Weighted Sum-Rate Maximization Using Weighted MMSE for MIMO-BC Beamforming Design

Christensen, Søren; De Carvalho, Elisabeth; Agarwal, Rajiv; Cioffi, John M.

Published in:
IEEE International Conference on Communications, 2009. ICC '09

DOI (link to publication from Publisher):
10.1109/ICC.2009.5199574

Publication date:
2009

Document Version
Publisher's PDF, also known as Version of record

Weighted Sum-Rate Maximization using Weighted MMSE for MIMO-BC Beamforming Design

Søren Skovgaard Christensen*, Rajiv Agarwal†, Elisabeth de Carvalho‡ and John M. Cioffi†

* Nokia Denmark
Modem Algorithm Design
Sydhavnsgade 17-19
DK-2450 Copenhagen, Denmark
skovgaard@ieee.org

†Stanford University
Packard Elect. Eng. Building
350 Serra Mall
Stanford, CA 94305, USA
{rajivag,cioffi}@stanford.edu

‡Aalborg University
Department of Electronic Systems
Fredrik Bajers Vej 7
DK-9220 Aalborg, Denmark
edc@es.aau.dk

***Abstract*—This paper studies linear transmit filter design for Weighted Sum-Rate (WSR) maximization in the Multiple Input Multiple Output Broadcast Channel (MIMO-BC). The problem of finding the optimal transmit filter is non-convex and intractable to solve using low complexity methods. Motivated by recent results highlighting the relationship between mutual information and Minimum Mean Square Error (MMSE), this paper establishes a relationship between weighted sum-rate and weighted MMSE in the MIMO-BC. The relationship is used to propose a low complexity algorithm for finding a local weighted sum-rate optimum based on alternating optimization. Numerical results studying sum-rate show that the proposed algorithm achieves high performance with few iterations.**

I. Introduction

Multiple Input Multiple Output (MIMO) systems have great potential to achieve high throughput in wireless systems [1]. In cellular systems, multiple antennas can easily be deployed at the base station to enhance the system capacity. When Channel State Information (CSI) is available at the transmitter, the base station can transmit to multiple users simultaneously to achieve a linear increase of system throughput in the number of transmit antennas. This can be done using linear or non-linear transmission techniques. For the Multiple Input Multiple Output Broadcast Channel (MIMO-BC), non-linear techniques have been shown to outperform linear techniques and achieve channel capacity. The capacity-achieving downlink strategy is non-linear and uses Dirty Paper Coding (DPC) [2]. However, practical techniques to implement DPC [3], [4], [5], are in preliminary states of development and adds implementation complexity due to non-linear operations at both transmitter and receivers. This makes linear downlink transmission techniques (also called beamforming) an attractive alternative because of their simplicity [6], [7], [8], [9]. Transmit beamforming design entails finding the linear transmit filter, through which the data intended for the different users is passed before transmission on the channel.

This paper focuses on transmit beamforming design to maximize Weighted Sum-Rate (WSR) subject to a transmit-power constraint, which is a non-convex and non-trivial problem. WSR is useful for prioritizing different users and thus finds different practical applications. For instance the weights can be chosen according to the state of the packet queues corresponding to a *max-stability* service [10] or, by using equal weights, to maximize sum-rate corresponding to a *best effort* service.

A recent paper [11] studies the same problem and proposes an iterative algorithm based on uplink-downlink Mean Square Error (MSE)-duality. From a given starting point, the algorithm converges to a local WSR-optimum. The principle in the algorithm is to iterate between the downlink system and a virtual uplink system in order to update filter structures, in addition to solving a Geometric Program (GP) for optimizing the transmit power distribution. In another recent paper [12], the authors attempt to solve the WSR problem using concepts from [8], however their algorithm is a 4-step iterative algorithm, two of which require solving a GP, which again is iterative and a Second-Order Cone Program (SOCP) respectively.

This paper takes a different approach to solving the WSR-problem which leads to an iterative algorithm that is guaranteed to converge to a local WSR-optimum. In the same line as recent results highlighting the relationship between information theoretic quantities (mutual information) and MMSE in single user MIMO channels [13], [14], we have established a relationship between WSR and Weighted sum-Minimum Mean Square Error (WMMSE) in the MIMO-BC. By comparing the gradients of resp. WSR and WMMSE cost functions we are able to show a simple relationship between the Karush-Kuhn-Tucker (KKT) conditions of the two problems. Essentially we show that the WSR-problem can be solved as a WMMSE-problem with optimized MSE-weights.

Using the derived correspondence we propose an iterative algorithm for WSR-optimization in the MIMO-BC. The algorithm iterates between WMMSE transmit filter design and MMSE receive filter computation using well-known closed-form expressions. In each iteration the MSE weights are updated using the derived correspondence. Each of the three steps is solved by evaluating closed-form expressions, and the proposed algorithm is less complex than state-of-the-art methods requiring multiple-level iterations. Numerical results comparing convergence rates and sum-rate performance to other recently proposed algorithms are presented.

Notation: m_{ij} denotes the (i, j)th entry of the matrix $\mathbf{M}$.

978-1-4244-3435-0/09/$25.00 ©2009 IEEE

$\mathbf{M}^{\mathrm{T}}/\mathbf{M}^{\mathrm{H}}/\mathrm{Tr}\,(\mathbf{M})$ denotes transpose/conjugate transpose/trace of a vector/matrix $\mathbf{M}$. The dimension of a matrix $\mathbf{M}$ is denoted by the subscript $\mathbf{M}_{[Q\times P]}$, where Q is the row dimension. $\mathbf{I}_K$ denotes an identity matrix of size $K \times K$. $||\mathbf{v}||$ denotes Euclidean-norm of a vector $\mathbf{v}$. $\mathbb{E}[\cdot]$ denotes statistical expectation.

