Iterative Algorithms for Transceiver Design in MIMO Broadcast Channels with Improper Signaling

Christoph Hellings and Wolfgang Utschick

10th International ITG Conference on Systems, Communications and Coding (SCC 2015)

Hamburg, Germany, 2nd - 5th February, 2015

Technische Universität München
Fachgebiet Methoden der Signalverarbeitung
http://www.msv.ei.tum.de

Iterative Algorithms for Transceiver Design in MIMO Broadcast Channels with Improper Signaling

Christoph Hellings and Wolfgang Utschick
Associate Institute for Signal Processing, Technische Universität München, 80290 Munich, Germany
Telephone: +49 89 289-28516, Fax: +49 89 289-28504, Email: {hellings,utschick}@tum.de

***Abstract*—Recent research has revealed that using improper (noncircular) transmit signals can be beneficial in multiple-input multiple-output broadcast channels if highly complex nonlinear interference cancellation schemes such as dirty paper coding are avoided. However, finding the optimal improper transmit strategy in such a scenario is a nonconvex optimization problem for which no globally optimal solution is known in general. In this paper, we study whether suboptimal solution approaches that are known from linear transceiver design with proper (circularly symmetric) transmit signals can be transferred to the case with improper signals via a composite real representation. It turns out that such optimization methods may get stuck in a proper solution even if an improper strategy would be better, but an adequate initialization can prevent this behavior. The theoretical analysis is based on the recently proposed framework of block-skew-circulant matrices in complex-valued signal processing.**

***Index Terms*—Block-skew-circulant matrices, broadcast channels, composite real representation, improper signals, multiuser MIMO systems, noncircular, widely linear transceivers.**

I. Introduction

The capacity-achieving transmit strategy in Gaussian multiple-input multiple-output (MIMO) broadcast channels is based on an interference precancellation scheme called dirty paper coding (DPC) [1], [2] which is prohibitively complex for practical implementation [3]. Therefore, researchers have also studied so-called linear transceivers where nonlinear operations such as encoding and decoding are applied only on a per-stream basis while all operations involving multiple data streams have to be linear (cf., e.g., [4], [5]). This notion can be extended to *widely* linear transceivers (e.g., [5], [6]), where the operations are linear functions of complex signals and their conjugates (see [7], [8]), or, equivalently, linear functions of the real and imaginary parts [8]. The system model with widely linear transceivers is introduced in Section III.

When using DPC, proper [9] (circularly symmetric) Gaussian transmit signals are the optimal input distribution for the MIMO broadcast channel [1]. However, it was recently shown that proper Gaussian per-user transmit signals are no longer optimal in MIMO broadcast channels if a restriction to (widely) linear transceivers is imposed [5], [10], [11]. Performance gains by using improper signals were observed both with time-sharing [10] and without time-sharing [5], [11]. Moreover, it was shown that improper signaling can enlarge the quality of service (QoS) feasibility region, i.e., data rates that cannot be achieved with proper per-user transmit signals even when spending arbitrarily high transmit power might be achievable with improper signals [5]. To benefit from these potential gains, we need appropriate optimization algorithms to find good improper transmit strategies.

The authors of [11] used the characterization of improper signals by means of their covariance matrices and their pseudocovariance matrices (cf., e.g., [8]), and they developed a specialized algorithm for multiple-input single-output (MISO) broadcast channels. Unfortunately, this approach cannot be easily extended to MIMO systems, which use multiple antennas at the transmitter and at the receivers.

In the numerical simulations in [5], it was instead exploited that improper signals can also be characterized by the covariance matrix of their composite real representation where real and imaginary parts of the signals are stacked onto each other in vectors (cf. Section II and, e.g., [8]). The important advantage of this composite real formulation is that optimization algorithms that have originally been developed for proper complex signals can be reused. Such algorithms can be transferred to real-valued scenarios, and via the composite real representation, they can then be applied to optimize improper transmit signals. That way, we can avoid having to develop completely new algorithms, which is necessary when using the complex representation of widely linear transceivers.

Composite real representations have also been used to study single-user MIMO systems [12], systems with space-time coding [13], broadcast channels with dirty paper coding [2], [14], interference channels [15], [16], and relaying scenarios [17], [18]. However, there are only few publications on improper signaling in MIMO broadcast channels with widely linear transceivers [5], [10], [11], and to the best of our knowledge, a systematic study of transceiver optimization via the composite real representation has not yet been performed.

In the main part of this paper in Section IV, we consider various iterative optimization algorithms known from the literature on MIMO broadcast channels with proper signals, and we study whether they are suitable for optimizing strategies involving improper signals. It turns out that these heuristic methods have to be applied and/or adapted carefully since they tend to produce solutions with proper signals even in cases where improper signals would be beneficial.

The theoretical analysis in this paper is based on the recently proposed framework of block-skew-circulant matrices in complex-valued signal processing from [19], which is briefly summarized in Section II. The main idea is to decompose composite real covariance matrices as well as composite real representations of widely linear filters into block-skew-circulant ($\mathcal{BSC}$) and block-Hankel-skew-circulant ($\mathcal{BHSC}$) matrices. For our analysis, we also derive some new properties of $\mathcal{BSC}$ matrices which are not included in [19].

We conclude the paper with exemplary simulation results (Section V), and with a comparison to another situation where algorithms for MIMO transceiver design have been applied to a new setting (Section VI), namely to the equivalent single-carrier representation of multicarrier broadcast channels [20].

Notation: We write $\mathbf{0}$ for the zero matrix or vector, $\mathbf{I}_L$ for the identity matrix of size L, $\bullet^{\mathrm{T}}$ for the transpose, and $\bullet^{\mathrm{H}}$ for the conjugate transpose. The vector $\boldsymbol{e}_i$ is the ith canonical unit vector of appropriate dimension. To easily distinguish real quantities from complex quantities, we use a tilde $\utilde{\bullet}$ below complex quantities. We use $\Re$, $\Im$, and $\bullet^*$ for real part, imaginary part, and complex conjugate, respectively. The shorthand notation $\breve{\boldsymbol{x}}$ is used for a vector $[\Re(\utilde{\boldsymbol{x}})^{\mathrm{T}}, \Im(\utilde{\boldsymbol{x}})^{\mathrm{T}}]^{\mathrm{T}}$.

