ALTERNATING OPTIMIZATION FOR MMSE BROADCAST PRECODING

Raphael Hunger, Wolfgang Utschick, David A. Schmidt, and Michael Joham

Institute for Circuit Theory and Signal Processing
Munich University of Technology, 80290 Munich, Germany
e-mail: {hunger, utschick, schmidt, joham}@nws.ei.tum.de

ABSTRACT

We address the problem of jointly optimizing the precoder and the receivers in a multi-user broadcast system for a linear and nonlinear transmitter under sum mean square error (MSE) minimization. By means of alternating optimization (AO), we find an iterative algorithm which always converges to the global optimum. In addition to an elegant initialization of the receivers' weights, we come up with expressions for the speed of convergence. Our algorithm is applicable to both single-antenna and multi-antenna receivers and can be combined with nonlinear Tomlinson-Harashima Precoding (THP). For THP, the precoding order can easily be included in the AO. Moreover, we show that for sum MSE minimization the user-wise channels are not necessarily diagonalized and give some ideas how this fact can be exploited.

1. INTRODUCTION

In a broadcast scenario, the decentralized receivers cannot cooperate. Since the receivers have not enough degrees of freedom, precoding techniques have to be applied in such point-to-multi-point connections to combat the multi-user interference. The most popular approaches are *zero-forcing* (ZF) [1, 2, 3], and *minimum* MSE (MMSE) precoding [4, 5, 6, 7]. In the special case of single-antenna receivers, the only available degrees of freedom at the receivers are scalar weights. However, only a few of above contributions allow for different receiver weights. Especially when the users experience different path losses, large gains can be expected by admitting different scalars, as shown in [3] for the ZF variant. In [6] the sum MSE minimization is considered for linear precoding by means of uplink-downlink duality and employing interior point methods to solve the power allocation problem. Interior point methods are difficult to handle and have a high complexity. Moreover, the optimum choice of the precoding order still remains a major problem in general as well as a tedious exhaustive search is required to determine the optimum power loading, see [5].

Our contribution is an iterative algorithm that converges *q-linearly* [8, 9] to the optimum set of precoder and receivers' weights for both single- and multi-antenna receivers. This algorithm which alternates between transmitter and receiver updates while keeping the respective counterpart *constant* is based on AO, and is also applicable to the nonlinear THP. Here, the precoding order can easily be included and updated in each transmitter update stage. No exhaustive search is required for the computation of the optimum THP filters. For multi-antenna receivers, we highlight the interesting fact that even in the single-user case (general MIMO with cooperating antennas), the optimum MMSE transmitter and receiver pair does not necessarily diagonalizes the channel. This can be exploited for stream balancing.

2. SYSTEM MODEL AND NOTATION

Fig. 1 shows the downlink of the broadcast system with K non-cooperative users served by a single base station. The complex-valued data symbols s_k of the K users are stacked in the column vector $\boldsymbol{\mathsf{s}} = [\mathsf{s}_1, \ldots, \mathsf{s}_K]^{\mathrm{T}} \in \mathbb{A}^K$, whose correlation matrix reads as $\mathrm{E}[\boldsymbol{\mathsf{s}}\boldsymbol{\mathsf{s}}^{\mathrm{H}}] = \sigma_{\mathsf{s}}^2 \boldsymbol{I}_K$; $\mathbb{A}$ denotes the symbol alphabet. The transmit filter $\boldsymbol{P} = [\boldsymbol{p}_1, \ldots, \boldsymbol{p}_K] \in \mathbb{C}^{N_\mathrm{a} \times K}$ consists of K spatial filters $\boldsymbol{p}_k \in \mathbb{C}^{N_\mathrm{a}}$. Here, N_a denotes the number of antennas deployed at the transmitter. The transmission over the frequency flat channel is characterized by the channel matrix $\boldsymbol{H} = [\boldsymbol{h}_1, \ldots, \boldsymbol{h}_K]^{\mathrm{T}} \in \mathbb{C}^{K \times N_\mathrm{a}}$, whose k-th row $\boldsymbol{h}_k^{\mathrm{T}}$ represents the channel coefficients from all antennas to user k. Throughout this paper, we assume perfect *channel state information* (CSI). The non-negative real-valued user weights b_k are stored in the diagonal matrix $\boldsymbol{B} = \mathbf{diag}\{b_k\}_{k=1}^K \in \mathbb{R}_{+,0}^{K \times K}$. With above definitions, the estimate $\hat{\boldsymbol{\mathsf{s}}}$ for the true symbol vector $\boldsymbol{\mathsf{s}}$ can then be expressed as $\hat{\boldsymbol{\mathsf{s}}} = \boldsymbol{BHP}\boldsymbol{\mathsf{s}} + \boldsymbol{B}\boldsymbol{\eta} \in \mathbb{C}^K$. The noise vector $\boldsymbol{\eta} \in \mathbb{C}^K$ represents the zero-mean additive noise of all users with covariance matrix $\mathrm{E}[\boldsymbol{\eta}\boldsymbol{\eta}^{\mathrm{H}}] = \boldsymbol{R}_{\boldsymbol{\eta}} \in \mathbb{C}^{K \times K}$. The k-th main diagonal entry of $\boldsymbol{R}_{\boldsymbol{\eta}}$ is $\sigma_{\eta_k}^2$. We compute the *sum mean square error* (MSE) $\varepsilon(\boldsymbol{P}, \boldsymbol{B}) := \mathrm{E}[\|\hat{\boldsymbol{\mathsf{s}}} - \boldsymbol{\mathsf{s}}\|_2^2]$ to