II. System model and main objective

A general narrowband point-to-multipoint MIMO system is considered. There are P transmit antennas and K users, each with Q receive antennas. The system has in total QK receive antennas across all users[1]. The MIMO channel between the transmitter and user k is described by a matrix $\mathbf{H}_k \in \mathbb{C}^{[Q\times P]}$ containing complex-valued channel gains of the different antenna-pairs. The signal observed at user k at sample time n can be represented by the complex vector

$$\mathbf{y}_k(n) = \mathbf{H}_k\mathbf{x}(n) + \mathbf{v}_k(n), \tag{1}$$

where $\mathbf{x}(n) \in \mathbb{C}^{[P\times 1]}$ is the complex-valued transmitted vector, and $\mathbf{v}_k(n) \in \mathbb{C}^{[Q\times 1]}$ is a noise vector containing circularly symmetric white[2] Gaussian noise with covariance $\mathbf{R}_{v_k v_k} = \mathbb{E}\left[\mathbf{v}_k(n)\mathbf{v}_k(n)^{\mathrm{H}}\right] = \mathbf{I}_Q$. The transmit vector $\mathbf{x}(n)$ is a linearly filtered version of the input data vectors $\mathbf{d}_1(n), \cdots, \mathbf{d}_K(n) \in \mathbb{C}^{[Q\times 1]}$:

$$\mathbf{x}(n) = \sum_{k=1}^{K} \mathbf{B}_k\mathbf{d}_k(n). \tag{2}$$

The matrices $\mathbf{B}_1, \cdots, \mathbf{B}_K \in \mathbb{C}^{[P\times Q]}$ are the transmit filters (beamformers). It is assumed that each user has Q parallel data streams, although some of the streams can have a rate of zero. Additionally it is assumed that each user receives Q independent streams such that $\mathbb{E}\left[\mathbf{d}_k(n)\mathbf{d}_k^{\mathrm{H}}(n)\right] = \mathbf{I}_Q$. The transmit vectors respect a total block power constraint for a block consisting of N transmissions i.e.

$$\frac{1}{N}\sum_{n=1}^{N} \mathbf{x}^{\mathrm{H}}(n)\mathbf{x}(n) \le E_{tx}. \tag{3}$$

Throughout the analysis it is assumed that N is large such that $\frac{1}{N}\sum_{n=1}^{N} \mathbf{x}^{\mathrm{H}}(n)\mathbf{x}(n)$ can be replaced by $\mathbb{E}\left[\mathbf{x}^{\mathrm{H}}(n)\mathbf{x}(n)\right] = \sum_k \mathrm{Tr}\left(\mathbf{B}_k\mathbf{B}_k^{\mathrm{H}}\right)$. It is also assumed that the channel changes in a quasi-static manner, and hence the channel matrices $\mathbf{H}_1, \cdots, \mathbf{H}_K$ are constant for the duration of the block. Furthermore it is assumed that CSI, i.e. $\mathbf{H}_1, \cdots, \mathbf{H}_K$ is perfectly known at the transmitter.

A. Main objective

The main objective is to find the transmit filters $\mathbf{B}_1 \cdots \mathbf{B}_K$ which maximize the weighted sum-rate. This can be written as the minimization problem:

$$\begin{aligned}\left[\mathbf{B}_1^{\mathrm{WSR}}, \cdots, \mathbf{B}_K^{\mathrm{WSR}}\right] &= \operatorname*{arg\,min}_{\mathbf{B}_1,\cdots,\mathbf{B}_K} \sum_k -u_{R_k} R_k \\ &\quad \text{s.t.} \quad \sum_{k=1}^{K} \mathrm{Tr}\left(\mathbf{B}_k\mathbf{B}_k^{\mathrm{H}}\right) = E_{tx},\end{aligned} \tag{4}$$

where $u_{R_k} \ge 0$ and R_k defines respectively a weight and the rate for the k^{th} user. Without loss of generality we have used an equality power constraint rather than the often used $\sum_{k=1}^{K} \mathrm{Tr}\left(\mathbf{B}_k\mathbf{B}_k^{\mathrm{H}}\right) \le E_{tx}$ because the WSR optimum is reached at maximum transmit power. Assuming Gaussian signaling, the achievable rate for user k is given as

$$R_k = \log\det\left(\mathbf{I}_k + \mathbf{B}_k^{\mathrm{H}}\mathbf{H}_k^{\mathrm{H}}\mathbf{R}_{\tilde{v}_k\tilde{v}_k}^{-1}\mathbf{H}_k\mathbf{B}_k\right), \tag{5}$$

where $\mathbf{R}_{\tilde{v}_k\tilde{v}_k}$ denotes the effective noise covariance matrix at user k:

$$\mathbf{R}_{\tilde{v}_k\tilde{v}_k} = \mathbf{I}_k + \sum_{i=1, i\neq k}^{K} \mathbf{H}_k\mathbf{B}_i\mathbf{B}_i^{\mathrm{H}}\mathbf{H}_k^{\mathrm{H}}. \tag{6}$$

We notice that R_k can be expressed as a function of the error covariance matrix after MMSE receive filtering. The MMSE-receive filter at user k is given as:

$$\begin{aligned}\mathbf{A}_k^{\mathrm{MMSE}} &= \operatorname*{arg\,min}_{\mathbf{A}_k} \mathbb{E}\left[||\mathbf{A}_k\mathbf{y}_k - \mathbf{d}_k||^2\right] \\ &= \mathbf{B}_k^{\mathrm{H}}\mathbf{H}_k^{\mathrm{H}}\left(\mathbf{H}_k\mathbf{B}_k\mathbf{B}_k^{\mathrm{H}}\mathbf{H}_k^{\mathrm{H}} + \mathbf{R}_{\tilde{v}_k\tilde{v}_k}\right)^{-1}.\end{aligned} \tag{7}$$

The MSE-matrix for user k given that the MMSE-receive filter is applied can be written as [15]:

$$\begin{aligned}\mathbf{E}_k &= \mathbb{E}\left[\left(\mathbf{A}_k^{\mathrm{MMSE}}\mathbf{y}_k - \mathbf{d}_k\right)\left(\mathbf{A}_k^{\mathrm{MMSE}}\mathbf{y}_k - \mathbf{d}_k\right)^{\mathrm{H}}\right] \\ &= \left(\mathbf{I}_k + \mathbf{B}_k^{\mathrm{H}}\mathbf{H}_k^{\mathrm{H}}\mathbf{R}_{\tilde{v}_k\tilde{v}_k}^{-1}\mathbf{H}_k\mathbf{B}_k\right)^{-1}.\end{aligned} \tag{8}$$

We refer to $\mathbf{E}_k$ as the MMSE-matrix. Given (5) and (8) the rate for user k can be written as:

$$R_k = \log\det\left(\mathbf{E}_k^{-1}\right). \tag{9}$$

III. Gradient expressions and KKT conditions

This section first studies the gradient for the WSR maximization problem. Next, the gradient for a new optimization problem, WMMSE, is computed and we are able to show that there is a simple relationship between the KKT conditions of the two problems.