II. Block-Skew-Circulant Matrices

In [19], it was proposed to study systems involving improper signals by means of real-valued block-skew-circulant ($\mathcal{BSC}$) and block-Hankel-skew-circulant ($\mathcal{BHSC}$) matrices. For the special case of 2×2 blocks, which is the relevant one for this application, the $\mathcal{BSC}_2$ and $\mathcal{BHSC}_2$ structures are given by

$$\grave{\boldsymbol{A}} = \begin{bmatrix} \boldsymbol{A}_1 & -\boldsymbol{A}_2 \\ \boldsymbol{A}_2 & \boldsymbol{A}_1 \end{bmatrix} \quad \text{and} \quad \acute{\boldsymbol{B}} = \begin{bmatrix} \boldsymbol{B}_1 & \boldsymbol{B}_2 \\ \boldsymbol{B}_2 & -\boldsymbol{B}_1 \end{bmatrix} \tag{1}$$

respectively [19], where the subscripts in $\mathcal{BSC}_2$ and $\mathcal{BHSC}_2$ refer to the number of blocks. Note that we use $\grave{\bullet}$ and $\acute{\bullet}$ to indicate the block-Toeplitz and block-Hankel structures of $\mathcal{BSC}$ and $\mathcal{BHSC}_2$ matrices, respectively. Where necessary, we indicate the block size as a superscript of the sets $\mathcal{BSC}_2^{K\times L} \subset \mathbb{R}^{2K\times 2L}$ and $\mathcal{BHSC}_2^{K\times L} \subset \mathbb{R}^{2K\times 2L}$.

The real-valued equivalent of a complex matrix given by

$$\utilde{\boldsymbol{A}} \in \mathbb{C}^{K\times L} \quad \leftrightarrow \quad \grave{\boldsymbol{A}} = \begin{bmatrix} \Re(\utilde{\boldsymbol{A}}) & -\Im(\utilde{\boldsymbol{A}}) \\ \Im(\utilde{\boldsymbol{A}}) & \Re(\utilde{\boldsymbol{A}}) \end{bmatrix} \in \mathcal{BSC}_2^{K\times L} \tag{2}$$

is a $\mathcal{BSC}_2$ matrix. For the real-valued equivalents $\grave{\boldsymbol{A}}$, $\grave{\boldsymbol{C}}$, and $\grave{\boldsymbol{D}}$ of complex matrices $\utilde{\boldsymbol{A}}$, $\utilde{\boldsymbol{C}}$, and $\utilde{\boldsymbol{D}}$, we have that $\grave{\boldsymbol{A}} = \grave{\boldsymbol{C}}\grave{\boldsymbol{D}} \Leftrightarrow \utilde{\boldsymbol{A}} = \utilde{\boldsymbol{C}}\utilde{\boldsymbol{D}}$, $\grave{\boldsymbol{A}} = \grave{\boldsymbol{C}} + \grave{\boldsymbol{D}} \Leftrightarrow \utilde{\boldsymbol{A}} = \utilde{\boldsymbol{C}} + \utilde{\boldsymbol{D}}$, and $\breve{\boldsymbol{y}} = \grave{\boldsymbol{A}}\breve{\boldsymbol{x}} \Leftrightarrow \utilde{\boldsymbol{y}} = \utilde{\boldsymbol{A}}\utilde{\boldsymbol{x}}$ [21, Lemma 1], where $\breve{\boldsymbol{x}} = [\Re(\utilde{\boldsymbol{x}})^{\mathrm{T}}, \Im(\utilde{\boldsymbol{x}})^{\mathrm{T}}]^{\mathrm{T}}$.

It can be shown that the subspace $\mathcal{BSC}_2^{K\times L}$ is the orthogonal complement of $\mathcal{BHSC}_2^{K\times L}$ in $\mathbb{R}^{2K\times 2L}$ [19, Lemma 5]. Using projections onto these linear spaces (see [19, Lemma 6]), any matrix $\boldsymbol{C}$ can be uniquely decomposed into a $\mathcal{BSC}_2$ and a $\mathcal{BHSC}_2$ component, i.e., $\boldsymbol{C} = \grave{\boldsymbol{C}} + \acute{\boldsymbol{C}}$. Now consider a complex widely linear mapping

$$\utilde{\boldsymbol{f}}(\utilde{\boldsymbol{x}}) = \utilde{\boldsymbol{A}}_{\mathrm{L}}\utilde{\boldsymbol{x}} + \utilde{\boldsymbol{A}}_{\mathrm{CL}}\utilde{\boldsymbol{x}}^* \tag{3}$$

where the subscripts of the constant factors $\utilde{\boldsymbol{A}}_{\mathrm{L}}$ and $\utilde{\boldsymbol{A}}_{\mathrm{CL}}$ stand for linear and conjugate linear, respectively. We can rewrite this in the composite real representation as

$$\breve{\boldsymbol{f}}_{\mathrm{WL}}(\breve{\boldsymbol{x}}) = \underbrace{(\grave{\boldsymbol{A}}_{\mathrm{L}} + \acute{\boldsymbol{A}}_{\mathrm{CL}})}_{\boldsymbol{A}}\breve{\boldsymbol{x}} = \grave{\boldsymbol{A}}_{\mathrm{L}}\breve{\boldsymbol{x}} + \grave{\boldsymbol{A}}_{\mathrm{CL}}\begin{bmatrix} \mathbf{I}_L & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_L \end{bmatrix}\breve{\boldsymbol{x}} \tag{4}$$

where the $\mathcal{BSC}_2$ matrices $\grave{\boldsymbol{A}}_{\mathrm{L}}$ and $\grave{\boldsymbol{A}}_{\mathrm{CL}}$ are the real-valued equivalents (2) of $\utilde{\boldsymbol{A}}_{\mathrm{L}}$ and $\utilde{\boldsymbol{A}}_{\mathrm{CL}}$, respectively [19, Theorem 2]. Only if $\acute{\boldsymbol{A}}_{\mathrm{CL}}$ vanishes, the matrix $\boldsymbol{A}$ corresponds to a linear operator in the complex representation [19, Corollary 1].

Similarly, the covariance matrix of the composite real representation $\breve{\boldsymbol{x}} = [\Re(\utilde{\boldsymbol{x}})^{\mathrm{T}}, \Im(\utilde{\boldsymbol{x}})^{\mathrm{T}}]^{\mathrm{T}}$ given by

$$\boldsymbol{C}_{\breve{\boldsymbol{x}}} = \mathrm{E}[\breve{\boldsymbol{x}}\breve{\boldsymbol{x}}^{\mathrm{T}}] - \mathrm{E}[\breve{\boldsymbol{x}}]\,\mathrm{E}[\breve{\boldsymbol{x}}^{\mathrm{T}}] = \begin{bmatrix} \boldsymbol{C}_{\Re\utilde{\boldsymbol{x}}} & \boldsymbol{C}_{\Re\utilde{\boldsymbol{x}}\Im\utilde{\boldsymbol{x}}} \\ \boldsymbol{C}_{\Re\utilde{\boldsymbol{x}}\Im\utilde{\boldsymbol{x}}}^{\mathrm{T}} & \boldsymbol{C}_{\Im\utilde{\boldsymbol{x}}} \end{bmatrix} \tag{5}$$

