The Geometry of the MIMO Broadcast Channel Rate Region Under Linear Filtering at High SNR

Raphael Hunger, Paul de Kerret, and Michael Joham
Associate Institute for Signal Processing, Technische Universität München, 80290 Munich, Germany
Telephone: +49 89 289-28508, Fax: +49 89 289-28504, Email: {hunger, joham}@tum.de

Abstract— **We present a high SNR analysis of the rate region geometry when linear filtering is applied in the MIMO broadcast channel and time sharing is not considered. In particular, the set of weakly Pareto optimal points of the asymptotic rate region is derived. While all different permutations of user sorting have to be investigated for the complete rate region when nonlinear dirty paper coding is applied, we show that the analogon to different sorting orders is the number of active streams per user when linear filtering is applied. Furthermore, we reveal that only a small fraction of weakly Pareto optimal points is in charge of maximizing an asymptotic weighted sum rate and points belonging to that fraction are all obtained when every user applies full multiplexing. In contrast, switching off some data streams may be optimal when it comes to balancing or maximizing a user's rate while keeping the other users' rates constant. Many of the derived results are not only applicable to systems with enough antennas at the base station, but also to configurations with too few degrees of freedom, which so far have never been discussed in any asymptotic analysis before.**

I. Introduction

The analysis of the capacity region of the MIMO broadcast channel has so far been limited to the case when dirty paper coding is applied with activated time sharing, see for example [1], [2]. Under this setup, the convex hull of the rate region can easily be obtained by means of a weighted sum rate maximization, since the weighted sum rate is then a concave utility, and many globally optimum algorithms solving this optimization are meanwhile available, see [3], [4], [5], and [6]. In case of linear filtering, even the convex hull of the rate region with enabled time sharing is hard to compute since the unweighted sum rate maximization is already nonconcave [7]. Although the algorithm in [7] can straightforwardly be extended to the case of different weights, the rate region itself *without* time sharing or convex hull operation is still an open problem. This deflating circumstance is fortunately alleviated in the high power regime. Early work focused on the maximization of the weighted sum rate in the ergodic case when multi-antenna terminals are involved [8], [9], and the extension for instantaneous channel realizations was given in [10], [11]. There, a novel rate duality for linear filtering was used [12], and the dual multiple access system was investigated instead of the original broadcast channel system. However, the weighted sum rate approach inherently allows only to compute the convex hull of the region and not the region itself. Our contribution is a description of the weakly Pareto optimal points of the asymptotic rate region without activated time sharing. We show that only specific stream allocations and transmit covariance matrices are relevant for the border of the rate region.

This work was supported by Deutsche Forschungsgemeinschaft (DFG) under fund Ut 36/9-1.

Notation: Matrices and vectors are upper and lower case bold, respectively. $\mathbb{S}_M$ denotes the set of $M \times M$ positive semidefinite matrices and $\mathbf{I}_M$ is the $M \times M$ identity matrix. The operators $\|\cdot\|_\mathrm{F}$, $|\cdot|$, and $(\cdot)^\mathrm{H}$ stand for Frobenius norm, determinant, and Hermitian transposition, respectively.

II. System Model and Definitions

Recent results on the rate duality between the MIMO broadcast channel (BC) and the MIMO multiple access channel (MAC) allow us to switch from the original broadcast channel to a dual multiple access channel with the same sum power constraint, see [12]. Since both rate regions are congruent, we can characterize the broadcast channel rate region by analyzing the multiple access channel region. This dual MAC has the advantageous property that less variables are required to express the transmit matrices than in the BC. Moreover, if a user transmits as many data streams as he has antennas, the MAC transmit covariance matrix will have full rank, which can be exploited in the rate expression formula. In the BC, however, the transmit covariance matrix for a particular user will never be full rank if the base station has more antennas than every user. Let $\boldsymbol{T}_k \in \mathbb{C}^{M_k \times B_k}$ denote the precoding matrix of user k in the dual MAC which maps the B_k dimensional symbol vector $\boldsymbol{s}_k$ onto his M_k antennas. The precoded symbol vector $\boldsymbol{T}_k\boldsymbol{s}_k$ then propagates over the frequency flat channel $\boldsymbol{H}_k \in \mathbb{C}^{N \times M_k}$. At the N-antenna receiver, zero-mean Gaussian noise $\boldsymbol{\eta} \in \mathbb{C}^N$ with identity covariance matrix is added. Unless otherwise noted, we make the assumption that the base station has at least as many antennas as the user terminals have in sum, i.e., $N \geq M := \sum_{k=1}^K M_k$. We will frequently make use of the following definitions:

Definition II.1. *Two functions f and g are said to be weakly asymptotically equivalent, if*

$$\lim_{P\to\infty} \frac{f(P)}{g(P)} = 1,$$

and we shall use the notation $f \doteq g$.