$$\begin{aligned}\varepsilon(\boldsymbol{P}, \boldsymbol{B}) = \sigma_{\mathsf{s}}^2 K - \sigma_{\mathsf{s}}^2 \operatorname{tr}(\boldsymbol{BHP}) - \sigma_{\mathsf{s}}^2 \operatorname{tr}(\boldsymbol{P}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{B}) \\ + \sigma_{\mathsf{s}}^2 \operatorname{tr}(\boldsymbol{BHPP}^{\mathrm{H}}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{B}) + \operatorname{tr}(\boldsymbol{B}\boldsymbol{R}_{\boldsymbol{\eta}}\boldsymbol{B}).\end{aligned} \quad (1)$$

Notation: Deterministic vectors and matrices are denoted by lower and upper case bold letters. The respective random variables are written in sans-serif font. The operators $\mathrm{E}[\cdot]$, $(\cdot)^{\mathrm{T}}$, $(\cdot)^{\mathrm{H}}$, $(\cdot)^*$, and $\operatorname{tr}(\cdot)$ stand for expectation with respect to symbols and noise, transposition, Hermitian transposition, complex conjugate, and trace of a matrix, respectively. $\boldsymbol{I}_K$ is the $K \times K$ identity matrix, and $\|\cdot\|_2$ is the Euclidean norm.

3. JOINT OPTIMIZATION OF THE LINEAR MMSE PRECODER AND THE RECEIVERS

The joint optimization of the MSE $\varepsilon(\boldsymbol{P}, \boldsymbol{B})$ with respect to *both* the precoder $\boldsymbol{P}$ *and* the receivers' weights in $\boldsymbol{B}$ reads as

$$\left\{\boldsymbol{P}_{\mathrm{DWF}}, b_{\mathrm{DWF}}, \boldsymbol{B}'_{\mathrm{DWF}}\right\} = \underset{\{\boldsymbol{P}, b, \boldsymbol{B}'\}}{\operatorname{argmin}}\, \varepsilon(\boldsymbol{P}, b\boldsymbol{B}') \quad \text{subject to: } \mathrm{E}\left[\|\boldsymbol{P}\boldsymbol{s}\|_2^2\right] \leq E_{\mathrm{tr}} \text{ and } \boldsymbol{B}' = \mathbf{diag}\{b'_k\}_{k=1}^K, \tag{2}$$

where $\boldsymbol{B} = b\boldsymbol{B}'$ has been split into $\boldsymbol{B}'$ and a common scaling factor b in order to easily fulfill the power constraint.

4. ITERATIVE SOLUTION OF THE LINEAR MMSE OPTIMIZATION

Since there is no closed form solution for the joint optimization w.r.t. the set $\{\boldsymbol{P}, b, \boldsymbol{B}'\}$ in (2), we optimize the precoder and the scaling factors separately in an *alternating fashion*. This procedure is termed *alternating optimization*, *block-nonlinear Gauss-Seidel iteration*, *grouped coordinate descent*, or *block-relaxation* [8]. This attractive iterative scheme splits up the joint optimization into two partial sub-optimizations, namely the optimum transmitter $\boldsymbol{P}$ and the optimum common scalar weight b for fixed receivers $\boldsymbol{B}'$,[1] and the optimum receiver weights $\boldsymbol{B}'$ given a fixed transmitter $\boldsymbol{P}$ and a fixed common scalar weight b. The benefit results from the fact that the sub-optimizations have closed form solutions [9]. Hence, we do not need to make use of the inconvenient interior point method in order to find a solution of (2) as proposed in [5, 6].