can be uniquely decomposed [19, Lemma 9] into

$$\boldsymbol{C}_{\breve{\boldsymbol{x}}} = \underbrace{\frac{1}{2}\begin{bmatrix} \Re(\utilde{\boldsymbol{C}}_{\utilde{\boldsymbol{x}}}) & -\Im(\utilde{\boldsymbol{C}}_{\utilde{\boldsymbol{x}}}) \\ \Im(\utilde{\boldsymbol{C}}_{\utilde{\boldsymbol{x}}}) & \Re(\utilde{\boldsymbol{C}}_{\utilde{\boldsymbol{x}}}) \end{bmatrix}}_{\grave{\boldsymbol{P}}_{\utilde{\boldsymbol{x}}}} + \underbrace{\frac{1}{2}\begin{bmatrix} \Re(\tilde{\utilde{\boldsymbol{C}}}_{\utilde{\boldsymbol{x}}}) & \Im(\tilde{\utilde{\boldsymbol{C}}}_{\utilde{\boldsymbol{x}}}) \\ \Im(\tilde{\utilde{\boldsymbol{C}}}_{\utilde{\boldsymbol{x}}}) & -\Re(\tilde{\utilde{\boldsymbol{C}}}_{\utilde{\boldsymbol{x}}}) \end{bmatrix}}_{\acute{\boldsymbol{N}}_{\utilde{\boldsymbol{x}}}} \tag{6}$$

(see [19, Theorem 3]), where $\utilde{\boldsymbol{C}}_{\utilde{\boldsymbol{x}}} = \mathrm{E}[\utilde{\boldsymbol{x}}\utilde{\boldsymbol{x}}^{\mathrm{H}}] - \mathrm{E}[\utilde{\boldsymbol{x}}]\,\mathrm{E}[\utilde{\boldsymbol{x}}^{\mathrm{H}}]$ is the complex covariance matrix, which determines the power shaping of $\utilde{\boldsymbol{x}}$, and $\tilde{\utilde{\boldsymbol{C}}}_{\utilde{\boldsymbol{x}}} = \mathrm{E}[\utilde{\boldsymbol{x}}\utilde{\boldsymbol{x}}^{\mathrm{T}}] - \mathrm{E}[\utilde{\boldsymbol{x}}]\,\mathrm{E}[\utilde{\boldsymbol{x}}^{\mathrm{T}}]$ is the pseudocovariance matrix, which determines the impropriety (noncircularity) of $\utilde{\boldsymbol{x}}$. In particular, $\utilde{\boldsymbol{x}}$ is proper only if $\acute{\boldsymbol{N}}_{\utilde{\boldsymbol{x}}} = \mathbf{0}$ [19, Corollary 2], i.e., if $\boldsymbol{C}_{\breve{\boldsymbol{x}}}$ is a $\mathcal{BSC}_2$ matrix.

The $\mathcal{BSC}_2$ structure is preserved by linear combinations, transposition, matrix products, and matrix inversion if all involved matrices are $\mathcal{BSC}_2$ matrices [19]. Moreover, the eigenvalue decomposition of a symmetric $\mathcal{BSC}_2$ matrix can be written as $\grave{\boldsymbol{A}} = \grave{\boldsymbol{Q}}\grave{\boldsymbol{\Lambda}}\grave{\boldsymbol{Q}}^{\mathrm{T}}$ where $\grave{\boldsymbol{Q}}$ is an orthogonal $\mathcal{BSC}_2$ matrix and $\grave{\boldsymbol{\Lambda}}$ is a diagonal $\mathcal{BSC}_2$ matrix [19, Lemma 11]. Such a decomposition was called standard eigenvalue decomposition in [19]. Similarly, a standard singular value decomposition $\grave{\boldsymbol{A}} = \grave{\boldsymbol{U}}\grave{\boldsymbol{\Sigma}}\grave{\boldsymbol{V}}^{\mathrm{T}}$ of general $\mathcal{BSC}_2$ matrices, which consists only of $\mathcal{BSC}_2$ factors, was derived in [19, Lemma 15].

III. System Model and Dual Uplink

We consider a MIMO broadcast channel with M transmit antennas, K users, and N_k receive antennas for user k, where the channel of user k is described by a channel matrix $\utilde{\boldsymbol{H}}_k \in \mathbb{C}^{N_k\times M}$ and an additive proper Gaussian noise component $\utilde{\boldsymbol{n}}_k \sim \mathcal{CN}(\mathbf{0}, \utilde{\boldsymbol{C}}_{\utilde{\boldsymbol{n}}_k})$. Transmission and estimation with widely linear transceivers can be described by

$$\utilde{\boldsymbol{y}}_k = \utilde{\boldsymbol{H}}_k \sum_{\ell=1}^{K} (\utilde{\boldsymbol{B}}_{\mathrm{L},\ell}\utilde{\boldsymbol{x}}_\ell + \utilde{\boldsymbol{B}}_{\mathrm{CL},\ell}\utilde{\boldsymbol{x}}_\ell^*) + \utilde{\boldsymbol{n}}_k \tag{7}$$

$$\hat{\utilde{\boldsymbol{x}}}_k = \utilde{\boldsymbol{G}}_{\mathrm{L},k}\utilde{\boldsymbol{y}}_k + \utilde{\boldsymbol{G}}_{\mathrm{CL},k}\utilde{\boldsymbol{y}}_k^* \tag{8}$$

where $S_k = \min\{N_k, M\}$ streams of i.i.d. proper Gaussian data symbols are intended for user k, i.e., $\utilde{\boldsymbol{x}}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{S_k})$. The matrices $\utilde{\boldsymbol{B}}_{\mathrm{L},\ell}, \utilde{\boldsymbol{B}}_{\mathrm{CL},\ell} \in \mathbb{C}^{M\times S_\ell}$ and $\utilde{\boldsymbol{G}}_{\mathrm{L},k}, \utilde{\boldsymbol{G}}_{\mathrm{CL},k} \in \mathbb{C}^{S_k\times N_k}$ describe the widely linear transmit filters (beamforming matrices) and receive filters (equalizers), respectively. If $\utilde{\boldsymbol{B}}_{\mathrm{CL},\ell} = \mathbf{0}$, the per-user transmit signal for user ℓ in (9) is a linear function of $\utilde{\boldsymbol{x}}_\ell$ and thus proper (cf. [8]). Otherwise, it is a widely linear function of $\utilde{\boldsymbol{x}}_\ell$ and can become improper.[1]

[1] If $\utilde{\boldsymbol{x}}_\ell$ is improper, $\utilde{\boldsymbol{B}}_{\mathrm{CL},\ell} \neq \mathbf{0}$ is not needed for improper signaling, but we assume proper $\utilde{\boldsymbol{x}}_\ell$ without loss of generality (only the distributions of the per-user transmit signals matter, not the distribution of $\utilde{\boldsymbol{x}}_\ell$).

