A Deterministic Initialization for Optimizing Carrier-Cooperative Transmit Strategies

Christoph Hellings, Stephan Herrmann, and Wolfgang Utschick

A Deterministic Initialization for Optimizing Carrier-Cooperative Transmit Strategies

Christoph Hellings, Stephan Herrmann, 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,s.herrmann,utschick}@tum.de

Abstract—Even though capacity can be achieved in parallel multiple-input multiple-output (MIMO) broadcast channels by coding separately on each carrier, joint coding across carriers (so-called carrier-cooperative transmission) can lead to performance gains in MIMO broadcast channels with a restriction to linear transceivers. In principle, carrier-cooperative transmission can be optimized in an equivalent single-carrier MIMO system, but it has been shown that most existing optimization algorithms lead to solutions that are equivalent to carrier-noncooperative transmission. The only exception discussed in the literature is the application of iterative algorithms with a random initialization. In this paper, we show that randomness is not a requirement to obtain good carrier-cooperative solutions. We do so by proposing a deterministic carrier-cooperative initialization that has good performance in numerical simulations as well as interesting interpretations: it can be interpreted in the context of equal gain transmission and based on the notion of mutual incoherence.

Index Terms—carrier-cooperative transmission, linear transceivers, multiple-input multiple-output (MIMO), multiuser multicarrier systems, parallel broadcast channels.

I. INTRODUCTION

So-called dirty paper coding (DPC, e.g., [1]) is known to be the capacity-achieving coding scheme for multiple-input multiple-output (MIMO) broadcast channels [1], [2]. Since this interference precompensation scheme has prohibitive complexity for practical implementation [3], linear transceivers are considered as a low-complexity alternative. Finding (close-to-)optimal linear transmit strategies for MIMO broadcast channels is a problem that has attracted the interest of many researchers (e.g., [4]–[7]).

The restriction to linear techniques changes the nature of the optimization fundamentally: it not only leads to nonconvex optimization problems, for which the globally optimal solution cannot be found efficiently, but it also renders well-established paradigms invalid. For instance, it is known that performing data transmission on each carrier separately using an optimized allocation of transmit power to carriers is optimal in MIMO broadcast channels with DPC [8]. However, in broadcast channels with linear transceivers, breaking with the paradigm of separate coding can lead to performance gains [9]. Another example for such a paradigm change, which is, however, not considered in this paper, is the use of improper (i.e., noncircular) transmit signals, which can lead to performance gains in MIMO broadcast channels without DPC [10], [11].

When deciding to employ joint coding across carriers (often referred to as carrier-cooperative transmission [12]–[14]), we have to find a way to optimize this joint transmission. In

[13], it was proposed to introduce an equivalent single-carrier broadcast channel with block-diagonal channel matrices and to apply existing optimization algorithms for MIMO broadcast channels to this equivalent setting. In the earlier work [12], a similar approach was pursued for the single-user case, i.e., for a point-to-point MIMO system.

The important particularity of the equivalent single-carrier broadcast channel is that the channel matrices are block-diagonal. The question of how conventional algorithms for the optimization of linear transceivers in MIMO broadcast channels (e.g., [4]–[7], [15]–[18]) behave when applied to channels with this special structure was studied in [13]. Unfortunately, it turns out that most of the algorithms lead to block-diagonal transmit covariance matrices, which is equivalent to carrier-noncooperative transmission (separate coding on each carrier). Among the optimization methods studied in [13], the only one that leads to carrier-cooperative solutions is the application of iterative algorithms (e.g., gradient methods [6], [15], alternating filter updates [4], [5], [7]) in combination with a random initialization.

In [19], an algorithm for power minimization under minimum rate constraints was developed for multicarrier MIMO broadcast channels with linear zero-forcing beamforming and carrier-cooperative transmission. In numerical simulations, this algorithm was shown to be able to close half of the gap between the carrier-noncooperative reference algorithm from [20] and the globally optimal strategy using dirty-paper-coding without zero-forcing. One of the main ingredients of this algorithm was again a gradient-based method with random initialization.