A. Gradient of Weighted Sum-Rate Maximization

To investigate stationary points of the problem (4) we formulate the Lagrangian:

$$f(\mathbf{B}_1, \cdots, \mathbf{B}_K) = \sum_k -u_{R_k}R_k + \lambda\left(\sum_k \mathrm{Tr}\left(\mathbf{B}_k\mathbf{B}_k^{\mathrm{H}}\right) - E_{tx}\right). \tag{12}$$

We define $\nabla_{\mathbf{B}_k} f = \frac{\partial f}{\partial \mathbf{B}_k^*}$ as the complex gradient operator. The gradient is a matrix with the $[n,m]^{\mathrm{th}}$ element defined as: $[\nabla_{\mathbf{B}_k} f]_{nm} = \nabla_{[\mathbf{B}_k]_{nm}} f = \frac{\partial f}{\partial [\mathbf{B}_k^*]_{nm}}$. From the KKT conditions a local optimum must satisfy for all k: $\nabla_{\mathbf{B}_k} f = \mathbf{0}$, and $\nabla_\lambda f = 0$.

[1] Without loss of generality each user is assumed to have Q receive antennas. Users with fewer antennas can be emulated by nulling the corresponding channel gains.

[2] The noise covariance matrix $\mathbf{R}_{vv}$ can be assumed to be white without loss of generality after an appropriate whitening transform on the channel matrix.

$$\nabla_{\mathbf{B}_k} f = -u_k \mathbf{H}_k^{\mathrm{H}} \mathbf{R}_{\tilde{v}_k\tilde{v}_k}^{-1} \mathbf{H}_k \mathbf{B}_k \mathbf{E}_k + \left(\sum_{i=1, i\neq k}^{K} u_i \mathbf{H}_i^{\mathrm{H}} \mathbf{R}_{\tilde{v}_i\tilde{v}_i}^{-1} \mathbf{H}_i \mathbf{B}_i \mathbf{E}_i \mathbf{B}_i^{\mathrm{H}} \mathbf{H}_i^{\mathrm{H}} \mathbf{R}_{\tilde{v}_i\tilde{v}_i}^{-1} \mathbf{H}_i \right) \mathbf{B}_k + \lambda \mathbf{B}_k \tag{10}$$

$$\nabla_{\mathbf{B}_k} g = -\mathbf{H}_k^{\mathrm{H}} \mathbf{R}_{\tilde{v}_k\tilde{v}_k}^{-1} \mathbf{H}_k \mathbf{B}_k \mathbf{E}_k \mathbf{W}_k \mathbf{E}_k + \left(\sum_{i=1, i\neq k}^{K} \mathbf{H}_i^{\mathrm{H}} \mathbf{R}_{\tilde{v}_i\tilde{v}_i}^{-1} \mathbf{H}_i \mathbf{B}_i \mathbf{E}_i \mathbf{W}_i \mathbf{E}_i \mathbf{B}_i^{\mathrm{H}} \mathbf{H}_i^{\mathrm{H}} \mathbf{R}_{\tilde{v}_i\tilde{v}_i}^{-1} \mathbf{H}_i \right) \mathbf{B}_k + \lambda \mathbf{B}_k \tag{11}$$

The gradient w.r.t. the transmit filter $\mathbf{B}_k$ is computed by dividing the summation into different parts. The gradient is computed element-wise and using the chain rule:

$$\begin{aligned} \nabla_{[\mathbf{B}_k]_{nm}} R_k &= \mathrm{Tr}\left(\left(\frac{\partial R_k}{\partial \mathbf{E}_k^{-1}} \right)^{\mathrm{T}} \frac{\partial \mathbf{E}_k^{-1}}{\partial [\mathbf{B}_k^*]_{nm}} \right) \\ &= \mathrm{Tr}\left(\mathbf{E}_k \mathbf{e}_m \mathbf{e}_n{}^{\mathrm{T}} \mathbf{H}_k^{\mathrm{H}} \mathbf{R}_{\tilde{v}_k\tilde{v}_k}^{-1} \mathbf{H}_k \mathbf{B}_k \right) \\ &= \mathbf{e}_n{}^{\mathrm{T}} \mathbf{H}_k^{\mathrm{H}} \mathbf{R}_{\tilde{v}_k\tilde{v}_k}^{-1} \mathbf{H}_k \mathbf{B}_k \mathbf{E}_k \mathbf{e}_m. \end{aligned} \tag{13}$$

Here $\mathbf{e}_n$ is a unity-vector with one at the n^{th} element and zeros elsewhere. As $\nabla_{[\mathbf{B}_k]_{nm}} R_k = [\nabla_{\mathbf{B}_k} R_k]_{nm}$ we conclude:

$$\nabla_{\mathbf{B}_k} R_k = \mathbf{H}_k^{\mathrm{H}} \mathbf{R}_{\tilde{v}_k\tilde{v}_k}^{-1} \mathbf{H}_k \mathbf{B}_k \mathbf{E}_k. \tag{14}$$

Next, $\nabla_{\mathbf{B}_k} R_i$ is computed. Define first a real-valued scalar function h which depends on $\mathbf{K}$ through $\mathbf{S} = \mathbf{X} + \mathbf{LKCC}^{\mathrm{H}}\mathbf{K}^{\mathrm{H}}\mathbf{L}^{\mathrm{H}}$, where $\mathbf{X}, \mathbf{L}, \mathbf{C}$ are fixed matrices independent of $\mathbf{K}$. Using the element-wise derivation method it can be shown that $\nabla_{\mathbf{K}} h = \mathbf{L}^{\mathrm{H}} (\nabla_{\mathbf{S}} h) \mathbf{LKCC}^{\mathrm{H}}$. Considering $\mathbf{S}$ as the effective noise covariance matrix $\mathbf{R}_{\tilde{v}_i\tilde{v}_i}$, we have

$$\nabla_{\mathbf{B}_k} R_i = \mathbf{H}_i^{\mathrm{H}} \left(\nabla_{\mathbf{R}_{\tilde{v}_i\tilde{v}_i}} R_i \right) \mathbf{H}_i \mathbf{B}_k. \tag{15}$$