A MIMO broadcast channel with widely linear transceivers can be described using the composite real representation

$$\check{\boldsymbol{y}}_k = \grave{\boldsymbol{H}}_k \sum_{\ell=1}^{K} \boldsymbol{B}_\ell \check{\boldsymbol{x}}_\ell + \check{\boldsymbol{n}}_k \tag{9}$$

$$\hat{\check{\boldsymbol{x}}}_k = \boldsymbol{G}_k \check{\boldsymbol{y}}_k \tag{10}$$

where $\boldsymbol{B}_\ell$ and $\boldsymbol{G}_k$ are defined just like the matrix $\boldsymbol{A}$ in (4). We assume $\boldsymbol{C}_{\check{\boldsymbol{n}}_k} = \mathbf{I}_{2N_k}$ without loss of generality (the channel matrices describe the effective channels after noise whitening). With information lossless receive filters, the rate of user k is

$$r_k = \frac{1}{2}\log_2 \det\left(\mathbf{I}_{2N_k} + \boldsymbol{Z}_{\mathrm{DL},k}^{-1} \grave{\boldsymbol{H}}_k \boldsymbol{B}_k \boldsymbol{B}_k^{\mathrm{T}} \grave{\boldsymbol{H}}_k^{\mathrm{T}}\right) \tag{11}$$

$$\boldsymbol{Z}_{\mathrm{DL},k} = \mathbf{I}_{2N_k} + \sum_{\ell \neq k} \grave{\boldsymbol{H}}_k \boldsymbol{B}_\ell \boldsymbol{B}_\ell^{\mathrm{T}} \grave{\boldsymbol{H}}_k^{\mathrm{T}}. \tag{12}$$

According to [10, Lemma 1], we can use a composite real representation to extend the uplink-downlink duality with linear transceivers from [22] to widely linear transceivers. With the uplink channel matrices $\grave{\boldsymbol{H}}_k^{\mathrm{T}}$, the uplink noise covariance matrix $\mathbf{I}_{2M}$, and the uplink transmit filters $\boldsymbol{T}_k$, we obtain the uplink rates (with information lossless receive filters)

$$R_k = \frac{1}{2}\log_2 \det\left(\mathbf{I}_{2M} + \boldsymbol{Z}_k^{-1} \grave{\boldsymbol{H}}_k^{\mathrm{T}} \boldsymbol{T}_k \boldsymbol{T}_k^{\mathrm{T}} \grave{\boldsymbol{H}}_k\right) \tag{13}$$

$$\boldsymbol{Z}_k = \mathbf{I}_{2M} + \sum_{\ell \neq k} \grave{\boldsymbol{H}}_\ell^{\mathrm{T}} \boldsymbol{T}_\ell \boldsymbol{T}_\ell^{\mathrm{T}} \grave{\boldsymbol{H}}_\ell. \tag{14}$$

The same rates $r_k = R_k$ are achievable in the downlink and in the dual uplink if the same sum transmit power is spent.

IV. Analysis of Iterative Algorithms

When applying an optimization algorithm to the composite real representation of a communication system, we have to keep in mind the particular structure of the composite real channel matrices: they are $\mathcal{BSC}_2$ matrices. Therefore, the question arises, how existing optimization algorithms behave when applied to a setting with this channel structure. In particular, we have to ask whether the obtained solutions can indeed correspond to improper signaling or if they correspond to strategies with linear filters and proper signals.

This question was first asked in [19] for a gradient-based method. After briefly sketching this existing result, we extend it to methods based on alternating filter updates, whose analysis is more involved. Finally, we study scaling of individual beamforming vectors, which is often done in combination with other iterative techniques, we analyze typical initializations, and we comment on effects caused by numerical inaccuracies.

A. Gradient-Based Filter Updates

In [23], [24], a gradient-projection algorithm was applied to the uplink transmit filters $\boldsymbol{T}_k$ to maximize the weighted sum rate in a MIMO broadcast channel with linear transceivers. Transferred to the composite real representation, we have

$$\frac{\partial \sum_{\ell=1}^{K} w_\ell R_\ell}{\partial \boldsymbol{T}_k^*} = \boldsymbol{A}_k \boldsymbol{T}_k \tag{15}$$

where the scalars w_ℓ are constant weighting factors, and

$$\boldsymbol{A}_k = \frac{1}{\ln 2} \grave{\boldsymbol{H}}_k \left(\sum_{\ell=1}^{K} w_\ell \boldsymbol{Z}^{-1} - \sum_{\ell \neq k} w_\ell \boldsymbol{Z}_\ell^{-1} \right) \grave{\boldsymbol{H}}_k^{\mathrm{T}} \tag{16}$$

with $\boldsymbol{Z}_k$ from (14) and

$$\boldsymbol{Z} = \mathbf{I}_{2M} + \sum_{k=1}^{K} \grave{\boldsymbol{H}}_k^{\mathrm{T}} \boldsymbol{T}_k \boldsymbol{T}_k^{\mathrm{T}} \grave{\boldsymbol{H}}_k. \tag{17}$$

The gradient-projection step is $\boldsymbol{T}_k \leftarrow (a_k \mathbf{I}_{N_k C} + b_k \boldsymbol{A}_k)\boldsymbol{T}_k$ where a_k and b_k are chosen according to a step size rule and subject to a sum power constraint [23], [24].

Since transposition, addition, multiplication, and matrix inversion preserve the $\mathcal{BSC}_2$ structure (see [19] and Section II), the filter matrices after the update are $\mathcal{BSC}_2$ if all filter matrices $\boldsymbol{T}_k$ have been $\mathcal{BSC}_2$ before the update. Thus, the gradient-projection algorithm converges to a solution that corresponds to proper signaling (i.e., a solution with $\mathcal{BSC}_2$ structure) if it is initialized with a strategy that corresponds to proper signaling (i.e., an initialization with $\mathcal{BSC}_2$ structure) [19]. On the other hand, a solution that corresponds to improper signaling can be obtained by using an initialization that corresponds to improper signaling (i.e., that does not have $\mathcal{BSC}_2$ structure).

B. Alternating Filter Updates

Another popular iterative optimization method consists of alternating updates of the transmit and receive filters in the downlink and the dual uplink [22] as proposed e.g., in [25], [26] for weighted sum rate maximization, in [25] for sum MSE minimization and SINR balancing, in [26] for rate balancing, and in [25]–[27] for power minimization. We study the application of the approach to the composite real representation.