A stricter definition is the following:

Definition II.2. *Two functions f and g are said to be strongly asymptotically equivalent, if*

$$\lim_{P\to\infty}\big[f(P)-g(P)\big]=0,$$

and we shall use the notation $f \cong g$.

Note that strong asymptotic equivalence always implies weak asymptotic equivalence (unless $g(P)=0$ independent of P). The converse, however, is usually not true.

III. Rate Region Geometry at High SNR

The rate region under linear filtering corresponds to the union of all rate tuples that are feasible under a sum power constraint P and a fixed number of active streams per user. Note that the union is taken over all meaningful active stream allocations that satisfy $B_k \leq M_k$ and $\sum_{k=1}^{K} B_k \leq N$. In other words, no user may have more active data streams than he has antennas and the total number of active data streams is upper bounded by the number N of antennas at the base station, which is then automatically fulfilled by our assumption $N \geq M$. Due to our focus on the high SNR regime, only few stream allocations actually contribute to the weakly Pareto optimal points of the rate region. In the two user case, those relevant allocations are characterized by the property that at least one user has to apply full multiplexing whereas the other user may have zero up to his number of antennas active streams. This can easily be understood by noticing that the set of feasible rate pairs in a two user system at high SNR where each user has only one active stream is a subset of the rate tuples that are feasible when both users have two active streams. Of course, both users need to be equipped with at least two antennas in that case.

For simplicity, we restrict ourselves to the rate region of the two user case. Under linear filtering and Gaussian signaling, user one sees interference from user two and vice versa. Thus, the rates of the two users can be expressed as

$$\begin{aligned}R_1&=\log_2\left|\mathbf{I}_{M_1}+\boldsymbol{H}_1^{\mathrm{H}}(\mathbf{I}_N+\boldsymbol{H}_2\boldsymbol{Q}_2\boldsymbol{H}_2^{\mathrm{H}})^{-1}\boldsymbol{H}_1\boldsymbol{Q}_1\right|,\\ R_2&=\log_2\left|\mathbf{I}_{M_2}+\boldsymbol{H}_2^{\mathrm{H}}(\mathbf{I}_N+\boldsymbol{H}_1\boldsymbol{Q}_1\boldsymbol{H}_1^{\mathrm{H}})^{-1}\boldsymbol{H}_2\boldsymbol{Q}_2\right|,\end{aligned}\tag{1}$$

where $\boldsymbol{Q}_k := \boldsymbol{T}_k\boldsymbol{T}_k^{\mathrm{H}} \in \mathbb{S}_{M_k}$ is the transmit covariance matrix of user k.

A. Full Multiplexing of All Data Streams

As will be shown later, full multiplexing is the optimum transmission strategy when the rates of both users are large, which can only be achieved if all eigenvalues of $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$ grow beyond all limits with increasing P. Under this assumption, the two inverses in (1) are strongly asymptotically equivalent to the projector

$$\begin{aligned}(\mathbf{I}_N+\boldsymbol{H}_k\boldsymbol{Q}_k\boldsymbol{H}_k^{\mathrm{H}})^{-1}&=\mathbf{I}_N-\boldsymbol{H}_k(\boldsymbol{Q}_k^{-1}+\boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{H}_k)^{-1}\boldsymbol{H}_k^{\mathrm{H}}\\ &\cong \mathbf{I}_N-\boldsymbol{H}_k(\boldsymbol{H}_k^{\mathrm{H}}\boldsymbol{H}_k)^{-1}\boldsymbol{H}_k^{\mathrm{H}}\end{aligned}\tag{2}$$

with $k\in\{1,2\}$ in the sense of Definition II.2. Using (2), the rate pair R_1 and R_2 is strongly asymptotically equivalent to the rate pair

$$\begin{aligned}R_1'&=\log_2\left|\mathbf{I}_{M_1}+\boldsymbol{H}_1^{\mathrm{H}}(\mathbf{I}_N-\boldsymbol{H}_2(\boldsymbol{H}_2^{\mathrm{H}}\boldsymbol{H}_2)^{-1}\boldsymbol{H}_2^{\mathrm{H}})\boldsymbol{H}_1\boldsymbol{Q}_1\right|,\\ R_2'&=\log_2\left|\mathbf{I}_{M_2}+\boldsymbol{H}_2^{\mathrm{H}}(\mathbf{I}_N-\boldsymbol{H}_1(\boldsymbol{H}_1^{\mathrm{H}}\boldsymbol{H}_1)^{-1}\boldsymbol{H}_1^{\mathrm{H}})\boldsymbol{H}_2\boldsymbol{Q}_2\right|.\end{aligned}\tag{3}$$

Obviously, $R_k' < R_k$, but in the asymptotic limit the difference $R_k - R_k'$ is zero, and therefore, $R_k \cong R_k'$. Allocating powers p_1 and $p_2 = P - p_1$ to users one and two, respectively, the optimum transmit covariance matrices $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$ share the same eigenbases as their respective projected channel Grams and their power allocation follows from the single-user water-filling principle with power p_1 or p_2, see [13] for example.