Furthermore, we compute

$$\nabla_{\mathbf{R}_{\tilde{v}_i\tilde{v}_i}} R_i = \mathbf{R}_{\tilde{v}_i\tilde{v}_i}^{-1} \mathbf{H}_i \mathbf{B}_i \mathbf{E}_i \mathbf{B}_i^{\mathrm{H}} \mathbf{H}_i^{\mathrm{H}} \mathbf{R}_{\tilde{v}_i\tilde{v}_i}^{-1}. \tag{16}$$

Combining (15) and (16) we obtain:

$$\nabla_{\mathbf{B}_k} R_i = -\mathbf{H}_i^{\mathrm{H}} \mathbf{R}_{\tilde{v}_i\tilde{v}_i}^{-1} \mathbf{H}_i \mathbf{B}_i \mathbf{E}_i \mathbf{B}_i^{\mathrm{H}} \mathbf{H}_i^{\mathrm{H}} \mathbf{R}_{\tilde{v}_i\tilde{v}_i}^{-1} \mathbf{H}_i \mathbf{B}_k. \tag{17}$$

The results (14) and (17) are finally used to obtain the WSR-gradient expression of equation (10).

B. Gradient of Weighted Minimum Mean Square Error Minimization

The new optimization problem we investigate is the WMMSE transmit filter design problem assuming that MMSE receive filtering is applied:

$$\begin{aligned} [\mathbf{B}_1^{\mathrm{WMMSE}}, \cdots, \mathbf{B}_K^{\mathrm{WMMSE}}] &= \arg\min_{\mathbf{B}_1, \cdots, \mathbf{B}_K} \sum_k \mathrm{Tr}\left(\mathbf{W}_k \mathbf{E}_k\right) \\ \text{s.t.} &\sum_k \mathrm{Tr}\left(\mathbf{B}_k \mathbf{B}_k^{\mathrm{H}}\right) = E_{tx}. \end{aligned} \tag{18}$$

The matrix $\mathbf{W}_k \in \mathbb{C}^{[Q_k \times Q_k]}$ is a constant weight matrix associated with user k. The Lagrangian reads:

$$g(\mathbf{B}_1 \cdots \mathbf{B}_K) = \sum_k \mathrm{Tr}\left(\mathbf{W}_k \mathbf{E}_k\right) + \lambda \left(\sum_k \mathrm{Tr}\left(\mathbf{B}_k \mathbf{B}_k^{\mathrm{H}}\right) - E_{tx} \right). \tag{19}$$

The gradient $\nabla_{\mathbf{B}_k} g$ is computed is a similar manner as $\nabla_{\mathbf{B}_k} f$ by considering the different parts of the summation. Firstly,

$$\nabla_{\mathbf{B}_k} \mathrm{Tr}\left(\mathbf{W}_k \mathbf{E}_k\right) = -\mathbf{H}_k^{\mathrm{H}} \mathbf{R}_{\tilde{v}_k\tilde{v}_k}^{-1} \mathbf{H}_k \mathbf{B}_k \mathbf{E}_k \mathbf{W}_k \mathbf{E}_k. \tag{20}$$

Secondly,

$$\nabla_{\mathbf{B}_k} \mathrm{Tr}\left(\mathbf{W}_i \mathbf{E}_i\right) = \mathbf{H}_i^{\mathrm{H}} \mathbf{R}_{\tilde{v}_i\tilde{v}_i}^{-1} \mathbf{H}_i \mathbf{B}_i \mathbf{E}_i \mathbf{W}_i \mathbf{E}_i \mathbf{B}_i^{\mathrm{H}} \mathbf{H}_i^{\mathrm{H}} \mathbf{R}_{\tilde{v}_i\tilde{v}_i}^{-1} \mathbf{H}_i \mathbf{B}_k. \tag{21}$$

Combining (20) and (21), we obtain the WMMSE-gradient expression of equation (11). Notice also that $\nabla_\lambda f = \nabla_\lambda g = \sum_k \mathrm{Tr}\left(\mathbf{B}_k \mathbf{B}_k^{\mathrm{H}}\right) - E_{tx}$.

C. Comparison of WSR and WMMSE gradients

Comparing (10) and (11) it is clear that the two problems are closely related. In fact, for a given set of transmit filters $\mathbf{B}_1, \cdots, \mathbf{B}_K$ and corresponding MMSE-matrices $\mathbf{E}_1, \cdots, \mathbf{E}_K$, the WMMSE-gradient can be made identical to the WSR-gradient, if the following MSE-weights are selected for all k:

$$\mathbf{W}_k = u_k \mathbf{E}_k^{-1}. \tag{22}$$

Consider next that we have a WSR optimal point, i.e. where $\forall k \nabla_{\mathbf{B}_k} f = \mathbf{0}$, $\nabla_\lambda f = 0$, and that the set of transmit filters and corresponding MMSE-matrices at this point are resp.: $\mathbf{B}_1^{\mathrm{WSR}}, \cdots, \mathbf{B}_K^{\mathrm{WSR}}$ and $\mathbf{E}_1^{\mathrm{WSR}}, \cdots, \mathbf{E}_K^{\mathrm{WSR}}$. If this set of MMSE-matrices is used to compute a set of MSE-weights according to (22), then the KKT-conditions for the WMMSE-problem are satisfied for the same set of transmit filters, i.e. the point is also a WMMSE optimal point with $\mathbf{B}_k^{\mathrm{WMMSE}} = \mathbf{B}_k^{\mathrm{WSR}} \forall k$.

The fact that the KKT-conditions of the two problems can be satisfied simultaneously suggests that it is possible to solve the WSR-problem through the use of WMMSE and a proper set of MSE-weights.

IV. Alternating Optimization for Finding a Local WSR-Optimum

The basic idea is to alternate between WMMSE optimization of $\mathbf{B}_1, \cdots, \mathbf{B}_K$ and the MSE weight update for $\mathbf{W}_1, \cdots, \mathbf{W}_K$ based on (22). If this iterative process converges, it converges to a fixed point, which is also a stationary point of the WSR-objective function.