For given uplink beamforming matrices $\boldsymbol{T}_k$, the optimal uplink receive filters in the MMSE sense $\boldsymbol{V}_k^{\mathrm{T}} = \boldsymbol{T}_k^{\mathrm{T}} \grave{\boldsymbol{H}}_k \boldsymbol{Z}^{-1}$ with $\boldsymbol{Z}$ from (17) are computed. Then, the downlink beamforming matrices are chosen according to

$$\boldsymbol{B}_k \leftarrow \boldsymbol{V}_k \boldsymbol{W}_k \operatorname{diag}_{i=1}^{2S_k}\{\alpha_{k,i}\} \tag{18}$$

where the orthogonal matrix $\boldsymbol{W}_k$ contains the eigenvectors of $\boldsymbol{T}_k^{\mathrm{T}} \grave{\boldsymbol{H}}_k^{\mathrm{T}} \boldsymbol{Z}_k^{-1} \grave{\boldsymbol{H}}_k \boldsymbol{T}_k$ with $\boldsymbol{Z}_k$ from (14) (see [22]), and the scalars $\alpha_{k,i} \in \mathbb{R}$ are chosen such that (based on [22])

$$\boldsymbol{M}[\alpha_{1,1}, \ldots, \alpha_{K,2S_k}]^{\mathrm{T}} = [\|\boldsymbol{T}_1 \boldsymbol{W}_1 \boldsymbol{e}_1\|_2^2, \ldots, \|\boldsymbol{T}_K \boldsymbol{W}_K \boldsymbol{e}_{2S_K}\|_2^2]^{\mathrm{T}} \tag{19}$$

with

$$\boldsymbol{M} = \begin{bmatrix} \boldsymbol{M}_{1,1} & \ldots & \boldsymbol{M}_{1,K} \\ \vdots & \ddots & \vdots \\ \boldsymbol{M}_{K,1} & \ldots & \boldsymbol{M}_{K,K} \end{bmatrix} \tag{20}$$

where $\boldsymbol{M}_{k,\ell}$ is defined via the Hadamard product

$$\boldsymbol{M}_{k,\ell} = -(\boldsymbol{W}_k^{\mathrm{T}} \boldsymbol{T}_k^{\mathrm{T}} \grave{\boldsymbol{H}}_k \boldsymbol{V}_\ell \boldsymbol{W}_\ell) \odot (\boldsymbol{W}_k^{\mathrm{T}} \boldsymbol{T}_k^{\mathrm{T}} \grave{\boldsymbol{H}}_k \boldsymbol{V}_\ell \boldsymbol{W}_\ell) \tag{21}$$

for $k \neq \ell$ and

$$[\boldsymbol{M}_{k,k}]_{i,i} = \|\boldsymbol{V}_k \boldsymbol{W}_k \boldsymbol{e}_i\|_2^2 - \sum_{\ell \neq k} \sum_{j=1}^{2S_\ell} [\boldsymbol{M}_{\ell,k}]_{j,i}. \tag{22}$$

Now assume that all uplink beamformers $\boldsymbol{T}_k = \grave{\boldsymbol{T}}_k$ are $\mathcal{BSC}_2$ matrices. Then, $\boldsymbol{T}_k^{\mathrm{T}} \grave{\boldsymbol{H}}_k^{\mathrm{T}} \boldsymbol{Z}_k^{-1} \grave{\boldsymbol{H}}_k \boldsymbol{T}_k$ is $\mathcal{BSC}_2$ since multiplication, inversion, summation, and transposition preserve the $\mathcal{BSC}_2$ structure. Thus, we can use the standard eigenvalue decomposition (cf. Section II) so that $\boldsymbol{W}_k = \grave{\boldsymbol{W}}_k$ is a $\mathcal{BSC}_2$ matrix. We explain later why this is without loss of generality.

In general, an individual rescaling of columns of a $\mathcal{BSC}_2$ matrix destroys the $\mathcal{BSC}_2$ structure, which can be easily seen from (1). However, this is not true for scaling with the coefficients $\alpha_{k,i}$ in (18). To show this, we need the following lemma on the Hadamard (element-wise) product of block-skew-circulant matrices.

Lemma 1: The Hadamard product of two $\mathcal{BSC}_2$ matrices is block-circulant ($\mathcal{BC}_2$).

Proof: Let $\grave{\boldsymbol{A}} \in \mathcal{BSC}_2^{K\times L}$ and $\grave{\boldsymbol{A}}' \in \mathcal{BSC}_2^{K\times L}$ with the block structure as in (1). Then,

$$\grave{\boldsymbol{A}} \odot \grave{\boldsymbol{A}}' = \begin{bmatrix} \boldsymbol{A}_1 \odot \boldsymbol{A}_1' & \boldsymbol{A}_2 \odot \boldsymbol{A}_2' \\ \boldsymbol{A}_2 \odot \boldsymbol{A}_2' & \boldsymbol{A}_1 \odot \boldsymbol{A}_1' \end{bmatrix} \tag{23}$$

where $\odot$ is the Hadamard product. ■

Due to Lemma 1, $\boldsymbol{M}_{k,\ell}$ is block-circulant ($\mathcal{BC}_2$) for all k, ℓ if all uplink beamformers $\boldsymbol{T}_k = \grave{\boldsymbol{T}}_k$ are $\mathcal{BSC}_2$ matrices. Moreover, $\|\boldsymbol{T}_k \boldsymbol{W}_k \boldsymbol{e}_i\|_2^2 = \|\boldsymbol{T}_k \boldsymbol{W}_k \boldsymbol{e}_{i+S_k}\|_2^2$ and $\|\boldsymbol{V}_k \boldsymbol{W}_k \boldsymbol{e}_i\|_2^2 = \|\boldsymbol{V}_k \boldsymbol{W}_k \boldsymbol{e}_{i+S_k}\|_2^2$ in this case. Therefore, it can be easily verified, that we obtain pairs of equal scalars $\alpha_{k,i+S_k} = \alpha_{k,i}$, i.e., the diagonal matrix $\mathrm{diag}_{i=1}^{2S_k}\{\alpha_{k,i}\}$ is $\mathcal{BSC}_2$.

Finally, since $\boldsymbol{B}_k$ is the product of three $\mathcal{BSC}_2$ matrices, it is a $\mathcal{BSC}_2$ matrix.

Similar observations hold when computing the optimal downlink receivers and performing the downlink-to-uplink transformation from [22].