4.1. Optimization of the Precoder for Fixed Receivers

Assume we are given *fixed* unscaled receiver weights $b_k'^{(n)}$, with the superscript $(\cdot)^{(n)}$ denoting the iteration number and $n=0$ representing the initialization step. *Grouped coordinate descent* means that we minimize the MSE w.r.t. the variables $\boldsymbol{P}$ and b of the first partitioning $\{\boldsymbol{P}, b\}$, keeping all variables from the second partitioning $\{\boldsymbol{B}'\}$ constant. *Given the fixed weights* $\boldsymbol{B}'^{(n)}$, the optimum precoding matrix $\boldsymbol{P}^{(n+1)}$ and the optimum common weight $b^{(n+1)}$ minimizing the sum MSE follow from

$$\{\boldsymbol{P}^{(n+1)}, b^{(n+1)}\} = \underset{\{\boldsymbol{P}, b\}}{\operatorname{argmin}}\, \varepsilon(\boldsymbol{P}, b\boldsymbol{B}'^{(n)}) \quad \text{subject to: } \mathrm{E}\left[\|\boldsymbol{P}\boldsymbol{s}\|_2^2\right] \leq E_{\mathrm{tr}}, \tag{3}$$

and $\boldsymbol{P}^{(n+1)}$ reads as [3]

$$\boldsymbol{P}^{(n+1)} = \frac{1}{b^{(n+1)}}\left(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{B}'^{(n)2}\boldsymbol{H} + \frac{\mathrm{tr}(\boldsymbol{B}'^{(n)2}\boldsymbol{R}_{\boldsymbol{\eta}})}{E_{\mathrm{tr}}}\boldsymbol{I}\right)^{-1}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{B}'^{(n)}. \tag{4}$$

[1] The common weight b is optimized during the transmitter computation since otherwise a closed form solution for the transmitter is not possible.

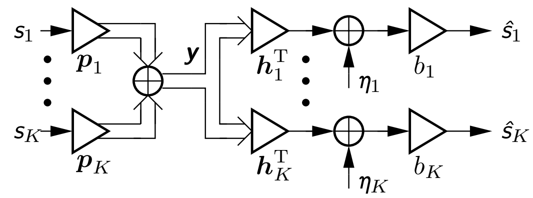

Figure 1. Downlink of the linear *multi-user multiple-input single-output* (MU-MISO) system.

The weight $b^{(n+1)}$ is computed to fulfill the power constraint in (2). Note that we do not optimize (3) with respect to $\boldsymbol{B}'$, and that (4) is invariant to a scaling of $\boldsymbol{B}'$ by α, i.e., $\boldsymbol{P}^{(n+1)}\left(\boldsymbol{B}'\right) = \boldsymbol{P}^{(n+1)}\left(\alpha\boldsymbol{B}'\right)$, because $b^{(n+1)}$ is proportional to α^{-1}. The mapping in (4) that generates the set $\{\boldsymbol{P}^{(n+1)}, b^{(n+1)}\}$ from the set $\{\boldsymbol{B}'^{(n)}\}$ is called a *minimization mapping* [10] since the updated set $\{\boldsymbol{P}^{(n+1)}, b^{(n+1)}\}$ is the unique optimum given fixed $\{\boldsymbol{B}'^{(n)}\}$. This property is favorable as no (possibly suboptimum) line search techniques with gradient related search directions have to be applied in order to find the updated set $\{\boldsymbol{P}^{(n+1)}, b^{(n+1)}\}$ for step $n+1$ [10]. Because $\{\boldsymbol{P}^{(n+1)}, b^{(n+1)}\}$ are optimum for fixed unscaled receiver weights $\boldsymbol{B}'^{(n)}$, we find for any $\boldsymbol{P}$ satisfying the power constraint in (2) and for any b, that

$$\varepsilon\Big(\boldsymbol{P}^{(n+1)}, b^{(n+1)}\boldsymbol{B}'^{(n)}\Big) \leq \varepsilon\Big(\boldsymbol{P}, b\boldsymbol{B}'^{(n)}\Big). \tag{5}$$