There are various ways to implement such an alternating optimization process and in particular, the weight matrices can be updated at different stages. As described above, one possibility is to update the MSE-weights after each update of the transmit filters. This method has been tested and gives good performance. The method we tested requires an inner loop to perform WMMSE optimization over $\mathbf{B}_1, \cdots, \mathbf{B}_K$ for fixed MSE-weights. Specifically, the inner loop performs alternating optimization between MMSE-receive filters and WMMSE transmit filters according to the method described in [16]. Overall, the disadvantage of this method is the number of required iterations, since inner iterations has to be performed for each weight update. Fortunately, it turns out that the inner iterations are not needed to obtain an algorithm which converges to a local WSR-optimum. Therefore we propose an algorithm which contains a single loop where respectively MSE-weights and transmit/receive filters are updated. In summary, the proposed algorithm, which we coin Weighted Sum-Rate maximization Beamforming using Weighted sum-Minimum Mean Square Error (WSRBF-WMMSE) is:

PROPOSED ALGORITHM: WSRBF-WMMSE
set $n = 0$
set $\mathbf{B}_k^n = \mathbf{B}_k^{\text{init}} \forall k$
iterate
  update $n = n + 1$
  I. *compute* $\mathbf{A}_k^n | \mathbf{B}_i^{n-1} \forall i$ for all k using (7)
  II. *compute* $\mathbf{W}_k^n | \mathbf{B}_i^{n-1} \forall i$ for all k using (22),(8)
  III. *compute* $\mathbf{B}^n | \mathbf{A}^n, \mathbf{W}^n$ using (23),(24)
until convergence

The first step updates the MMSE-receive filters given the transmit filters from previous iteration. The second step updates the MSE-weights given the transmit filters as a function of the MMSE-matrix. The third step computes the WMMSE transmit filters given the receive filters and MSE-weights. The problem of computing the MMSE transmit filter $\mathbf{B}_{[P \times QK]} = [\mathbf{B}_1, \cdots, \mathbf{B}_K]$ for fixed receive filters was treated in [17] for the unweighed MMSE-case, but the extension to WMMSE is straightforward. The WMMSE transmit filter structure is computed as:

$$\bar{\mathbf{B}} = (\mathbf{H}^{\text{H}}\mathbf{A}^{\text{H}}\mathbf{W}\mathbf{A}\mathbf{H} + \frac{\text{Tr}\left(\mathbf{W}\mathbf{A}\mathbf{A}^{\text{H}}\right)}{E_{tx}}\mathbf{I}_P)^{-1}\mathbf{H}^{\text{H}}\mathbf{A}^{\text{H}}\mathbf{W}, \quad (23)$$

where $\mathbf{W}_{[QK \times QK]} = \text{diag}\{\mathbf{W}_1, \cdots, \mathbf{W}_K\}$ and $\mathbf{A}_{[QK \times QK]} = \text{diag}\{\mathbf{A}_1, \cdots, \mathbf{A}_K\}$ are block-diagonal matrices, and $\mathbf{H}_{[QK \times P]} = \left[\mathbf{H}_1{}^{\text{T}}, \cdots, \mathbf{H}_K{}^{\text{T}}\right]^{\text{T}}$ contains the different channel matrices stacked row-wise. Similar to the unweighted MMSE-case [17], the final transmit filter is computed as:

$$\mathbf{B}^{\text{WMMSE}} = b\bar{\mathbf{B}}, \quad (24)$$

where $b = \sqrt{\frac{E_{tx}}{\text{Tr}\left(\bar{\mathbf{B}}\bar{\mathbf{B}}^{\text{H}}\right)}}$ is a gain factor which scales the signal so as to satisfy the transmit power constraint. Following [17], $\mathbf{B}^{\text{WMMSE}}$ is derived under the assumption that all receive filters are rescaled by $\frac{1}{b}$.

A. Convergence analysis

Following the reasoning of section III-C it is clear that if the proposed algorithm is initialized by $\mathbf{B}_k^{\text{WSR}} \forall k$, then after one iteration it will return the optimal filter again, i.e. $\mathbf{B}_k^n = \mathbf{B}_k^{\text{WSR}} \forall k$. Convergence in the general case is proven by proving monotonic convergence of an equivalent optimization problem, which is based on expanding the cost $-u_{R_k} \log\det\left(\mathbf{E}_k^{-1}\right)$ to include the MSE-weights and receive filters as optimization variables in addition to the transmit filters.

We define $\tilde{\mathbf{E}}_k = \mathbb{E}\left[\left(\mathbf{A}_k\mathbf{y}_k - \mathbf{d}_k\right)\left(\mathbf{A}_k\mathbf{y}_k - \mathbf{d}_k\right)^{\text{H}}\right]$, the MSE-matrix. Note that it is different from $\mathbf{E}_k$ defined in (8). Consider the following cost function:

$$\tilde{l}_k(\mathbf{W}_k, \mathbf{A}_k, \mathbf{B}_i \forall i) = \text{Tr}\left(\mathbf{W}_k\tilde{\mathbf{E}}_k\right) - u_{R_k}\log\det\left(u_{R_k}^{-1}\mathbf{W}_k\right) - u_{R_k}Q. \quad (25)$$

and the following objective:

$$[\mathbf{B}_k^{\text{WSR}} \forall k] = \underset{\mathbf{B}_k \forall k, \mathbf{A}_k \forall k, \mathbf{W}_k \forall k}{\arg\min} \sum_k \tilde{l}_k(\mathbf{W}_k, \mathbf{A}_k, \mathbf{B}_i \forall i) \quad \text{s.t.} \sum_k \text{Tr}\left(\mathbf{B}_k\mathbf{B}_k^{\text{H}}\right) \leq E_{tx}. \quad (26)$$

