{L}_i^{-1}\boldsymbol{T}_{i-1}\boldsymbol{R}_0\left(\prod_{k=1}^{i-1}\boldsymbol{B}_k^\mathrm{H}\right)\left(\prod_{k=i-1}^{1}\boldsymbol{B}_k\right). \tag{15}$$

The rows of $\boldsymbol{T}^{(D)}$ are in general not orthogonal. For orthonormality,

$$\boldsymbol{T}_i\boldsymbol{T}_j^\mathrm{H} = \mathbf{1}_K\delta[i-j] \tag{16}$$

has to hold. Plugging Eq. (8) into Eq. (16), we find that for orthonormality, $\boldsymbol{Q}_i\boldsymbol{Q}_i^\mathrm{H} = \mathbf{1}_K$ and $\boldsymbol{B}_i\boldsymbol{B}_i^\mathrm{H} = \mathbf{1}_{N-iK}$. Now, both $\mathbf{1} - \boldsymbol{Q}_i^\mathrm{H}\boldsymbol{Q}_i$ and $\boldsymbol{B}_i^\mathrm{H}\boldsymbol{B}_i$ define an orthogonal projector onto $\mathrm{null}(\boldsymbol{Q}_i)$. From the uniqueness of projectors it can be concluded that $\boldsymbol{B}_i^\mathrm{H}\boldsymbol{B}_i = \mathbf{1} - \boldsymbol{Q}_i^\mathrm{H}\boldsymbol{Q}_i$. Plugging this result into Eq. (15), using Eq. (8) and finally Eq. (13), we find that Eq. (15) turns into the block-Lanczos algorithm [12], [13]:

$$\boldsymbol{T}_i = \boldsymbol{L}_i^{-1}\bar{\boldsymbol{T}}_i = \boldsymbol{L}_i^{-1}\Big(\boldsymbol{T}_{i-1}\boldsymbol{R}_0 - \sum_{k=i-2}^{i-1}\boldsymbol{T}_{i-1}\boldsymbol{R}_0\boldsymbol{T}_k^\mathrm{H}\boldsymbol{T}_k\Big),$$

with $\bar{\boldsymbol{T}}_1 = \boldsymbol{V}_0^\mathrm{H}$ and a matrix $\boldsymbol{L}_i^{-1}$ that orthonormalizes $\bar{\boldsymbol{T}}_i$. An obvious way to obtain $\boldsymbol{T}_i$ and $\boldsymbol{L}_i$ from $\bar{\boldsymbol{T}}_i$ is to compute the reduced LQ-factorization of $\bar{\boldsymbol{T}}_i$. With the block-Lanczos algorithm, we have found a very efficient way for computing a subspace basis. Moreover, from the fact that the Lanczos algorithm is a method for computing a Krylov-subspace basis of a Hermitian matrix, it follows that if the filters $\boldsymbol{T}_i$ are orthonormal, the reduced rank D-stage MSTxWF is the approximation of the TxWF in the Krylov subspace

$$\mathrm{span}\left(\left[\boldsymbol{V}_0^*, \boldsymbol{R}_0^*\boldsymbol{V}_0^*, \ldots, \boldsymbol{R}_0^{*,D-1}\boldsymbol{V}_0^*\right]\right).$$