Theorem III.1: *For a system where every user applies full multiplexing and the base station is equipped with enough antennas, the weakly asymptotically optimum mode power allocation is uniform and therefore, the weakly asymptotically optimum transmit covariance matrices are scaled identities whose traces grow beyond all limits with increasing total transmit power.*

For the proof, we use the fact that the two rate expressions R_1' and R_2' in (3) are not coupled by the two transmit covariance matrices $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$, as long as all eigenvalues of $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$ grow beyond all limits when $P\to\infty$. Given this property, the water-filling power allocation is optimum with respect to (3), and the water-levels grow beyond all limits as long as both $p_1\to\infty$ and $p_2\to\infty$. Since the difference of two mode powers for a particular user does not depend on the water-level but solely on the inverse eigenvalues of the projected channel Gram $\bar{\boldsymbol{H}}_i^{\mathrm{H}}\bar{\boldsymbol{H}}_i$ (see [13]), the power allocation of different modes is never strongly asymptotically equivalent to a uniform power allocation (unless all eigenvalues of $\bar{\boldsymbol{H}}_i^{\mathrm{H}}\bar{\boldsymbol{H}}_i$ are equal). Note that $\bar{\boldsymbol{H}}_1 = (\mathbf{I}_N-\boldsymbol{H}_2(\boldsymbol{H}_2^{\mathrm{H}}\boldsymbol{H}_2)^{-1}\boldsymbol{H}_2^{\mathrm{H}})\boldsymbol{H}_1$ represents the channel $\boldsymbol{H}_1$ of user one that has been projected into the null space of $\boldsymbol{H}_2$ and the same holds for $\bar{\boldsymbol{H}}_2$ with reversed indices. However, the mode power allocation is affine and scales linearly with the water-level. Therefore, the ratio of two mode powers tends to one when the water-level goes to infinity and the power allocations is thus weakly asymptotically equivalent to a uniform power allocation, for which the eigenbasis of $\boldsymbol{Q}_i$ may be chosen freely. ■

Although a uniform mode power allocation for $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$ according to

$$\boldsymbol{Q}_1=\frac{p_1}{M_1}\mathbf{I}_{M_1}\quad\text{and}\quad \boldsymbol{Q}_2=\frac{p_2}{M_2}\mathbf{I}_{M_2}$$

is only weakly asymptotically equivalent to the optimum one, the resulting rate expressions

$$R_k''=M_k\log_2 p_k - M_k\log_2 M_k+\log_2\left|\bar{\boldsymbol{H}}_k^{\mathrm{H}}\bar{\boldsymbol{H}}_k\right|\tag{4}$$

are strongly asymptotically equivalent to R_1 and R_2, which is due to the properties of the log operator. Of course, (4) is only meaningful for those power allocations p_1 and $p_2 = P - p_1$, where the resulting expressions R_1'' and R_2'' are not negative. Since we neglected the identity in (3) to get (4), small values for p_1 and p_2 are not allowed in (4). However, we are again only interested in large rate pairs R_1'' and R_2'' that increase with P beyond all limits. For those, the difference $R_k - R_k''$ again goes down to zero as $P \to \infty$.

B. Less Active Streams Than Antennas

Not all possible stream configurations contribute to the weakly Pareto optimal points of the asymptotic rate region, when less streams are active than antennas are available at the user terminals. In the two user case, at least one user has to multiplex as many data streams as he can, whereas the other user may have zero up to his number of antennas minus one active streams. Without loss of generality, we now assume that user one applies full multiplexing with $B_1 = M_1$ whereas user two has $B_2 < M_2$ active streams. The reversed case can easily be obtained by interchanging indices. We focus on that part of the rate region where the rate of user one is very large with a multiplexing gain larger than $M_1 - 1$ (for nonlinear power allocations, the multiplexing gain can be noninteger), whereas the rate of user two varies from zero to the maximum value which is achievable with B_2 active streams. For this part, the covariance matrix $\boldsymbol{Q}_1$ will have eigenvalues that grow beyond all limits with increasing sum power P, and the rate of user two is therefore strongly asymptotically equivalent to

$$R_2' = \log_2 \left|\mathbf{I}_{B_2} + \boldsymbol{T}_2^{\mathrm{H}} \bar{\boldsymbol{H}}_2^{\mathrm{H}} \bar{\boldsymbol{H}}_2 \boldsymbol{T}_2\right|, \tag{5}$$

with $\boldsymbol{Q}_2 = \boldsymbol{T}_2 \boldsymbol{T}_2^{\mathrm{H}}$ and the projected channel matrix $\bar{\boldsymbol{H}}_2$, cf. the second equation in (3). Hence, the asymptotic rate of user two again depends only on his own beamforming matrix $\boldsymbol{T}_2$, and not on $\boldsymbol{Q}_1$. The asymptotic rate of user one, however, also depends on $\boldsymbol{T}_2$ because the rank of $\boldsymbol{T}_2$ is smaller than M_2. Hence, $\boldsymbol{Q}_2$ is not invertible and the projector approximation from (2) does not work. Using the first equation in (1) and neglecting the identity inside the determinant due to the assumption that $\boldsymbol{Q}_1$ has very large eigenvalues, the rate of user one $R_1 \cong R_1'$ is now strongly asymptotically equivalent to