We first prove that the optimization wrt the transmit filters $\mathbf{B}_k$ using this criterion as the same as the original WSR optimization (4). First we minimize $\tilde{l}_k(\mathbf{W}_k, \mathbf{A}_k, \mathbf{B}_i \forall i)$ w.r.t. $\mathbf{A}_k$ considering weights and transmit filters fixed. The minimizing value is unique and is denoted as $\mathbf{A}_k^{\text{MMSE}}(\mathbf{B}_i \forall i)$ (see eq. (7)). The variable $\mathbf{A}_k$ intervene only in $\tilde{\mathbf{E}}_k$ and substituting by $\mathbf{A}_k^{\text{MMSE}}(\mathbf{B}_i \forall i)$, $\tilde{\mathbf{E}}_k$ becomes equal to $\mathbf{E}_k$. Hence, we get a new cost function for the transmit filters and the weights:

$$l_k(\mathbf{W}_k, \mathbf{B}_i \forall i) = \text{Tr}\left(\mathbf{W}_k\mathbf{E}_k\right) - u_{R_k}\log\det\left(u_{R_k}^{-1}\mathbf{W}_k\right) - u_{R_k}Q. \quad (27)$$

Minimizing $l_k(\mathbf{W}_k, \mathbf{B}_i \forall i)$ w.r.t. $\mathbf{W}_k$ leads to $\mathbf{W}_k^{\min}(\mathbf{B}_i \forall i) = u_{R_k}\mathbf{E}_k^{-1}(\mathbf{B}_i \forall i)$. Substituting $\mathbf{W}_k$ in $l_k(\mathbf{W}_k, \mathbf{B}_i \forall i)$ by $\mathbf{W}_k^{\min}$ we get the cost function $-u_{R_k}\log\det\left(\mathbf{E}_k^{-1}\right)$, which corresponds to the original WSR-cost.

Now, we prove that alternating minimization of the cost $\sum_k \tilde{l}_k(\cdot)$ in (26) corresponds to the steps I,II,III of WSRBF-WMMSE. When $\mathbf{W}_k \forall k$ is constant, the cost function is $\sum_k \text{Tr}\left(\mathbf{W}_k\tilde{\mathbf{E}}_k(\mathbf{A}_k, \mathbf{B}_i \forall i)\right)$, so in the alternating minimization process: finding $\mathbf{A}_k$ given $\mathbf{B}_i^{n-1} \forall i$ gives the same result as in step I and finding $\mathbf{B}_k \forall k$ given $\mathbf{A}_k^n \forall k$ and $\mathbf{W}_k^n \forall k$ gives the same results as step III. Optimization w.r.t. $\mathbf{W}_k$ with $\mathbf{A}_k$ and $\mathbf{B}_i \forall i$ fixed gives $\mathbf{W}_k = u_{R_k}\tilde{\mathbf{E}}_k^{-1}(\mathbf{A}_k^n, \mathbf{B}_i^{n-1} \forall i) = \mathbf{E}_k^{-1}(\mathbf{B}_i^{n-1} \forall i)$. So we have same result as step II.

Due to the alternating minimization process, the cost $\sum_k \tilde{l}_k(\cdot)$ decreases monotonically. Furthermore, assuming that the minimal value of the cost (max WSR-value) exists, the cost function is lower bounded. We conclude that we have convergence to a local optimum. Note that the original WSR-cost does not necessarily experience a monotonic convergence, although simulation results show that this is often the case.

V. NUMERICAL EXAMPLES

This section evaluates sum-rate performance for the MIMO downlink for different system settings using Monte-Carlo simulation. In all simulations the number of transmit antennas is fixed to $P = 4$. The elements of the channel matrices are generated as i.i.d. Gaussian random variables $\mathcal{CN}(0, \sigma_h^2)$ and the receive noise covariances are normalized, i.e. $\mathbf{R}_{v_k v_k} = \mathbf{I}_Q$. Since both noise and data covariance are normalized, we define SNR as σ_h^2.

First we study the convergence properties by comparing WSRBF-WMMSE to the recently proposed ALGORITHM 1 of [11] which also converges to a local WSR-optimum. For both algorithms we choose a simple initialization by the transmit matched filter, i.e. $\forall k \mathbf{B}_k^{\text{init}} = b\mathbf{H}_k^{\text{H}}$, where b is selected so as to satisfy the transmit power constraint. Sum-rate convergence results for four different scenarios are shown in Figure 1. Overall the plots indicate that convergence speed is comparable for the two methods, although the convergence speed varies for the individual channel realizations. From a complexity point of view the proposed algorithm has the advantage that it does

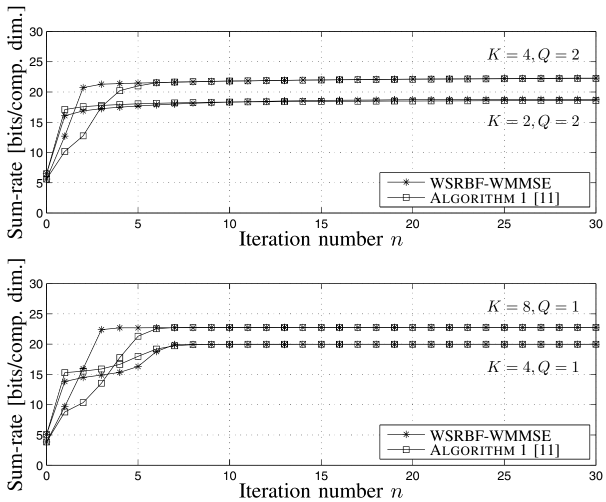

Fig. 1. Sum-rate maximization convergence examples with $P = 4$ transmit antennas and with different combinations of users K and receive antennas Q. Each combination is simulated for a random channel realization (SNR=20dB).

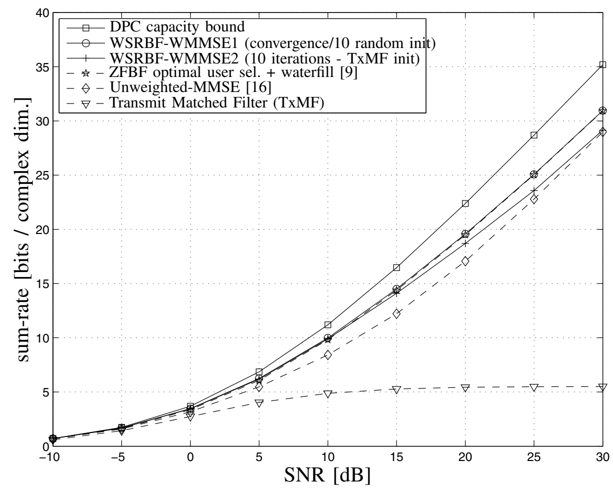

Fig. 2. Average sum-rate performance in 1000 random channels with $P = 4$ transmit antennas and $K = 4$ single receive antenna users ($Q = 1$).