Choose maximum dimension D
$\boldsymbol{T}_0 = \boldsymbol{0}_{K\times N}, \quad \boldsymbol{T}_1 = \mathrm{orth}\left(\boldsymbol{V}_0^{\mathrm{H}}\right)$
$\boldsymbol{U} = \boldsymbol{T}_1\boldsymbol{R}_0$
$\boldsymbol{R}_{1,1} = \boldsymbol{U}\boldsymbol{T}_1^{\mathrm{H}}, \quad \boldsymbol{R}_{1,0} = \boldsymbol{0}_{K\times K}$
$\boldsymbol{C}_{\mathrm{L}}^{(1)} = \boldsymbol{R}_{1,1}^{-1}, \quad \boldsymbol{C}_{\mathrm{F}}^{(1)} = \boldsymbol{C}_{\mathrm{L}}^{(1)}$
$\Delta = 1$
for $i = 2, \ldots, D$:
  $\bar{\boldsymbol{T}}_i = \boldsymbol{U} - \boldsymbol{R}_{i-1,i-1}\boldsymbol{T}_{i-1} - \boldsymbol{R}_{i-1,i-2}\boldsymbol{T}_{i-2}$
  if $\|\bar{\boldsymbol{T}}_i\|_F = 0$ then break
  $\boldsymbol{T}_i = \mathrm{orth}\left(\bar{\boldsymbol{T}}_i\right)$
  $\boldsymbol{U} = \boldsymbol{T}_i\boldsymbol{R}_0$
  $\boldsymbol{R}_{i,i} = \boldsymbol{U}\boldsymbol{T}_i^{\mathrm{H}}, \quad \boldsymbol{R}_{i,i-1} = \boldsymbol{U}\boldsymbol{T}_{i-1}^{\mathrm{H}}$
  $\boldsymbol{\Theta}_i = \boldsymbol{R}_{i,i} - \boldsymbol{R}_{i,i-1}\boldsymbol{\Theta}_{i-1}\boldsymbol{R}_{i,i-1}^{\mathrm{H}}$
  $\boldsymbol{C}_{\mathrm{L}}^{(i)} = \boldsymbol{\Theta}_i^{-1}\left[-\boldsymbol{R}_{i,i-1}\boldsymbol{C}_{\mathrm{L}}^{(i-1)} \quad \boldsymbol{1}_K\right]$
  $\boldsymbol{C}_{\mathrm{F}}^{(i)} = \left[\boldsymbol{C}_{\mathrm{F}}^{(i-1)} \quad \boldsymbol{0}_{K\times K}\right] - \boldsymbol{S}_i\boldsymbol{C}_{\mathrm{L}}^{(i-1),\mathrm{H}}\boldsymbol{R}_{i,i-1}^{\mathrm{H}}\boldsymbol{C}_{\mathrm{L}}^{(i)}$
  $\Delta = i$
$\boldsymbol{T}^{(D)} = \left[\boldsymbol{T}_1^{\mathrm{T}}, \ldots, \boldsymbol{T}_\Delta^{\mathrm{T}}\right]^{\mathrm{T}}$
$\tilde{\boldsymbol{P}}_{\mathrm{WF}} = \boldsymbol{V}_0^{\mathrm{H}}\boldsymbol{T}_1^{\mathrm{H}}\boldsymbol{C}_{\mathrm{F}}^{(\Delta)}$
$\tilde{\beta}_{\mathrm{WF}} = \sqrt{E_{\mathrm{tr}}/\sigma_s^2}\,\|\tilde{\boldsymbol{P}}_{\mathrm{WF}}\|_{\mathrm{F}}^{-1}$
$\tilde{\boldsymbol{P}}_{\mathrm{WF}} = \tilde{\beta}_{\mathrm{WF}}\tilde{\boldsymbol{P}}_{\mathrm{WF}}$
$\boldsymbol{P}_{\mathrm{WF}}^{(D)} = \tilde{\boldsymbol{P}}_{\mathrm{WF}}\boldsymbol{T}^{(D)}$

TABLE I
BLOCK-LANCZOS MSTXWF

B. Reduced Rank Solution

The D-stage MSTxWF of rank DK is found by plugging $\boldsymbol{P} = \tilde{\boldsymbol{P}}\boldsymbol{T}^{(D)}$ into (1) and minimizing over $\tilde{\boldsymbol{P}} \in \mathbb{C}^{K\times DK}$:

$$\left\{\tilde{\boldsymbol{P}}_{\mathrm{WF}}, \tilde{\beta}_{\mathrm{WF}}\right\} = \underset{\{\tilde{\boldsymbol{P}},\tilde{\beta}\}}{\mathrm{argmin}}\, \sigma_{\boldsymbol{\varepsilon}}^2\Big|_{\boldsymbol{P}=\tilde{\boldsymbol{P}}\boldsymbol{T}^{(D)}}$$
$$\text{s.t.} \quad \sigma_s^2\left\|\tilde{\boldsymbol{P}}\boldsymbol{T}^{(D)}\right\|_{\mathrm{F}}^2 \leq E_{\mathrm{tr}}.$$

With $\boldsymbol{R}_0^{(D)} = \boldsymbol{T}^{(D)}\boldsymbol{R}_0\boldsymbol{T}^{(D),\mathrm{H}}$ and $\boldsymbol{V}_0^{(D)} = \boldsymbol{T}^{(D)}\boldsymbol{V}_0$, the solution is given by

$$\tilde{\boldsymbol{P}}_{\mathrm{WF}} = \tilde{\beta}_{\mathrm{WF}}\boldsymbol{V}_0^{(D),\mathrm{H}}\boldsymbol{R}_0^{(D),-1},$$
$$\tilde{\beta}_{\mathrm{WF}} = \sqrt{E_{\mathrm{tr}}/\sigma_s^2}\left\|\boldsymbol{V}_0^{(D),\mathrm{H}}\boldsymbol{R}_0^{(D),-1}\right\|_{\mathrm{F}}^{-1}.$$