4.2. Optimization of the Receivers for a Fixed Precoder

Having computed the updated set $\{\boldsymbol{P}^{(n+1)}, b^{(n+1)}\}$, we now optimize the MSE w.r.t. the second partitioning $\{\boldsymbol{B}'\}$, in order to compute the new set $\{\boldsymbol{B}'^{(n+1)}\}$. We can distinguish two types of connections [10]: *Parallel connection* implies that $\boldsymbol{B}'^{(n+1)}$ is updated on the basis of $\{\boldsymbol{P}^{(n)}, b^{(n)}\}$, i.e., the results of the previous transmitter update are not incorporated. However, as the set $\{\boldsymbol{B}'^{(n)}\}$ of the previous iteration step has been computed by a *minimization mapping*, $\{\boldsymbol{B}'\}$ would be updated only every second step. *Serial connection* stands for incorporating the updates of all previous variable sets for the computation of the next set. In our case, this implies that $\{\boldsymbol{B}'^{(n+1)}\}$ is determined on the basis of $\{\boldsymbol{P}^{(n+1)}, b^{(n+1)}\}$. Since we apply the *minimization mapping*, we clearly make use of the *serial connection*. The receiver weights $b^{(n+1)}\boldsymbol{B}'^{(n)}$ are in general not optimum for the precoder $\boldsymbol{P}^{(n+1)}$, except for the case that they have coincidentally been chosen to be $b^{(n+1)}\boldsymbol{B}'^{(n)} = \boldsymbol{B}_{\mathrm{DWF}}$. From the k-th receiver's point of view, Fig. 2 shows the received signal model for the second intermediate update stage in step $n+1$. If receiver k does not scale its received signal by $b^{(n+1)}b_k'^{(n)}$ but allows for an unscaled weight b'_k different from $b_k'^{(n)}$ instead, its resulting MSE $\varepsilon_k(\boldsymbol{P}^{(n+1)}, b^{(n+1)}b'_k)$ reads as

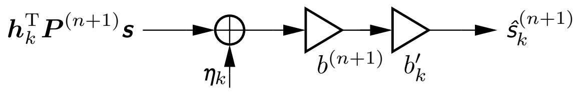

Figure 2. Received signal model of user k.

$$
\begin{aligned}
\varepsilon_k\left(\boldsymbol{P}^{(n+1)}, b^{(n+1)} b'_k\right) &= \mathrm{E}\left[\left|s_k - \hat{s}_k^{(n+1)}\right|^2\right] \\
&= \sigma_s^2 - b^{(n+1)} b'_k \boldsymbol{h}_k^{\mathrm{T}} \boldsymbol{p}_k^{(n+1)} \sigma_s^2 - b^{(n+1)} b'_k \boldsymbol{p}_k^{(n+1)\,\mathrm{H}} \boldsymbol{h}_k^* \sigma_s^2 \\
&\quad + b^{(n+1)2} b_k'^2 \boldsymbol{h}_k^{\mathrm{T}} \boldsymbol{P}^{(n+1)} \boldsymbol{P}^{(n+1)\,\mathrm{H}} \boldsymbol{h}_k^* \sigma_s^2 + b^{(n+1)2} b_k'^2 \sigma_{\eta_k}^2
\end{aligned} \tag{6}
$$

with $\boldsymbol{p}_k^{(n+1)} = \boldsymbol{P}^{(n+1)} \boldsymbol{e}_k$ representing the k-th column of $\boldsymbol{P}^{(n+1)}$, and the estimate $\hat{s}_k^{(n+1)}$ being defined as

$$
\hat{s}_k^{(n+1)} = b^{(n+1)} b'_k \left(\boldsymbol{h}_k^{\mathrm{T}} \boldsymbol{P}^{(n+1)} \boldsymbol{s} + \eta_k\right). \tag{7}
$$