In fact, the eigenvalue decomposition used to compute $\boldsymbol{W}_k$ is ambiguous with respect to permutations of the eigenvalues and eigenvectors, and to the choice of the bases of eigenspaces corresponding to eigenvalues with multiplicity larger than one (see the discussion in [19]). To account for these ambiguities, we can replace $\boldsymbol{W}_k$ by $\grave{\boldsymbol{W}}_k \boldsymbol{Q}_k \notin \mathcal{BSC}_2$, where $\grave{\boldsymbol{W}}_k$ is $\mathcal{BSC}_2$ and $\boldsymbol{Q}_k$ is orthogonal. It is easy to verify that the matrices $\boldsymbol{B}_k$, $\boldsymbol{V}_k$, $\boldsymbol{T}_k$, etc. are no longer $\mathcal{BSC}_2$ in this case, but their Gramians still are since $\boldsymbol{Q}_k$ cancels out when computing the Gramians. Moreover, this property is then preserved from one iteration to the next. Since the downlink transmit covariance matrix is equal to the Gramian $\boldsymbol{B}_k \boldsymbol{B}_k^{\mathrm{T}}$, the matrix $\boldsymbol{Q}_k$ does not have an influence on the distribution of the transmit signal, and we can assume $\boldsymbol{W}_k = \grave{\boldsymbol{W}}_k$ without loss of generality.

We have established that alternating filter updates preserve the $\mathcal{BSC}_2$ structure. The implications are the same as for gradient-based methods, namely that an initialization with $\mathcal{BSC}_2$ matrices leads to a solution consisting of $\mathcal{BSC}_2$ matrices while a non-$\mathcal{BSC}_2$ initialization allows for general solutions.

C. Individual Scaling of Beamforming Vectors

Alternating filter updates are often combined with additional scaling steps that adapt the individual powers of the data streams, for instance, in order to fulfill rate targets (e.g., [26], [27]). This corresponds to scaling individual columns of transmit filter matrices. Since scaling the columns individually can destroy a $\mathcal{BSC}_2$ structure, we have to include such scaling steps in the analysis when studying whether or not an iterative algorithm preserves the $\mathcal{BSC}_2$ structure.

As an example, we consider the algorithm from [27], which introduces per-stream rate targets $\rho_k^{(s)}$ as auxiliary variables, applies a gradient-projection update to the targets $\rho_k^{(s)}$ in each iteration, and then scales the filter vectors such that the per-stream rate targets are achieved. Due to space constraints, we cannot reproduce the details here and refer the reader to [27].

Analyzing the scaling procedure step by step, it turns out that a set of filters where all filter vectors occur in pairs of

$$\boldsymbol{t}_{k,i} \quad \text{and} \quad \boldsymbol{t}_{k,i+S_k} = \grave{\boldsymbol{J}}_{N_k} \boldsymbol{t}_{k,i} \tag{24}$$

with

$$\grave{\boldsymbol{J}}_N = \begin{bmatrix} \mathbf{0} & -\mathbf{I}_N \\ \mathbf{I}_N & \mathbf{0} \end{bmatrix} \tag{25}$$

leads to a set of scaling factors where all factors form pairs of $\beta_{k,i} = \beta_{k,i+S_k}$ (similar as for the factors $\alpha_{k,i}$ in the previous section). The formal proof of this statement again relies on the framework of $\mathcal{BSC}$ matrices—in particular on the fact that transposition, addition, multiplication, and inversion preserve the $\mathcal{BSC}_2$ structure, on the properties of the matrix $\grave{\boldsymbol{J}}_N$ given in [19, Lemma 10], and on the exploitation of a similar block-circulant structure as in Lemma 1. The result $\beta_{k,i} = \beta_{k,i+S_k}$ implies that this scaling preserves the $\mathcal{BSC}_2$ structure.

A particularity in the scaling steps in [27] is that the algorithm allows that inactive streams with zero power can become active again. Therefore, filters of inactive streams are updated in [27] using a special update procedure. Due to space constraints, we cannot reproduce this procedure here and have to refer the reader to [27]. To show that the updated inactive filters form $\mathcal{BSC}_2$ structures if all matrices have $\mathcal{BSC}_2$ structure in the previous step, we have to analyze the procedure step by step and apply similar arguments as above. However, since one of the steps (see [27]) is a generalized eigenvalue decomposition, we additionally need the following new lemma on $\mathcal{BSC}_2$ matrices to complete the analysis.

Lemma 2: The generalized eigenvectors of a pair of symmetric matrices $\grave{\boldsymbol{A}}, \grave{\boldsymbol{C}} \in \mathcal{BSC}_2^{M\times M}$ can be arranged in a $\mathcal{BSC}_2$ matrix.

Proof: With $\grave{\boldsymbol{J}}_M$ from (25), we have $\grave{\boldsymbol{J}}_M \grave{\boldsymbol{A}} \grave{\boldsymbol{J}}_M^{\mathrm{T}} = \grave{\boldsymbol{A}}$ for symmetric $\grave{\boldsymbol{A}} \in \mathcal{BSC}_2^{M\times M}$ and $\boldsymbol{q}^{\mathrm{T}} \grave{\boldsymbol{J}}_M \boldsymbol{q} = 0$ for any $\boldsymbol{q} \in \mathbb{R}^{2M}$ [19, Lemma 10]. Thus, for any generalized eigenvalue ϕ and generalized eigenvector $\boldsymbol{q}$ of $(\grave{\boldsymbol{A}}, \grave{\boldsymbol{C}})$, we have ($\grave{\boldsymbol{J}}_M^{\mathrm{T}} = \grave{\boldsymbol{J}}_M^{-1}$)

$$\begin{aligned} \grave{\boldsymbol{A}} \boldsymbol{q} = \grave{\boldsymbol{C}} \boldsymbol{q} \phi \;&\Leftrightarrow\; \grave{\boldsymbol{J}}_M \grave{\boldsymbol{A}} \grave{\boldsymbol{J}}_M^{\mathrm{T}} \grave{\boldsymbol{J}}_M \boldsymbol{q} = \grave{\boldsymbol{J}}_M \grave{\boldsymbol{C}} \grave{\boldsymbol{J}}_M^{\mathrm{T}} \grave{\boldsymbol{J}}_M \boldsymbol{q} \phi \\ &\Leftrightarrow\; \grave{\boldsymbol{A}} \grave{\boldsymbol{J}}_M \boldsymbol{q} = \grave{\boldsymbol{C}} \grave{\boldsymbol{J}}_M \boldsymbol{q} \phi \end{aligned} \tag{26}$$

i.e., for each ϕ, we have a pair of orthogonal generalized eigenvectors $\boldsymbol{q}$ and $\grave{\boldsymbol{J}}_M \boldsymbol{q}$. Properly arranging them in a matrix of generalized eigenvectors, we obtain a $\mathcal{BSC}_2$ matrix. ■

D. Commonly Used Initializations

Above, we have seen that in order to obtain an algorithm that is capable of finding solutions that correspond to improper signaling, we have to make sure to initialize the iterative procedure with such an improper strategy.