Note that only the first K rows of $\boldsymbol{V}_0^{(D)}$ are non-zero. Thus, only the first K rows of the inverse of $\boldsymbol{R}_0^{(D)}$ are actually needed for the computation of the solution. Due to Eq. (13), the matrix $\boldsymbol{R}_0^{(D)}$ is tri-diagonal and can be constructed from $\boldsymbol{R}_0^{(D-1)}$ in the following way:

$$\boldsymbol{R}_0^{(D)} = \left[\begin{array}{c|c} \boldsymbol{R}_0^{(D-1)} & \begin{matrix}\boldsymbol{0}_{(D-2)K\times K} \\ \boldsymbol{R}_{D-1,D}\end{matrix} \\ \hline \boldsymbol{0}_{K\times(D-2)K} \quad \boldsymbol{R}_{D-1,D}^{\mathrm{H}} & \boldsymbol{R}_{D,D} \end{array}\right],$$

with $\boldsymbol{R}_{i,j} = \boldsymbol{T}_i\boldsymbol{R}_0\boldsymbol{T}_j^{\mathrm{H}}$. Now define the matrices $\boldsymbol{C}_{\mathrm{F}}^{(D)}$ and $\boldsymbol{C}_{\mathrm{L}}^{(D)}$, where $\boldsymbol{C}_{\mathrm{F}}^{(D)}$ and $\boldsymbol{C}_{\mathrm{L}}^{(D)}$ contain the first K and the last K rows of $\boldsymbol{R}_0^{(D),-1}$, respectively. Using the inversion lemma for partitioned matrices [11], we find that $\boldsymbol{C}_{\mathrm{F}}^{(D)}$ and $\boldsymbol{C}_{\mathrm{L}}^{(D)}$ can be computed iteratively:

$$\boldsymbol{C}_{\mathrm{L}}^{(D)} = \boldsymbol{\Theta}_D^{-1}\left[-\boldsymbol{R}_{D-1,D}^{\mathrm{H}}\boldsymbol{C}_{\mathrm{L}}^{(D-1)} \quad \boldsymbol{1}_K\right],$$
$$\boldsymbol{C}_{\mathrm{F}}^{(D)} = \left[\boldsymbol{C}_{\mathrm{F}}^{(D-1)} \quad \boldsymbol{0}_{K\times K}\right] - \boldsymbol{S}_D\boldsymbol{C}_{\mathrm{L}}^{(D-1),\mathrm{H}}\boldsymbol{R}_{D-1,D}\boldsymbol{C}_{\mathrm{L}}^{(D)},$$

with $\boldsymbol{\Theta}_D = \boldsymbol{R}_{D,D} - \boldsymbol{R}_{D-1,D}^{\mathrm{H}}\boldsymbol{\Theta}_{D-1}\boldsymbol{R}_{D-1,D}$ and a selection matrix $\boldsymbol{S}_D = \left[\boldsymbol{1}_K \quad \boldsymbol{0}_{K\times K(D-2)}\right]$. With this iterative algorithm we can efficiently compute the relevant first K rows of $\boldsymbol{R}_0^{(D),-1}$. Under the assumption that $N \gg K$, the complexity of each iteration is governed by the matrix-matrix multiplication $\boldsymbol{T}_i\boldsymbol{R}_0$, which is of complexity $\mathrm{O}\left(KN^2\right)$. As a result, the D-stage solution results in a complexity of $\mathrm{O}\left(DKN^2\right)$. Recall that the terms $\boldsymbol{T}_i\boldsymbol{R}_0$ are also needed in the Lanczos algorithm for iteratively computing the basis $\boldsymbol{T}^{(D)}$. As a result, the computation of $\boldsymbol{T}^{(D)}$ and the computation of $\tilde{\boldsymbol{P}}_{\mathrm{WF}}$ can be efficiently combined in a single algorithm of overall complexity $\mathrm{O}\left(DKN^2\right)$, see Table I. Compared to the full rank solution with $\mathrm{O}\left(N^3\right)$, complexity is reduced if a close-enough approximation is achieved for $D < N/K$.