Now, receiver k can apply a scalar MMSE filter

$$
b_k'^{(n+1)} = \underset{b'_k}{\operatorname{argmin}}\, \varepsilon_k\left(\boldsymbol{P}^{(n+1)}, b^{(n+1)} b'_k\right), \quad \forall k \tag{8}
$$

minimizing (6). This leads to the unscaled receiver weights

$$
b_k'^{(n+1)} = \frac{\sigma_s^2 \boldsymbol{h}_k^{\mathrm{T}} \boldsymbol{p}_k^{(n+1)}}{b^{(n+1)} (\sigma_s^2 \boldsymbol{h}_k^{\mathrm{T}} \boldsymbol{P}^{(n+1)} \boldsymbol{P}^{(n+1)\,\mathrm{H}} \boldsymbol{h}_k^* + \sigma_{\eta_k}^2)}. \tag{9}
$$

Note that the optimizations in (8) are decoupled. It is obvious that the MSE decreases or stays constant compared to (5), i.e.,

$$
\varepsilon\left(\boldsymbol{P}^{(n+1)}, b^{(n+1)} \boldsymbol{B}'^{(n+1)}\right) \leq \varepsilon\left(\boldsymbol{P}^{(n+1)}, b^{(n+1)} \boldsymbol{B}'^{(n)}\right). \tag{10}
$$

4.3. Alternating Optimization

The transmitter iterates through several intermediate steps of both the precoder and the receivers' weights updates, see Algorithm 1. In Line 1, the maximum number of iterations is chosen, Line 2 allows to set a relative MSE descent threshold c_{th} in order to terminate the algorithm before n–1 has reached $n_{\max}$ if the relative MSE descent is below c_{th}. Starting with some initialization weights $\boldsymbol{B}'^{(0)}$, Line 6 computes the precoder $\boldsymbol{P}^{(n+1)}$ and the common receiver weight $b^{(n+1)}$, Line 8 the new receiver weights $\boldsymbol{B}'^{(n+1)}$. If the relative MSE descent is below c_{th}, Line 11 terminates the program. In case of a fixed number of iterations, Lines 7 and 9–12 can be skipped.

4.4. Choice of Initialization Weights

The choice of the initialization weights $\boldsymbol{B}'^{(0)}$ in Line 3 drastically influences the speed of convergence. It is essential to choose all initial weights $b_k'^{(0)} \neq 0\, \forall k$, otherwise, the weights would stay zero during all iterations, see the update rules (4) and (9). Nonetheless, the case might occur, where some users may be switched off in order to minimize the *sum* MSE, see [3]. Our algorithm then asymptotically switches off these streams. In [11], Serbetli et al. propose the use of random weights in their first step. We instead propose to initialize the AO with the weights of the diagonal ZF filter [3], which leads to dramatically increased convergence especially in the medium to high SNR-regime.

4.5. Convergence Analysis

Combining (5) and (10), we find that the MSE reduces monotonically. In combination with the fact that the MSE is lower bounded, convergence follows. We are able to prove that Algorithm 1 *always* converges to the global optimum by showing that only one stationary point is a minimum and trajectories to saddle-points or maxima are impossible. The latter is proven by relating the behavior of Alg. 1 in the vicinity of a saddle-point to the steepest descent variant, which never converges to such a saddle-point. The MSE series $\{\varepsilon^{(n)}\} := \{\varepsilon(\boldsymbol{P}^{(n)}, b^{(n)} \boldsymbol{B}'^{(n)})\}$ of the optimization is locally *q-linear* convergent if the joint optimization is split into the receiver weights update and into the precoding filter update, see [8, 9]. The *q-linear* convergence implies that for the MSE series $\{\varepsilon^{(n)}\}$ with limit $\varepsilon_{\mathrm{DWF}}$, there exists an integer m and a real valued scalar $\varrho \in [0, 1)$, such that $\|\varepsilon^{(n+1)} - \varepsilon_{\mathrm{DWF}}\| \leq \varrho \|\varepsilon^{(n)} - \varepsilon_{\mathrm{DWF}}\|$ is valid for $n \geq m$ [12]. The value ϱ governs the speed of convergence and depends on the spectral radius of a transformed Hessian in the global optimum [8]. A clear advantage of the alternating optimization compared to gradient methods is that there is no need to compute optimum step-sizes for line search algorithms and no gradient projections have to be evaluated.