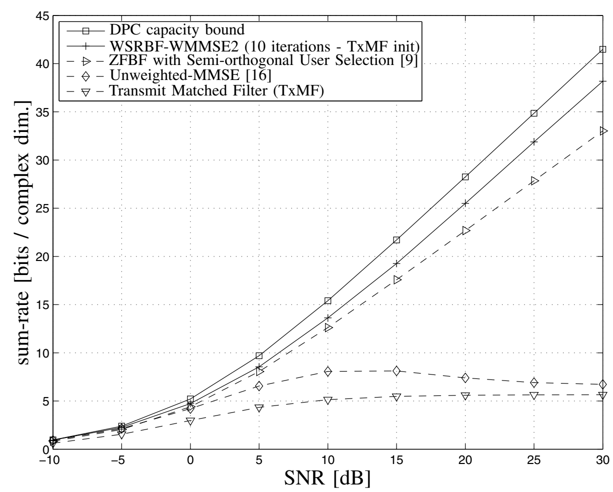

Fig. 3. Average sum-rate performance in 1000 random channels with $P = 4$ transmit antennas and $K = 20$ single receive antenna users ($Q = 1$).

not require solving a GP in each iteration as ALGORITHM 1 of [11]. Solving a GP[3] has a worst-case polynomial time complexity in the number of variables (KQ) [11].

In the following three simulations we consider sum-rate performance in different MIMO-BC scenarios. Since the WSR-problem is non-convex, the initialization $\mathbf{B}_k^{\text{init}} \forall k$ determines if the WSR-optimum obtained after iterations will be local or global. Currently it is unknown how to choose the initialization such that the global optimum is guaranteed. In our simulations we use two versions of WSRBF-WMMSE: 1) Using 10 random filter initializations and eventually selecting the best result, and 2) Using the simple transmit matched filter initialization and allowing only 10 iterations. The first version is chosen to obtain a high performing solution although there is no guarantee that the global optimum is found. The second version is chosen to study the performance of a potentially practical low complexity solution. The presented results are averaged over 1000 channel realizations. DPC sum-capacity reference bounds are produced using ALGORITHM 2 from [18]. In the first two simulations the number of users is varied as $K = \{4, 20\}$ and each user has a single receive antenna. In the third simulation the number of users is varied as $K = \{1, 2, 6\}$ and each user has two receive antennas.

Figure 2 shows the average sum-rate performance in a $K = 4$ setting. As performance reference we have used a Zero-Forcing Beamforming (ZFBF)-based algorithm [9]. This algorithm tries all combinations of scheduled users, and computes the ZF-filter with the optimal power levels (waterfilling) for each combination. The combination with highest sum-rate is selected. The plots show that sum-rate maximization leads to only a marginal improvement of WSRBF-WMMSE1 as compared to the ZFBF-based algorithm. WSRBF-WMMSE2 sees a loss at high SNR's but performs well at SNR's up to 15 dB. In the simulations, it was noticed that at low SNR, WSRBF-WMMSE allocates all transmit power to the user with the best channel. This phenomenon is similar to selection of the best user in the single-antenna degraded-broadcast channel to maximize sum-capacity [19]. As the SNR increases, more users are gradually supported simultaneously. To illustrate the role of the MSE-weights we have included the unweighted MMSE beamformer which is also computed using alternating optimization [16]. In fact, the MSE-weight update is the only difference between WSRBF-WMMSE and the unweighted MMSE algorithm [16], but as seen by Figure 2 and in the following Figure 3 the selection of weights at each iteration is crucial.

Figure 3 shows the sum-rate performance when the system has 20 users. Due to the system size, it is considered impractical to use several random initializations and we focus only on the algorithm version with fixed initialization. As a reference we have instead used the ZFBF with Semi-orthogonal User Selection (SUS) algorithm [9] which uses a

[3]We have used Matlab and CVX to solve the GP: "http://www.stanford.edu/ boyd/cvx/". Generating the curves for Figure 1 took less than a second for WSRBF-WMMSE, whereas it took approximately one hour for ALGORITHM 1 of [11].

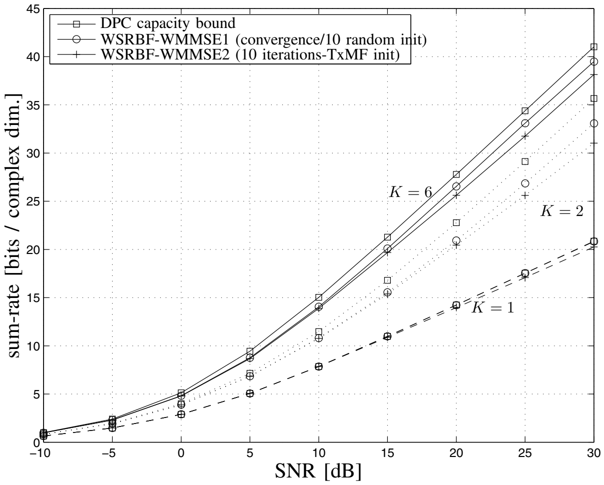

Fig. 4. Average sum-rate performance in 1000 random channels with $P = 4$ transmit antennas and $K = \{1, 2, 6\}$ users with $Q = 2$ receive antennas.