$$R_1' = \log_2 \left|\boldsymbol{Q}_1\right| + \log_2 \left|\boldsymbol{H}_1^{\mathrm{H}} (\mathbf{I}_N + \boldsymbol{H}_2 \boldsymbol{T}_2 \boldsymbol{T}_2^{\mathrm{H}} \boldsymbol{H}_2^{\mathrm{H}})^{-1} \boldsymbol{H}_1\right|.$$

Some modifications lead to

$$R_1' = \log_2 \left|\boldsymbol{Q}_1\right| + \log_2 \left|\boldsymbol{H}_1^{\mathrm{H}} \boldsymbol{H}_1\right| + \log_2 \frac{\left|\mathbf{I} + \boldsymbol{T}_2^{\mathrm{H}} \bar{\boldsymbol{H}}_2^{\mathrm{H}} \bar{\boldsymbol{H}}_2 \boldsymbol{T}_2\right|}{\left|\mathbf{I} + \boldsymbol{T}_2^{\mathrm{H}} \boldsymbol{H}_2^{\mathrm{H}} \boldsymbol{H}_2 \boldsymbol{T}_2\right|} \tag{6}$$

where R_2' from (5) can be identified to be part of R_1' in (6). Letting user two have B_2 active streams, the rate pairs that are weakly Pareto optimal are obtained by finding the maximum asymptotic rate R_1' given R_2' and afterwards varying R_2'. For the asymptotic target rate $R_{2,\text{target}}'$, the optimization reads as

$$\begin{aligned} \underset{\boldsymbol{Q}_1, \boldsymbol{T}_2}{\text{maximize}}\, R_1' \quad & \text{s.t.: } R_2' = R_{2,\text{target}}',\ \boldsymbol{T}_2 \in \mathbb{C}^{M_2 \times B_2}, \\ & \boldsymbol{Q}_1 \succ \mathbf{0},\ \operatorname{tr}(\boldsymbol{Q}_1) + \|\boldsymbol{T}_2\|_{\mathrm{F}}^2 \leq P. \end{aligned} \tag{7}$$

Since only the determinant of $\boldsymbol{Q}_1$ is relevant, we may choose $\boldsymbol{Q}_1 = (P - \|\boldsymbol{T}_2\|_{\mathrm{F}}^2)/M_1 \cdot \mathbf{I}_{M_1}$ without loss of optimality. Then, (7) reduces to

$$\begin{aligned} \underset{\boldsymbol{T}_2 \in \mathbb{C}^{M_2 \times B_2}}{\text{maximize}}\ & \frac{(P - \|\boldsymbol{T}_2\|_{\mathrm{F}}^2)^{M_1}}{\left|\mathbf{I} + \boldsymbol{T}_2^{\mathrm{H}} \boldsymbol{H}_2^{\mathrm{H}} \boldsymbol{H}_2 \boldsymbol{T}_2\right|} \\ \text{s.t.:}\quad & \left|\mathbf{I} + \boldsymbol{T}_2^{\mathrm{H}} \bar{\boldsymbol{H}}_2^{\mathrm{H}} \bar{\boldsymbol{H}}_2 \boldsymbol{T}_2\right| = 2^{R_{2,\text{target}}'},\ \|\boldsymbol{T}_2\|_{\mathrm{F}}^2 \leq P. \end{aligned} \tag{8}$$

In this contribution, we cover the case with one active stream ($B_2 = 1$) for user two in detail, whereas the setup with more active streams ($B_2 \geq 2$) can be solved analogously by means of a projected gradient algorithm. However, the identities in the determinants in (8) have to be neglected then such that the multi-stream precoder solution is only applicable to large target rates $R_{2,\text{target}}$, for which the eigenvalues of the resulting matrix product $\boldsymbol{T}_2^{\mathrm{H}} \bar{\boldsymbol{H}}_2^{\mathrm{H}} \bar{\boldsymbol{H}}_2 \boldsymbol{T}_2$ are large compared to one. Note that the following analysis can also be applied if the base station has only $N = M_1 + 1$ antennas, which is less than $M = M_1 + M_2$ in general. When $B_2 = 1$, the beamforming matrix $\boldsymbol{T}_2$ reduces to the vector $\boldsymbol{t}_2$ which we split into its unit norm part $\boldsymbol{u}_2$ and its norm $\sqrt{p_2}$ via $\boldsymbol{t}_2 = \sqrt{p_2} \boldsymbol{u}_2$. The rate constraint in (8) is then satisfied by choosing