Optimization of carrier-cooperative strategies using random initializations proved to be beneficial also in other system models. In particular, the application of this concept was studied for various relay-assisted multicarrier interference channel scenarios in [14], [21].

Despite its good performance in numerical studies, random initialization is not completely satisfying from a theoretical point of view [13]. The quality of the solutions obtained by iterative methods for nonconvex problems can depend strongly on the initialization, and we might want to develop algorithms with a deterministic outcome (same solution when applied twice to the same channel realization). Furthermore, and even more important, blindly using a random initialization is not very insightful when trying to understand which initialization leads to good results. Therefore, pseudorandom initializations

or arbitrarily chosen deterministic initializations without a clear rationale are not satisfying, either.

After introducing the system model and the notions of carrier-cooperative and carrier-noncooperative transmission in Section II, we summarize some key points of [13] in Section III in order to explain the need for a new deterministic initialization. Then, we propose a deterministic initialization based on the Fourier matrix (discrete Fourier transform matrix, DFT matrix), which not only leads to good performance in numerical studies (Section IV), but also to interesting interpretations, which are discussed in Section V.

Notation: Vectors and matrices are typeset in boldface lowercase and uppercase letters, respectively. We use $\bullet^{\mathrm{H}}$ for the conjugate transpose, $\mathbf{I}_L$ for the identity matrix of size L, and $\mathbf{0}$ for the zero vector. The absolute value of a scalar is denoted by $|\bullet|$.

II. SYSTEM MODEL AND PROBLEM FORMULATION

Consider a set of C parallel K-user broadcast channels[1] with M transmit antennas and N_k antennas at receiver k. We collect the channel matrices $\boldsymbol{H}_k^{(c),\mathrm{H}} \in \mathbb{C}^{N_k \times M}$ and noise covariance matrices $\boldsymbol{C}_{\boldsymbol{\eta}_k}^{(c)} \in \mathbb{C}^{N_k \times N_k}$ on the subchannels (e.g., carriers) $c \in \{1, \ldots, C\}$ in the block diagonal matrices

$$\boldsymbol{H}_k^{\mathrm{H}} = \mathrm{blockdiag}\left(\boldsymbol{H}_k^{(1),\mathrm{H}}, \ldots, \boldsymbol{H}_k^{(C),\mathrm{H}}\right) \in \mathbb{C}^{N_k C \times MC} \quad (1)$$

$$\boldsymbol{C}_{\boldsymbol{\eta}_k} = \mathrm{blockdiag}\left(\boldsymbol{C}_{\boldsymbol{\eta}_k}^{(1)}, \ldots, \boldsymbol{C}_{\boldsymbol{\eta}_k}^{(C)}\right) \in \mathbb{C}^{N_k C \times N_k C}. \quad (2)$$

The additive noise of all carriers is collected in the noise vector $\boldsymbol{\eta}_k \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{C}_{\boldsymbol{\eta}_k})$.

Using linear transceivers, the transmission of circularly symmetric Gaussian symbol vectors[2] $\boldsymbol{x}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{S_k})$ (containing $S_k \leq C \min\{N_k, M\}$ independent data streams for user k) can be described in an equivalent single-carrier setting

$$\hat{\boldsymbol{x}}_k = \boldsymbol{V}_k^{\mathrm{H}} \boldsymbol{H}_k^{\mathrm{H}} \sum_{k'=1}^{K} \boldsymbol{B}_{k'} \boldsymbol{x}_{k'} + \boldsymbol{V}_k^{\mathrm{H}} \boldsymbol{\eta}_k. \quad (3)$$

Here, we have employed beamforming matrices $\boldsymbol{B}_{k'} \in \mathbb{C}^{MC \times S_{k'}}$ and receive filters $\boldsymbol{V}_k^{\mathrm{H}} \in \mathbb{C}^{S_k \times N_k C}$.