This is an important observation, since commonly used initialization indeed have the $\mathcal{BSC}_2$ structure. For instance, (scaled) identity matrices (e.g., [23], [24], [27]) are obviously block-skew-circulant. Moreover, an initialization based on a singular value decomposition (e.g., [26]) of the composite real channel matrices is equivalent to a $\mathcal{BSC}_2$ initialization since the standard singular value decomposition (cf. Section II) can be used without loss of generality.

However, it is easy to break the $\mathcal{BSC}_2$ structures of these initializations by rescaling their columns with individual factors. A special case of this is setting a beamforming vector to zero, i.e., switching a real-valued data streams off. This corresponds to switching either the inphase or the quadrature component of a complex data stream off and obviously corresponds to improper signaling. Even if the algorithm might reactivate the real-valued data stream in the course of its execution, the solution after convergence can then be an improper strategy since we have avoided a $\mathcal{BSC}_2$ initialization. In fact, this is what the initialization with truncated identity matrices from [24] and the initialization with single data streams from [27] do when they are applied to the composite real representation.

Another possibility would be to use a random initialization as in [25], [27]. For instance, the entries of the beamforming matrices can be chosen to be i.i.d. Gaussian or, if it is desired that the $2S_k$ columns of each initial beamforming matrix are orthogonal to each other, we can use a (possibly truncated) eigenbasis of $\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{H}}$, where $\boldsymbol{Y}$ is a square random matrix with i.i.d. Gaussian elements [28]. In both cases, the initial filter matrices have a nonzero $\mathcal{BHSC}_2$ component almost surely.

E. Effects Caused by Numerical Inaccuracies

So far, we have assumed that all operations are performed with infinite precision. However, in an implementation, a function that theoretically delivers a $\mathcal{BSC}_2$ matrix might yield a result with a very small, but nonzero $\mathcal{BHSC}_2$ component. Then, we can no longer assume that all inputs for the next step are $\mathcal{BSC}_2$ matrices, and the line of argumentation is interrupted, i.e., the final result no longer needs to have $\mathcal{BSC}_2$ structure. Finding out how strong such numerical inaccuracies have to be in order to leave the neighborhood of the set of $\mathcal{BSC}_2$ solutions is an interesting question for future research. In the next section, we see that this effect can indeed happen in numerical simulations.

V. Simulation Results

As an example, we consider the power minimization algorithm for MIMO broadcast channels with linear transceivers proposed in [27]. On the one hand, we apply this algorithm directly to the complex-valued formulation, which is the application it has originally been developed for. On the other hand, we apply it to the composite real representation in order to optimize widely linear transceivers. Note that this algorithm was also used in [5] to optimize widely linear transceivers, but only in a MISO setting.

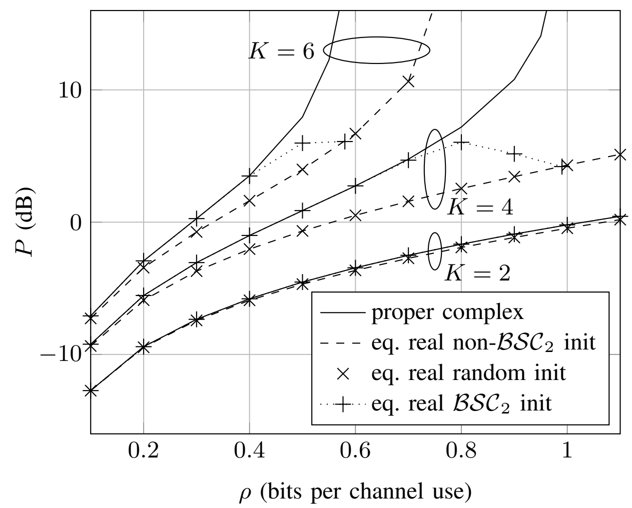

Fig. 1. Transmit power needed to serve K users with data rate $r_k = \rho$ in a MIMO broadcast channel with $M = 2$ transmit antennas and $N_k = 2$ receive antennas per user. Powers are averaged (in the dB domain) over 1000 realization of i.i.d. circularly symmetric Gaussian channel coefficients.

In the equivalent real representation, we use three different initializations. The first one, which does not correspond to a $\mathcal{BSC}_2$ structure, starts with only one real-valued data stream for which the transmit filter is a canonical unit vector. It can be seen in Fig. 1 that this initialization, which allows for solutions that correspond to improper transmit strategies, leads to lower sum transmit power than the complex-valued implementation, which applies proper transmit signals. Nearly the same average performance is obtained with random vectors as initial filters.

As an initialization that has the $\mathcal{BSC}_2$ structure, we use a pair of canonical unit vectors $\boldsymbol{e}_i$ and $\boldsymbol{e}_{i+S_k}$ for user k and initialize both real-valued data streams with the same powers and data rates. As predicted by the theoretical analysis, this $\mathcal{BSC}_2$ initialization leads to solutions that also have the $\mathcal{BSC}_2$ structure and have the same performance as the optimization with proper complex signals, i.e., cannot achieve the performance gains of improper signaling.

In the plot, we can also observe the effects caused by numerical inaccuracies described in Section IV-E. To understand this, we first have to note that the optimization can become infeasible for high numbers of users and high data rates (see [5]), which corresponds to diverging curves in the plot. Note that the feasibility region with improper signaling is larger than with proper signaling [5]. If the rate requirements are close to the feasibility boundary, intermediate results with very large powers can occur before the algorithm converges. Apparently, the precision is impaired by these large values strong enough to push the algorithm away from the set of $\mathcal{BSC}_2$ solutions, i.e., it no longer yields the same result as the complex version.

Even though this might seem counterintuitive at the first glance, the simulation results show that the numerical inaccuracies can be helpful instead of harmful in this context. However, since we cannot expect that such an effect always happens, it is preferable to break the $\mathcal{BSC}_2$ structure explicitly, e.g., by using an appropriate initialization.

VI. Discussion

While it is straightforward that iterative heuristic algorithm for optimizing proper signaling in MIMO broadcast channels with linear transceivers can be applied to the composite real representation in order to optimize settings with widely linear transceivers, the interesting question is whether the obtained solutions indeed correspond to improper signaling. In this paper, we have shown that this can be the case only if we use an initialization that corresponds to improper signaling. The reason for this is that the existing iterative optimization algorithms tend to preserve the so-called $\mathcal{BSC}_2$ structure, which is the characterizing structure of the composite real equivalents of proper complex transmit strategies.