simplified procedure for finding a good subset of the users since exhaustive search is considered impractical[4]. Finding the best subset of the 20 users is a non-trivial problem, but the simple initialization by the transmit matched filter combined with only 10 iterations (WSRBF-WMMSE2) finds a good solution. Compared to the ZFBF-SUS algorithm, WSRBF-WMMSE2 clearly performs better in spite of its low complexity. At high Signal-to-Noise Ratio (SNR), WSRBF-WMMSE2 allocates non-zero rates to four users, corresponding to the rank of the channel. In this way WSRBF-WMMSE2 selects 4 of 20 users, rather than attempting to transmit data to all 20 users. The user selection is done automatically by the algorithm which nulls out some users through the weight update. In general the algorithm finds the subset of the users which has a good combination of having high channel gains while simultaneously being spatially compatible (nearly orthogonal). In contrast, the unweighted MMSE-solution (without initial user selection) transmits data to more than 4 users simultaneously which results in interference limitation at high SNR's. Figure 4 shows the sum-rate performance with a varying number of users where each user has $Q = 2$ receive antennas. For $K = 1$, i.e. the single user case, WSRBF-WMMSE1 achieves capacity (achieved by waterfilling over channel singular values). The single user problem is convex and therefore the transmit filter initialization does not matter. For $K = \{2, 6\}$, WSRBF-WMMSE1 obtains a slope comparable to the DPC capacity bounds with a loss on the order of 1-2 dB at high SNR. WSRBF-WMMSE2 performs equally well up to ≈10 dB, but is degraded by ≈1dB at high SNR. Notice that the ZF-based solutions [9] are developed only for single antenna receivers and are therefore not included in the plot.

VI. Conclusion

This paper studied beamforming design for the MIMO-BC to maximize weighted sum-rate. The paper found its motivation in recent results highlighting a relationship between mutual information and MMSE, and established a simple relation between weighted sum-rate and weighted MMSE in the MIMO-BC. As a result, a simple alternating optimization algorithm based on well-known transmit/receive MMSE-designs was proposed for finding a local weighted sum-rate optimum. Numerical results studying sum-rate show that the proposed algorithm achieves high performance, even when initiated by the simple transmit matched filter and allowing only few iterations. The algorithm is therefore a potential candidate for practical low complexity transmit beamforming implementations.

[4]For the ZFBF-SUS algorithm [9], the optimal value of the semi-orthogonality angle α, used in the SUS procedure, was found.

References

[1] E. Telatar, "Capacity of Multi-antenna Gaussian Channels," *Euro. Trans. Telecomm.*, vol. 10, no. 6, pp. 585–596, Nov. 1999.

[2] H. Weingarten, Y. Steinberg, and S. Shamai, "The Capacity Region of the Gaussian Multiple-Input Multiple-Output Broadcast Channel," *Information Theory, IEEE Transactions on*, vol. 52, no. 9, pp. 3936–3964, Sept. 2006.

[3] U. Erez, S. Shamai, and R. Zamir, "Capacity and Lattice-Strategies for Cancelling known interference," in *ISITA 2000*, Honolulu Hawaii USA, Nov. 2000, pp. 681–684.

[4] W. Yu and J. Cioffi, "Sum Capacity of Gaussian Vector Broadcast Channels," *IEEE Tran. on Info. Th.*, vol. 50, no. 9, pp. 1875–1892, Sep. 2004.

[5] J. Kusuma and K. Ramchandran, "Communicating by cosets and applications to broadcast," in *CISS*, March 2002.

[6] M. Schubert and H. Boche, "Iterative multiuser uplink and downlink beamforming under SINR constraints," *IEEE Trans. Sig. Proc.*, vol. 53, no. 7, pp. 2324–2334, July 2005.

[7] W. Yu and T. Lan, "Transmitter Optimization for the Multi-Antenna Downlink With Per-Antenna Power Constraints," *Signal Processing, IEEE Transactions on*, vol. 55, no. 6, pp. 2646–2660, June 2007.

[8] A. Wiesel, Y. Eldar, and S. Shamai, "Linear precoding via conic optimization for fixed MIMO receivers," *IEEE Trans. on Sig. Proc.*, vol. 54, no. 1, pp. 161–176, Jan 2006.

[9] T. Yoo and A. Goldsmith, "On the Optimality of Multi-Antenna Broadcast Scheduling Using Zero-Forcing Beamforming," in *IEEE JSAC, Special Issue on 4G Wireless Systems*, vol. 24, no. 3, March 2006, pp. 528–541.

[10] M. Kobayashi and G. Caire, "An Iterative Water-Filling Algorithm for Maximum Weighted Sum-rate of Gaussian MIMO-BC," *Selected Areas in Communications, IEEE Journal on*, vol. 24, no. 8, pp. 1640–1646, Aug. 2006.

[11] S. Shi, M. Schubert, and H. Boche, "Rate Optimization for Multiuser MIMO Systems With Linear Processing," *Signal Processing, IEEE Transactions on*, vol. 56, no. 8, pp. 4020–4030, Aug. 2008.

[12] M. Codreanu, A. Tolli, M. Juntti, and M. Latva-aho, "MIMO Downlink Weighted Sum Rate Maximization with Power Constraint per Antenna Groups," *IEEE VTC Spring*, April 2007.

[13] D. Guo, S. S. (Shitz), and S. Verdu, "Mutual Information and Minimum Mean-Square Error in Gaussian Channels," *Information Theory, IEEE Transactions on*, vol. 51, no. 4, pp. 1261–1282, April 2005.

[14] H. Sampath, P. Stoica, and A. Paulraj, "Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion," *Communications, IEEE Transactions on*, vol. 49, no. 12, pp. 2198–2206, Dec 2001.

[15] D. Palomar and S. Verdu, "Gradient of mutual information in linear vector Gaussian channels," *Information Theory, IEEE Transactions on*, vol. 52, no. 1, pp. 141–154, Jan. 2006.

[16] R. Hunger, W. Utschick, D. Schmidt, and M. Joham, "Alternating Optimization for MMSE Broadcast Precoding," in *Proc. IEEE ICASSP*, vol. 4, 2006, pp. IV–757–IV–760.

[17] M. Joham, K. Kusume, M. Gzara, W. Utschick, and J. Nossek, "Transmit Wiener filter for the downlink of TDDDS-CDMA systems," vol. 1, September 2002, pp. 9–13 vol.1.

[18] N. Jindal, W. Rhee, S. Vishwanath, S. Jafar, and A. Goldsmith, "Sum power iterative water-filling for multi-antenna Gaussian broadcast channels," *Information Theory, IEEE Transactions on*, vol. 51, no. 4, pp. 1570–1580, April 2005.

[19] T. Cover and J. Thomas, *Elements of Information Theory*. Wiley & Sons, Inc., 1991.