5. EXTENSION TO THP AND MULTI-USER MIMO

The alternating optimization scheme can conveniently be applied to nonlinear THP. The solution of the conventional MMSE THP precoder [7], which assumes identical user weights $\boldsymbol{B}' = \boldsymbol{I}_K$ can again be utilized, when we substitute $\boldsymbol{H} \leftarrow \boldsymbol{B}'\boldsymbol{H}$ and $\boldsymbol{R}_{\boldsymbol{\eta}} \leftarrow \boldsymbol{B}'\boldsymbol{R}_{\boldsymbol{\eta}}\boldsymbol{B}'$. The receiver update follows from $b_k'^{(n+1)} = \operatorname{argmin}_{b'_k} \mathrm{E}[|d_k - \hat{d}_k|^2]$, with

$$
\begin{aligned}
\hat{d}_k &= b^{(n+1)} b'_k \left(\boldsymbol{h}_k^{\mathrm{T}} \boldsymbol{P}^{(n+1)} \boldsymbol{v}^{(n+1)} + \eta_k\right), \\
\boldsymbol{v}^{(n+1)} &= \left(\boldsymbol{I}_K - \boldsymbol{F}^{(n+1)}\right)^{-1} \boldsymbol{\Pi}^{(n+1)} \boldsymbol{d}.
\end{aligned} \tag{11}
$$

See [7] for the definition of the respective variables. In contrast to [5], where sorting is impossible and an exhaustive search has to be accomplished to find the optimum receiver weights, we can easily incorporate the sorting order in the precoder update stage and find the weights by the alternating optimization.

For MU-MIMO Systems, the AO technique also leads to the jointly optimized precoder and receiver matrices. However, in contrast to the commonly held assumption that the channel has to be diagonalized, the resulting transmitter-channel-receiver chain is not diagonal in general. The sum MSE is invariant to any unitary matrix $\boldsymbol{Q}_k^{\mathrm{H}}$ and $\boldsymbol{Q}_k$ multiplied to RHS and LHS of the transmitter and receiver matrices of user k, respectively. However, the individual streams of user k have different MSEs then. Diagonalization is achieved when $\boldsymbol{Q}_k$ follows from an *eigenvalue decomposition*, stream-balanced MSEs and, according to the Schur-Horn theorem, any MSE

Algorithm 1 Iterative computation of filters.

choose $n_{\max}$ {maximum iterations}
choose c_{th} {threshold constant}
get $\boldsymbol{B}'^{(0)}$ {initialization weights}
$n = 0$ {current iteration counter}
while $n \leq n_{\max}$ **do**
  update $\boldsymbol{P}^{(n+1)}$ and $b^{(n+1)}$ via (4) {new precoder}
  $\varepsilon_{\boldsymbol{P}}^{(n+1)} := \varepsilon(\boldsymbol{P}^{(n+1)}, b^{(n+1)}\boldsymbol{B}'^{(n)})$ {MSE Tx-update}
  update $\boldsymbol{B}'^{(n+1)}$ via (9) {new weights}
  $\varepsilon_{\boldsymbol{B}'}^{(n+1)} := \varepsilon(\boldsymbol{P}^{(n+1)}, b^{(n+1)}\boldsymbol{B}'^{(n+1)})$ {MSE Rx-update}
  if $\frac{|\varepsilon_{\boldsymbol{P}}^{(n+1)} - \varepsilon_{\boldsymbol{B}'}^{(n+1)}|}{\varepsilon_{\boldsymbol{P}}^{(n+1)}} < c_{\text{th}}$ **then**
    break {stop if rel. descent below threshold}
  end if
  $n \leftarrow n + 1$ {increase iteration counter}
end while

set which majorizes the eigenvalues of the MSE matrix can be achieved by applying the generalized *Bendel-Mickey* algorithm from [13], which carries out plane Givens rotations.

6. SIMULATION RESULTS

We study the performance of the linear MMSE precoder in terms of *uncoded bit-error-ratios* (BERs). $K = 4$ users are served by a base-station with $N_{\text{a}} = 4$ antennas. Users 1 and 2 have the same *average* channel powers which are ten times smaller than the average channel powers of users 3 and 4 for example due to pathloss. The total average channel power is normalized such that for the Frobenius norm, $\mathrm{E}[\|\boldsymbol{H}\|_{\mathrm{F}}^2] = N_{\text{a}}K = 16$ holds. The MMSE precoders are shown in Fig. 3. Here, the conventional WF from [4] with identical user weights allocates different MSEs and different BERs to the users. The stronger users 3 and 4 (plus-marker) exhibit smaller BER values than the weaker users (star-marker). The user averaged BER has the dot marker. When we allow for different user weights, the DWF variant yields only slight BER improvements for the *weak* users (triangle down marker). Nonetheless, the stronger users (triangle-up) feature a BER reduction compared to the conventional counterpart. The user averaged BER curve (square marker) exhibits a 1 dB gain to the average BER curve of the MMSE filter with identical weights (dot marker) at a BER of 0.1. We assumed $c_{\text{th}} = 0.001$, see Algorithm 1. The THP versions have dashed lines, only the strong users are shown here, as for the weak users, there is almost no gain when we allow for different weights. Again, a gain of nearly 2 dB can be achieved.