$$p_2 = \frac{c_2}{\boldsymbol{u}_2^{\mathrm{H}} \bar{\boldsymbol{H}}_2^{\mathrm{H}} \bar{\boldsymbol{H}}_2 \boldsymbol{u}_2} \tag{9}$$

with the substitution $c_2 := 2^{R_{2,\text{target}}'} - 1$. Reinserting this into the utility in (8) and dividing by P^{M_1} yields

$$\mu(\boldsymbol{u}_2) = \frac{[\boldsymbol{u}_2^{\mathrm{H}} (\bar{\boldsymbol{H}}_2^{\mathrm{H}} \bar{\boldsymbol{H}}_2 - \frac{c_2}{P} \mathbf{I}) \boldsymbol{u}_2]^{M_1}}{(\boldsymbol{u}_2^{\mathrm{H}} \bar{\boldsymbol{H}}_2^{\mathrm{H}} \bar{\boldsymbol{H}}_2 \boldsymbol{u}_2)^{M_1 - 1} \boldsymbol{u}_2^{\mathrm{H}} (\bar{\boldsymbol{H}}_2^{\mathrm{H}} \bar{\boldsymbol{H}}_2 + c_2 \boldsymbol{H}_2^{\mathrm{H}} \boldsymbol{H}_2) \boldsymbol{u}_2}$$

which is independent of the norm of $\boldsymbol{u}_2$, so the optimum $\boldsymbol{u}_2$ that maximizes $\mu(\boldsymbol{u}_2)$ can afterwards be scaled such that $\|\boldsymbol{u}_2\|_2 = 1$. The rate of user one then asymptotically reads as [cf. (6)]

$$R_1' = M_1 \log_2 \frac{P}{M_1} + \log_2 \left|\boldsymbol{H}_1^{\mathrm{H}} \boldsymbol{H}_1\right| + R_2' + \log_2 \mu(\boldsymbol{u}_2). \tag{10}$$

If user one has only a single antenna, i.e., $M_1 = 1$, then the optimum unit norm beamformer $\check{\boldsymbol{u}}_2$ maximizing $\mu(\boldsymbol{u}_2)$ is the principal eigenvector of the matrix

$$\left[\bar{\boldsymbol{H}}_2^{\mathrm{H}} \bar{\boldsymbol{H}}_2 + c_2 \boldsymbol{H}_2^{\mathrm{H}} \boldsymbol{H}_2\right]^{-1} \left(\bar{\boldsymbol{H}}_2^{\mathrm{H}} \bar{\boldsymbol{H}}_2 - \frac{c_2}{P} \mathbf{I}\right), \tag{11}$$

and the resulting power allocation for user two follows from (9) with $\boldsymbol{u}_2$ replaced by its optimum $\check{\boldsymbol{u}}_2$. When user one is equipped with more than one antenna, the utility $\mu(\boldsymbol{u}_2)$ can be maximized by computing a sequence of principal eigenvectors, which will be explained in the following. With appropriate matrix substitutes, $\mu(\boldsymbol{u}_2)$ can be expressed via

$$\mu(\boldsymbol{u}_2) = \frac{(\boldsymbol{u}_2^{\mathrm{H}} \boldsymbol{A} \boldsymbol{u}_2)^{M_1}}{(\boldsymbol{u}_2^{\mathrm{H}} \boldsymbol{B} \boldsymbol{u}_2)^{M_1 - 1} \boldsymbol{u}_2^{\mathrm{H}} \boldsymbol{C} \boldsymbol{u}_2}. \tag{12}$$

Maximizing $\mu(\boldsymbol{u}_2)$ with respect to $\boldsymbol{u}_2$ leads to the Karush-Kuhn-Tucker condition for the optimum unit-norm beamformer $\check{\boldsymbol{u}}_2$, which reads as

$$\left[M_1 \beta^{M_1 - 1} \boldsymbol{C}^{-1} \boldsymbol{A} - (M_1 - 1) \beta^{M_1} \boldsymbol{C}^{-1} \boldsymbol{B}\right] \check{\boldsymbol{u}}_2 = \mu(\check{\boldsymbol{u}}_2) \check{\boldsymbol{u}}_2, \tag{13}$$

where the substitute β is defined via

$$\beta := \frac{\check{\boldsymbol{u}}_2^{\mathrm{H}} \boldsymbol{A} \check{\boldsymbol{u}}_2}{\check{\boldsymbol{u}}_2^{\mathrm{H}} \boldsymbol{B} \check{\boldsymbol{u}}_2}.$$