Without structural constraints on these filters, transmission is allowed to be carrier-cooperative (joint coding across carriers). In this case, the sum transmit power P and the per-user rate r_k have to be computed using

$$P = \sum_{k=1}^{K} \mathrm{trace}\left[\boldsymbol{B}_k \boldsymbol{B}_k^{\mathrm{H}}\right]. \quad (4)$$

and

$$r_k = \log \det\left(\mathbf{I}_{N_k C} + \boldsymbol{R}_k^{-1} \boldsymbol{H}_k^{\mathrm{H}} \boldsymbol{B}_k \boldsymbol{B}_k^{\mathrm{H}} \boldsymbol{H}_k\right) \quad (5)$$

$$\text{with} \quad \boldsymbol{R}_k = \boldsymbol{C}_{\boldsymbol{\eta}_k} + \sum_{j \neq k} \boldsymbol{H}_k^{\mathrm{H}} \boldsymbol{B}_j \boldsymbol{B}_j^{\mathrm{H}} \boldsymbol{H}_k. \quad (6)$$

Due to the inverse of $\boldsymbol{R}_k$, the rate r_k is not a concave function of the beamforming matrices $\boldsymbol{B}_{k'}$, $k' = 1, \ldots, K$. Consequently, an optimization problem with r_k occurring in the objective function or in the constraints is a nonconvex problem. An example is the power minimization with minimum rate constraints (12) considered in Section IV. For such nonconvex problems, many different local optima can exist.

On the other hand, in case of block-diagonal filter matrices, the transmission is carrier-noncooperative[3] (separate coding on each carrier), and (3) can be decomposed as

$$\hat{\boldsymbol{x}}_k^{(c)} = \boldsymbol{V}_k^{(c),\mathrm{H}} \boldsymbol{H}_k^{(c),\mathrm{H}} \sum_{k'=1}^{K} \boldsymbol{B}_{k'}^{(c)} \boldsymbol{x}_{k'}^{(c)} + \boldsymbol{V}_k^{(c),\mathrm{H}} \boldsymbol{\eta}_k^{(c)}. \quad (7)$$

In this case, the subchannels are coupled only by the per-user rates $r_k = \sum_{c=1}^{C} r_k^{(c)}$ and the sum power $P = \sum_{c=1}^{C} P^{(c)}$. Here,

$$P^{(c)} = \sum_{k=1}^{K} \mathrm{trace}\left[\boldsymbol{B}_k^{(c)} \boldsymbol{B}_k^{(c),\mathrm{H}}\right] \quad (8)$$

is the transmit power on carrier c. The data rate

$$r_k^{(c)} = \log \det\left(\mathbf{I}_{N_k} + \boldsymbol{R}_k^{(c),-1} \boldsymbol{H}_k^{(c),\mathrm{H}} \boldsymbol{B}_k^{(c)} \boldsymbol{B}_k^{(c),\mathrm{H}} \boldsymbol{H}_k^{(c)}\right) \quad (9)$$

$$\text{with} \quad \boldsymbol{R}_k^{(c)} = \boldsymbol{C}_{\boldsymbol{\eta}_k^{(c)}} + \sum_{j \neq k} \boldsymbol{H}_k^{(c),\mathrm{H}} \boldsymbol{B}_j^{(c)} \boldsymbol{B}_j^{(c),\mathrm{H}} \boldsymbol{H}_k^{(c)} \quad (10)$$

of user k on carrier c is again a nonconcave function of the beamforming matrices.

III. CARRIER-COOPERATIVE AND CARRIER-NONCOOPERATIVE INITIALIZATIONS

In [13], the application of iterative optimization algorithms to the equivalent single-carrier MIMO broadcast channel (3) was studied. In particular, gradient-based methods [6], [15] and methods based on alternating filter updates in the downlink and the dual uplink [4], [5], [7] were investigated. It was shown that these methods can converge to carrier-cooperative solutions only if carrier-cooperative filter matrices are used as initialization for the iterative procedures [13].