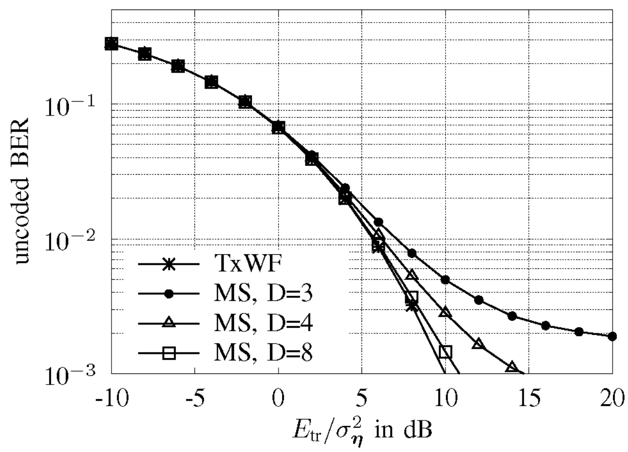

Fig. 2. MSTxWF, $K = 2$, $N_{\mathrm{a}} = 4$, $L = 14$

V. REDUCED RANK WIENER T-THP

In contrast to spatial THP and spatio-temporal THP, in Wiener T-THP the matrix $\check{\boldsymbol{R}}_0$, which is needed in the computation of the feedforward filter, is user-independent [10]. Based on this property, the results from Section IV are directly applicable to Wiener T-THP, i.e. the complexity of Wiener T-THP is reduced by computing a reduced rank approximation of the optimum feedforward filter $\boldsymbol{P}_{\mathrm{WTHP}}$. According to Eq. (3), it suffices to replace $\boldsymbol{R}_0$ by $\check{\boldsymbol{R}}_0$ in the derivation of the MSTxWF.

Eq. (5) defines a lower bound for the latency time ν. An upper bound follows from the fact that in order to collect the desired symbol's energy from all channel taps, it is necessary to choose $L \geq Q$ and $Q \leq \nu \leq L$. It follows that Eq. (5) restricts our considerations to systems with $N_{\mathrm{a}}/K \leq 1$, where equality can be achieved for $\nu = L$. Note that the case $N_{\mathrm{a}}/K = 1$ already corresponds to a highly loaded scenario.

VI. SIMULATION RESULTS

In TD-CDMA cellular systems, high-data rates require a small spreading factor (SF). Here, the special case of SF = 1 is considered. In this case, user separation has to be accomplished by space-time processing at the transmitter. We present *uncoded bit error rate* (uncoded BER) results for QPSK modulation and a system with $N_{\mathrm{a}} = 4$ antennas at the transmitter and a variable number of receivers. We assume a chip rate of $f_{\mathrm{c}} = 3.84$Mcps. The channel has an exponential

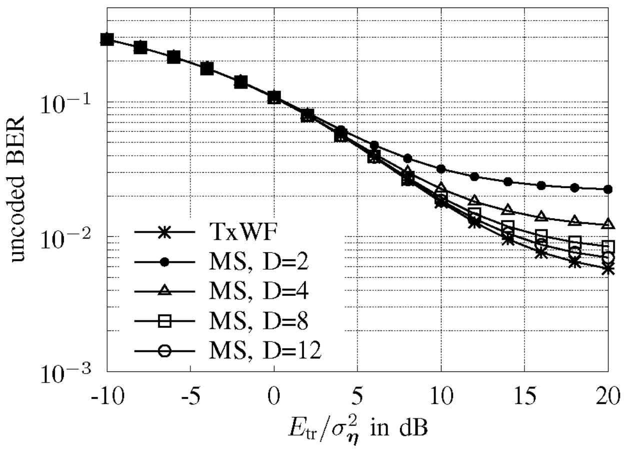

Fig. 3. MSTxWF, $K = 4$, $N_\text{a} = 4$, $L = 28$

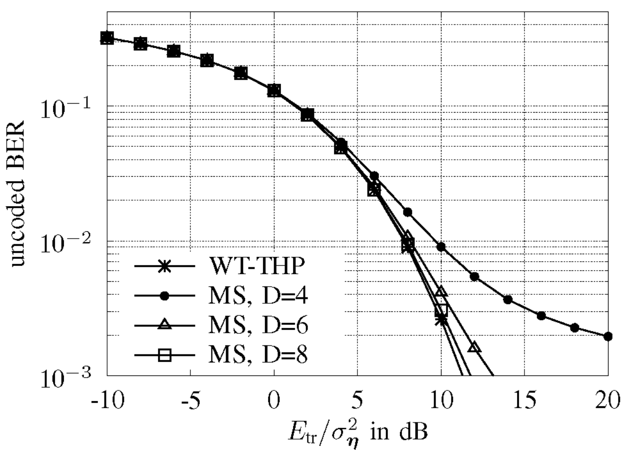

Fig. 4. MS WT-THP, $K = 4$, $N_\text{a} = 4$, $L = 14$

power delay profile with 6 paths and maximum delay spread $T_D = 3.9\mu$s. This is incorporated in our system model by setting $Q+1 = \lceil 3.84 * 3.9 \rceil = 15$. We assume temporally and spatially uncorrelated Rayleigh fading.