Due to the dependency of β on $\check{\boldsymbol{u}}_2$, the scalar weights $\beta^{M_1 - 1}$ and β^{M_1} are not known in advance, and the optimum beamformer $\check{\boldsymbol{u}}_2$ thus cannot be found in closed form. However, an iterative principal eigenvector computation leads to the desired beamformer. In iteration n, the matrix on the left hand side of (13) is evaluated for $\boldsymbol{u}_2 = \boldsymbol{u}_2^{(n)}$ instead of $\boldsymbol{u}_2 = \check{\boldsymbol{u}}_2$, and the corresponding principal eigenvector of this matrix is then chosen as the beamformer $\boldsymbol{u}_2^{(n+1)}$ for the next iteration:

$$\boldsymbol{K}^{(n)}\boldsymbol{u}_2^{(n+1)} = \lambda^{(n+1)}\boldsymbol{u}_2^{(n+1)}.$$

In above equation, $\boldsymbol{K}^{(n)}$ is defined via

$$\boldsymbol{K}^{(n)} = M_1\big(\beta^{(n)}\big)^{M_1-1}\boldsymbol{C}^{-1}\boldsymbol{A} - (M_1-1)\big(\beta^{(n)}\big)^{M_1}\boldsymbol{C}^{-1}\boldsymbol{B},$$

with the substitute

$$\beta^{(n)} := \frac{\boldsymbol{u}_2^{(n)\,\mathrm{H}}\boldsymbol{A}\boldsymbol{u}_2^{(n)}}{\boldsymbol{u}_2^{(n)\,\mathrm{H}}\boldsymbol{B}\boldsymbol{u}_2^{(n)}}.$$

In the limit $n \to \infty$, the maximum eigenvalue $\lambda^{(n+1)}$ converges to the objective $\mu(\check{\boldsymbol{u}}_2)$ and $\boldsymbol{u}_2^{(n+1)}$ converges to the beamformer $\check{\boldsymbol{u}}_2$. The monotonic convergence of the utility $\mu(\boldsymbol{u}_2)$ can be shown in a similar fashion as it has been done in [14].

We will now present a functional relationship between the asymptotic rates R_1' and R_2' for small R_2'. Close to the axis $R_2 = 0$, the target rate $R'_{2,\text{target}}$ for user two is very small, and consequently, c_2 is close to zero as well. If c_2 is furthermore divided by a large P, we can safely approximate $\bar{\boldsymbol{H}}_2^{\mathrm{H}}\bar{\boldsymbol{H}}_2 - c_2/P\mathbf{I}$ by $\bar{\boldsymbol{H}}_2^{\mathrm{H}}\bar{\boldsymbol{H}}_2$, and the maximizer of $\mu(\boldsymbol{u}_2)$ then corresponds to the principal eigenvector $\boldsymbol{z}$ of the matrix

$$\big[\boldsymbol{H}_2^{\mathrm{H}}\boldsymbol{H}_2\big]^{-1}\bar{\boldsymbol{H}}_2^{\mathrm{H}}\bar{\boldsymbol{H}}_2 \tag{14}$$

whose eigenvalues are upper bounded by one, since $\bar{\boldsymbol{H}}_2$ is a projected version of $\boldsymbol{H}_2$ and $\max\{0, M_2 - M_1\}$ eigenvalues are equal to one with probability one. Note that the choice $\boldsymbol{u}_2 = \boldsymbol{z}$ maximizes the utility $\mu(\boldsymbol{u}_2)$ for all target rates for which the approximation that neglects $c_2/P\mathbf{I}$ is accurate. Let λ denote the maximum eigenvalue of the matrix in (14). Close to the axis $R_2 = 0$, the utility $\mu(\boldsymbol{z})$ follows the strong asymptotic equivalence

$$\log_2 \mu(\boldsymbol{z}) \cong \log_2 \frac{1}{1+\frac{c_2}{\lambda}} = -R'_{2,\text{target}} - \log_2\Big(\frac{1}{\lambda} + \frac{1-\frac{1}{\lambda}}{2^{R'_{2,\text{target}}}}\Big),$$

and the functional relationship between R_1' and R_2' can be expressed for $\boldsymbol{u}_2 = \boldsymbol{z}$ and $R_2' = R'_{2,\text{target}}$ as [see (10)]

$$R_1' \cong M_1 \log_2 \frac{P}{M_1} + \log_2\big|\boldsymbol{H}_1^{\mathrm{H}}\boldsymbol{H}_1\big| - \log_2\Big(\frac{1}{\lambda} + \frac{1-\frac{1}{\lambda}}{2^{R'_{2,\text{target}}}}\Big).$$

Deriving the right hand side of above equation with respect to $R'_{2,\text{target}}$ and evaluating the result at $R'_{2,\text{target}} = 0$ yields $1 - \frac{1}{\lambda}$. Therefore, the border of the rate region hits the axis $R_2 = 0$ with the angle

$$\varphi = \arctan\Big(\frac{\lambda}{1-\lambda}\Big). \tag{15}$$

For $\lambda = 1$, which happens if $M_2 > M_1$ as said before, the border of the asymptotic rate region is perpendicular to the axis $R_2 = 0$.