A question similar to the one considered here was studied in [20] for multicarrier MIMO broadcast channels with proper signaling. It was investigated whether existing algorithms can be applied to an equivalent single-carrier formulation of multicarrier communication systems to optimize carrier-cooperative transmission (coding across carriers, for a formal definition of this concept see [20]). In that case, the matter of interest was not the $\mathcal{BSC}_2$ structure. Instead, the particularity was that the channel matrices had a block-diagonal structure. However, a similar result was obtained: it turned out that existing algorithm tend to yield solutions that match the block structure of the channel matrices. In particular, it was shown that gradient methods and alternating filter updates converge to block-diagonal solutions if the initialization is block-diagonal.

An important difference between these two applications is that individual scaling of the columns of the filter matrices can destroy the $\mathcal{BSC}_2$ structure, but preserves block-diagonality. Therefore, while rescaling is a simple way to construct adequate initializations for the optimization of improper transmit strategies, more thought has to be put into finding a sensible initialization when trying to optimize carrier-cooperative transmission via an equivalent block-diagonal formulation. On the other hand, the sensitivity of the $\mathcal{BSC}_2$ structure to rescaling of individual columns made the theoretical analysis in this paper more involved than the one in [20].

References

[1] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.

[2] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3936–3964, Sep. 2006.

[3] U. Erez and S. ten Brink, "A close-to-capacity dirty paper coding scheme," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3417–3432, October 2005.

[4] D. P. Palomar and Y. Jiang, "MIMO transceiver design via majorization theory," *Foundations and Trends® Commun. and Inf. Theory*, vol. 3, no. 4–5, pp. 331–551, 2006.

[5] C. Hellings, M. Joham, and W. Utschick, "QoS feasibility in MIMO broadcast channels with widely linear transceivers," *IEEE Signal Process. Lett.*, vol. 20, no. 11, pp. 1134–1137, Nov. 2013.

[6] C. Hellings and W. Utschick, "Linear precoding in parallel MIMO broadcast channels: Separable and inseparable scenarios," presented at the Int. ITG Workshop on Smart Antennas (WSA) 2013, Stuttgart, Germany, Mar., 13–14, 2013.

[7] B. Picinbono and P. Chevalier, "Widely linear estimation with complex data," *IEEE Trans. Signal Process.*, vol. 43, no. 8, pp. 2030–2033, Aug. 1995.

[8] T. Adalı, P. J. Schreier, and L. L. Scharf, "Complex-valued signal processing: The proper way to deal with impropriety," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5101–5125, Nov. 2011.

[9] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1293–1302, Jul. 1993.

[10] C. Hellings and W. Utschick, "Performance gains due to improper signals in MIMO broadcast channels with widely linear transceivers," in *Proc. Int. Conf. Acoust., Speech, and Signal Process. (ICASSP) 2013*, May 2013, pp. 4379–4383.

[11] Y. Zeng, R. Zhang, E. Gunawan, and Y. L. Guan, "Optimized transmission with improper Gaussian signaling in the K-user MISO interference channel," *IEEE Trans. Wireless Commun.*, vol. 12, no. 12, pp. 6303–6313, Dec. 2013.

[12] D. Mattera, L. Paura, and F. Sterle, "Widely linear MMSE transceiver for real-valued sequences over MIMO channel," presented at the 14th European Signal Process. Conf. (EUSIPCO), Florence, Italy, Sep. 4–8, 2006.

[13] S. Buzzi, M. Lops, and S. Sardellitti, "Widely linear reception strategies for layered space-time wireless communications," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2252–2262, Jun. 2006.

[14] C. Hellings, L. Weiland, and W. Utschick, "Optimality of proper signaling in Gaussian MIMO broadcast channels with shaping constraints," in *Proc. Int. Conf. Acoust., Speech, and Signal Process. (ICASSP) 2014*, May 2014, pp. 3498–3502.

[15] V. R. Cadambe, S. A. Jafar, and C. Wang, "Interference alignment with asymmetric complex signaling—settling the Høst-Madsen–Nosratinia conjecture," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4552–4565, Sep. 2010.

[16] Z. K. M. Ho and E. Jorswieck, "Improper Gaussian signaling on the two-user SISO interference channel," *IEEE Trans. Wireless Commun.*, vol. 11, no. 9, pp. 3194–3203, Sep. 2012.

[17] C. Hellings, L. Gerdes, L. Weiland, and W. Utschick, "On optimal Gaussian signaling in MIMO relay channels with partial decode-and-forward," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3153–3164, Jun. 2014.

[18] Z. Ho, A. Zappone, and E. Jorswieck, "Optimal non-regenerative relay processing with improper signals," in *Proc. 10th Int. Symp. Wireless Commun. Syst. (ISWCS)*, Aug. 2013, pp. 567–571.

[19] C. Hellings and W. Utschick, "Block-skew-circulant matrices in complex-valued signal processing," Jun. 2014, submitted to *IEEE. Trans. Signal Process.*

[20] C. Hellings and W. Utschick, "Carrier-cooperative transmission in parallel MIMO broadcast channels: Potential gains and algorithms," in *Proc. 8th Int. Symp. Wireless Commun. Syst. (ISWCS)*, Nov. 2011, pp. 774–778.

[21] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov./Dec. 1999.

[22] R. Hunger and M. Joham, "A general rate duality of the MIMO multiple access channel and the MIMO broadcast channel," presented at the IEEE GLOBECOM 2008, New Orleans, LA, USA, Nov. 30–Dec. 4, 2008.

[23] R. Böhnke and K.-D. Kammeyer, "Weighted sum rate maximization for MIMO-OFDM systems with linear and dirty paper precoding," presented at the 7th Int. ITG Conf. Source and Channel Coding (SCC), Ulm, Germany, Jan. 14–16, 2008.

[24] R. Hunger, D. Schmidt, and M. Joham, "A combinatorial approach to maximizing the sum rate in the MIMO BC with linear precoding," in *Conf. Record 42nd Asilomar Conf. Signals, Syst., and Comput. (ACSSC)*, Oct. 2008, pp. 316–320.

[25] M. Codreanu, A. Tölli, M. Juntti, and M. Latva-aho, "Joint design of Tx-Rx beamformers in MIMO downlink channel," *IEEE Trans. Signal Process.*, vol. 55, no. 9, pp. 4639–4655, Sep. 2007.

[26] S. Shi, M. Schubert, and H. Boche, "Rate optimization for multiuser MIMO systems with linear processing," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 4020–4030, Aug. 2008.

[27] C. Hellings, M. Joham, and W. Utschick, "Gradient-based power minimization in MIMO broadcast channels with linear precoding," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 877–890, Feb. 2012.

[28] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Commun. Inf. Theory*, vol. 1, no. 1, pp. 1–182, 2004.