7. REFERENCES

[1] R. Fischer, C. Windpassinger, A. Lampe, and J. Huber, "MIMO Precoding for Decentralized Receivers," in *ISIT '02*, 2002, p. 496.

[2] B. R. Vojčić and W. M. Jang, "Transmitter Precoding in Synchronous Multiuser Communications," *IEEE Trans. Commun.*, vol. 46, no. 10, pp. 1346–1355, October 1998.

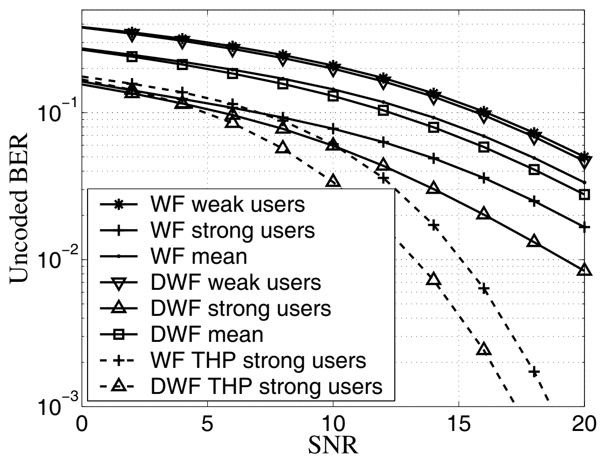

Figure 3. Uncoded BER vs. SNR for the conventional [4] and the diagonal linear Wiener filter (solid curves), and for the conventional [7] and diagonal THP (dashed curves).

[3] R. Hunger, M. Joham, and W. Utschick, "Extension of Linear and Nonlinear Transmit Filters for Decentralized Receivers," in *Proc. European Wireless*, vol. 1, April 2005, pp. 40–46.

[4] M. Joham, W. Utschick, and J. A. Nossek, "Linear Transmit Processing in MIMO Communications Systems," *IEEE Trans. Signal Processing*, vol. 53, no. 8, pp. 2700–2712, August 2005.

[5] M. Schubert and S. Shi, "MMSE Transmit Optimization with Interference Pre-Compensation," in *Proc. VTC 2005 Spring*, Stockholm, May 2005.

[6] S. Shi and M. Schubert, "MMSE Transmit Optimization for Multiuser Multiantenna Systems," in *ICASSP 05*, March 2005.

[7] M. Joham, J. Brehmer, and W. Utschick, "MMSE Approaches to Multiuser Spatio-Temporal Tomlinson-Harashima Precoding," in *Proc. ITG SCC'04*, January 2004, pp. 387–394.

[8] J. C. Bezdek and R. J. Hathaway, "Convergence of alternating optimization," *Neural, Parallel and Scientific Computation*, vol. 11, no. 4, pp. 351–368, 2003.

[9] J. C. Bezdek, R. J. Hathaway, R. Howard, C. Wilson, and M. P. Windham, "Local convergence analysis of a grouped variable version of coordinate descent," *Journal of Optimization Theory and Applications*, vol. 54, no. 3, pp. 471–477, 1987.

[10] L. Grippo and M. Sciandrone, "Globally convergent block-coordinate techniques for unconstrained optimization," *Optimization Methods and Software*, vol. 10, pp. 587–637, 1999.

[11] S. Serbetli and A. Yener, "Transceiver Optimization for Multiuser MIMO Systems," *IEEE Trans. Signal Processing*, vol. 52, no. 1, pp. 214–226, January 2004.

[12] J. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*. Academic Press, 1970.

[13] I. S. Dhillon, R. W. H. Jr., M. A. Sustik, and J. A. Tropp, "Generalized Finite Algorithms for Constructing Hermitian Matrices with Prescribed Diagonal and Spectrum," *SIAM Journal on Matrix Analysis and Applications*, vol. 27, no. 1, 2005.