IV. Asymptotic Weighted Sum Rate Maximization

In [11], the following theorem is proven:

Theorem IV.1: *For any strictly positive constant weight vector $\boldsymbol{w} > \mathbf{0}$, the weighted sum rate is asymptotically maximized when every user multiplexes as many data streams as he has antennas given the base station has at least as many antennas as the users have in sum. Furthermore, a linear power allocation and scaled identities are the optimum transmission strategy.*

Intuitively, this can be explained as follows. Since every active stream contributes to the weighted sum rate via $w_i \log_2 P$ for a linear power allocation if the stream belongs to user i [7], there is a threshold power P_{Th} above which the weighted sum rate is larger when all data streams are active compared to the case when some users do not apply full multiplexing. Of course, this requires the base station to have at least as many antennas as the user terminals have in sum.

Only a small fraction of the asymptotic rate region boundary constitutes the maximizers of the asymptotic weighted sum rate, and as mentioned before, these rate tuples are obtained by the full multiplexing stream configuration with a linear power allocation. A consequence of the linear power allocation is that any weighted sum rate maximizing transmission strategy for a positive weight vector $\boldsymbol{w} > \mathbf{0}$ achieves a multiplexing gain of $\sum_{k=1}^{K} M_k$. The graphical interpretation is that the rate region whose axes are divided by $\log_2 P$ is weakly asymptotically equivalent to a hyper-cuboid with edge length M_k for axis k. However, strong asymptotic equivalence does not hold there, since first, the 'edge' with maximum individual rates is rounded off even for infinite transmit power, and second, close to the axis $R_2 = 0$ in the two user case, user one can achieve a larger rate when user two applies single stream beamforming instead of full multiplexing, see Fig. 1. Summing up, all weighted sum rate maximizers asymptotically achieve rate tuples that lie on the rounded off corner and achieve the maximum multiplexing gain.

V. Graphical Visualization

To illustrate the geometry of the border of the rate region in the asymptotic case, we choose a scenario where a base station with $N = 5$ antennas receives data in the dual MAC from two users with two antennas each. In Fig. 1, the sum power is set to $10\log_{10} P = 40$dB. The dashed curve corresponds to the exact rate pairs that are achieved for transmit covariance matrices $\boldsymbol{Q}_1 = p_1/M_1 \cdot \mathbf{I}_{M_1}$ and $\boldsymbol{Q}_2 = (P - p_1)/M_2 \cdot \mathbf{I}_{M_2}$ according to (1) when p_1 is swept from zero to P. Obviously, the stream allocation for this curve is $[2/2]$, so both users have two active streams. The high power approximation of the full multiplexing case from (4) shows the rate pairs R_1'' and R_2'' as the solid black curve which perfectly matches the dashed one if both rates are larger than approximately 14 bits per channel use. The deviation for smaller rates does not matter because there, the border of the asymptotic rate region is given by a stream allocation where only one user applies full multiplexing, the other one has only one active stream. In other words, the blue curve which corresponds to the stream allocation [1/2] is weakly Pareto optimal as long as the rate of user one is smaller than about 17 bits per channel use. Conversely, if user two has only one active stream denoted by [2/1], the red curve is weakly Pareto optimal if the rate of user two is below about 15 bits per channel use. Otherwise, full multiplexing all four data streams is optimum. Since both users have the same number of antennas, the largest eigenvalue of (14) is smaller than one and the angle φ from (15) is smaller than 90 degrees if the two channels are not orthogonal, although at the given zoom level, this does not seem to be the case. The exact angle under which the red curve hits the axis $R_2 = 0$ is 82 degrees for the given channels, which is also the angle under which the axis $R_1 = 0$ is hit, since the eigenvalues of (14) do not change when changing the index from 2 to 1 in (14) if $M_1 = M_2$.

The case in which the base station does not have enough degrees of freedom to support the overall full multiplexing configuration is shown in Fig. 2. There, only the two stream configurations [1/2] and [2/1] contribute to the boundary of the asymptotic rate region. Again, $10 \log_{10} P = 40$dB holds but the base station is equipped with $N = 3$ antennas only, so the last two rows of $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ have been removed. The lack of the full multiplexing configuration leads to the dent in the region boundary, and φ has decreased to 72 degrees.

VI. Conclusion

We have derived the boundary of the asymptotic rate region under linear filtering in the MIMO broadcast channel. For the stream configurations that contribute to the border of the region, we have found the optimum transmission strategy in terms of transmit covariance matrices and precoders. Besides the case when the base station has enough degrees of freedom, we investigated the scenario where not enough antennas are available at the base station, which so far has never been analyzed before.