Fig. 2 shows the uncoded BER performance of the linear MSTxWF in a scenario with $K = 2$ receivers and a feedforward filter of order $L = 14$. The latency time is set to $\nu = 14$. Note that $N/K = 60/2$, i.e., the TxWF is equivalent to a MSTxWF with 30 stages. The MSTxWF can achieve a significant reduction in complexity: For a target BER of 10^{-2}, $D = 4$ stages provide optimum performance; $D = 6$ stages are needed at a target BER of 10^{-3}.

In the second scenario, the number of receivers is increased to $K = 4$. Note that this corresponds to a highly loaded scenario, as $K = N_\text{a}$. In order to provide the linear precoder with additional degrees of freedom, the filter order is set to $L = 28$. As can be observed from Fig. 2, even the BER curve of the optimum TxWF saturates. Due to $K = N_\text{a}$, the data streams cannot be completely separated by a linear precoder, resulting in an error floor. In this scenario, $N/K = 29$. Again, the MSTxWF can provide a significant reduction in complexity, depending on the target BER. However, it can also be noticed that the number of stages required to obtain a close to optimum solution increases with the system load.

Fig. 4 demonstrates the superiority of nonlinear *Wiener T-THP* (WT-THP) over the linear TxWF in highly loaded scenarios. The order of the feedforward filter is set to $L = 14$, i.e. the complexity required to compute the optimum feedforward filter solution is the same as in Fig. 2, although twice the number of users is served. The overall number of filter coefficients (including the feedback filter) is approximately equal to the linear TxWF with $L = 28$. The latency time is again chosen as $\nu = 14$. With the given parameters, $N/K = 15$ for the WT-THP system. Considering the performance of reduced rank multi-stage WT-THP, the results in Fig. 4 show that 6 to 8 stages are sufficient to get almost optimum performance in the BER range under consideration. Compared to WT-THP, again a considerable reduction in complexity.

VII. Conclusions

We derived the multi-user MSTxWF, showed its relationship to block-Krylov methods and provided an efficient algorithm for computing a low complexity reduced rank solution. Moreover, we extended our results to nonlinear precoding by developing reduced-complexity multi-stage Wiener T-THP. Simulation results demonstrate that the presented MSTxWF approach can achieve a substantial complexity reduction for both linear and nonlinear MSE-optimum precoding in frequency-selective multi-user scenarios.

References

[1] M. Joham, K. Kusume, M. H. Gzara, W. Utschick, and J. A. Nossek, "Transmit Wiener Filter for the Downlink of TDD DS-CDMA Systems," in *Proc. ISSSTA 2002*, September 2002, vol. 1, pp. 9–13.

[2] M. Schubert and H. Boche, "Solution of the Multiuser Downlink Beamforming Problem with Individual SINR constraints," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 1, pp. 18–28, January 2004.

[3] A. Wiesel, Y.C. Eldar, and S. Shamai, "Linear MIMO Precoders for Fixed Receivers," in *Proc. ICASSP 2004*, May 2004, vol. IV, pp. 481–484.

[4] 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, November 1998.

[5] J. Brehmer, M. Joham, G. Dietl, and W. Utschick, "Reduced Complexity Transmit Wiener Filter Based on a Krylov Subspace Multi-Stage Decomposition," in *Proc. ICASSP 2004*, May 2004, vol. IV, pp. 805–808.

[6] H. Harashima and H. Miyakawa, "Matched-Transmission Technique for Channels With Intersymbol Interference," *IEEE Transactions on Communications*, vol. 20, no. 4, pp. 774–780, August 1972.

[7] M. Tomlinson, "New automatic equaliser employing modulo arithmetic," *Electronics Letters*, vol. 7, no. 5/6, pp. 138–139, March 1971.

[8] G. Ginis and J. M. Cioffi, "A Multi-user Precoding Scheme achieving Crosstalk Cancellation with Application to DSL Systems," in *Proc. Asilomar Conference on Signals, Systems, and Computers*, October 2000, vol. 2, pp. 1627–1631.

[9] R. F. H. Fischer, C. Windpassinger, A. Lampe, and J. B. Huber, "Space-Time Transmission using Tomlinson-Harashima Precoding," in *Proc. 4th ITG Conference on Source and Channel Coding*, January 2002, pp. 139–147.

[10] M. Joham, J. Brehmer, A. Voulgarelis, and W. Utschick, "Multiuser Spatio-Temporal Tomlinson-Harashima Precoding for Frequency Selective Vector Channels," in *Proc. ITG Workshop on Smart Antennas*, March 2004.

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

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