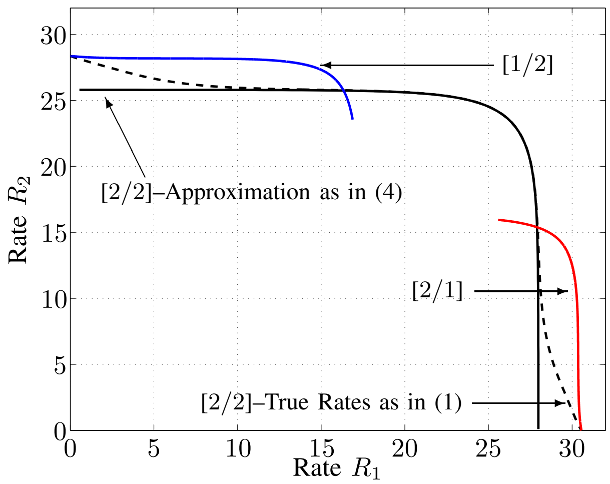

Fig. 1. Border of the rate region with $N = 5$ and $M_1 = M_2 = 2$ for which the relevant three different stream allocations [1/2], [2/1], and [2/2] contribute.

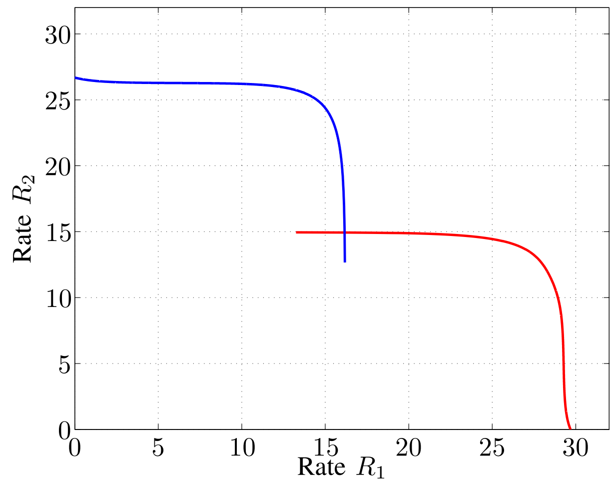

Fig. 2. Border of the rate region with $N = 3$ and $M_1 = M_2 = 2$ for which the relevant two different stream allocations [1/2] and [2/1] contribute.

References

[1] S. Vishwanath, N. Jindal, and A. Goldsmith, "On the Capacity of Multiple Input Multiple Output Broadcast Channels," in *International Conference on Communications (ICC 2002)*, vol. 3, 2002, pp. 1444–1450.

[2] ——, "Duality, Achievable Rates, and Sum-Rate Capacity of Gaussian MIMO Broadcast Channels," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2658–2668, October 2003.

[3] R. Böhnke and K. Kammeyer, "Weighted Sum Rate Maximization for the MIMO-Downlink Using a Projected Conjugate Gradient Algorithm," in *First International Workshop on Cross Layer Design (IWCLD 07)*, Jinan, China, September 2007.

[4] J. Liu and Y. T. Hou, "Maximum Weighted Sum Rate of Multi-Antenna Broadcast Channels," in *International Conference on Communications (ICC 2008)*, Beijing, China, May 2008.

[5] M. Kobayashi and G. Caire, "An Iterative Water-Filling Algorithm for Maximum Weighted Sum-Rate of Gaussian MIMO-BC," *IEEE J. Select. Areas Commun.*, vol. 24, no. 8, pp. 1640–1646, August 2006.

[6] R. Hunger, D. A. Schmidt, M. Joham, and W. Utschick, "A General Covariance-Based Optimization Framework Using Orthogonal Projections," in *International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Recife, Brazil, July 2008.

[7] R. Hunger, D. A. Schmidt, and M. Joham, "A Combinatorial Approach to Maximizing the Sum Rate in the MIMO BC with Linear Precoding," in *42nd Asilomar Conference on Signals, Systems, and Computers (Asilomar 2008)*, October 2008.

[8] J. Lee and N. Jindal, "Dirty Paper Coding vs. Linear Precoding for MIMO Broadcast Channels," in *40th Asilomar Conference on Signals, Systems, and Computers (Asilomar 2006)*, October 2006, pp. 779–783.

[9] ——, "High SNR Analysis for MIMO Broadcast Channels: Dirty Paper Coding Versus Linear Precoding," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4787–4792, December 2007.

[10] R. Hunger and M. Joham, "On the Asymptotic Optimality of Block-Diagonalization for the MIMO BC under Linear Filtering," in *Workshop on Smart Antennas (WSA 09)*, Berlin, February 2009.

[11] ——, "An Asymptotic Analysis of the MIMO Broadcast Channel under Linear Filtering," in *Conference on Information Sciences and Systems (CISS 2009)*, March 2009.

[12] ——, "A General Rate Duality of the MIMO Multiple Access Channel and the MIMO Broadcast Channel," in *Proc. IEEE Globecom 2008*, New Orleans, November 2008.

[13] E. Telatar, "Capacity of multi-antenna gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–596, November/December 1999.

[14] R. Hunger, P. de Kerret, and M. Joham, "An Algorithm for Maximizing a Quotient of Two Hermitian Form Determinants with Different Exponents," *Submitted to ICASSP 2010.*