Choices for the initial filter matrices that are well established in the existing literature are, e.g., (truncated) identity matrices (e.g., [6]) and matrices consisting of singular vectors of the channels (e.g., [5]). Unfortunately, these choices match the block-structure of the channels and are, thus, carrier-noncooperative initializations [13].

As a carrier-cooperative initialization, it was proposed to use random initial filter matrices [13], [14], [19]. For instance, the filter vectors (i.e., the columns of the beamforming matrices or the rows of the receive filter matrices) can be random orthonormal vectors [13]. This choice corresponds to carrier-cooperative transmission almost surely. However, as discussed

---

[1]Not to be confused with interfering broadcast channels, i.e., several interfering base stations. We consider a classical broadcast scenario with only one base station, but parallel channels (e.g., carriers).

[2]Note that $\boldsymbol{x}_k$ is a concatenation of all symbols intended for user k no matter across which subchannel(s) they are transmitted.

[3]In fact, transmission is mathematically equivalent to carrier-noncooperative transmission if the transmit covariance matrices $\boldsymbol{B}_k \boldsymbol{B}_k^{\mathrm{H}}$ are block-diagonal, even if the beamforming matrices $\boldsymbol{B}_k$ are not. However, for simplicity, we stick to the nomenclature of [12], [13], where the term carrier-noncooperative is defined based on the transmit filters.

in the introduction, blindly using random vectors is not insightful when it comes to the question of which initializations lead to preferable outcomes of the iterative methods.

To the best of our knowledge, a sensible choice for a deterministic carrier-cooperative initialization has not yet been discussed in the literature. A straightforward choice of a deterministic filter vector corresponding to carrier-cooperative transmission is a (scaled) all-ones vector. However, it is necessary to decide for further filter vectors for the same user, which are linearly independent from the first one.

We therefore propose using scaled columns of the Fourier matrix (DFT matrix)

$$\boldsymbol{F}_N = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & 1 & 1 & \ldots & 1 \\ 1 & \omega & \omega^2 & \ldots & \omega^{(N-1)} \\ 1 & \omega^2 & \omega^4 & \ldots & \omega^{2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{(N-1)} & \omega^{2(N-1)} & \ldots & \omega^{(N-1)(N-1)} \end{bmatrix}$$
$$\in \mathbb{C}^{N \times N} \quad (11)$$

with $\omega = \mathrm{e}^{-\frac{\mathrm{j}2\pi}{N}}$ (e.g., [22, Section 6.12]) as initial filter vectors: considering only the absolute values $|\omega^i| = 1 \; \forall i$ of the elements, each of these vectors resembles an all-ones vector, but due to the different phases, all these vectors are linearly independent. In fact, when normalized, they even form an orthonormal basis, just like the random orthonormal vectors discussed above.

In the next section, we show that using the columns of the Fourier matrix as initial filter vectors leads to approximately the same performance as the random initialization used in [13]. Then, in Section V, we discuss two interesting interpretations of the proposed initialization.

IV. NUMERICAL EVALUATION

In the following, we consider two examples of iterative algorithms: one that is based on alternating filter updates and one that is based on gradient updates of the filters. In both cases, we compare the carrier-cooperative solutions obtained with a random initialization and with the Fourier initialization, and we include a reference value obtained with a carrier-noncooperative algorithm. Both methods optimize the sum transmit power in a broadcast channel subject to minimum rate constraints:

$$\min \; P \quad \text{s.t.} \quad r_k \geq \rho_k \; \forall k \quad (12)$$

with P and r_k from (4) and (5), respectively, where the optimization is performed over all transmit and receive filters. Extending our experiments to other optimization problems and to other system models is left open for future research.

A. Power Minimization in Parallel MISO Broadcast Channels

As an example for a scenario where carrier-cooperative transmission outperforms carrier-noncooperative transmission,
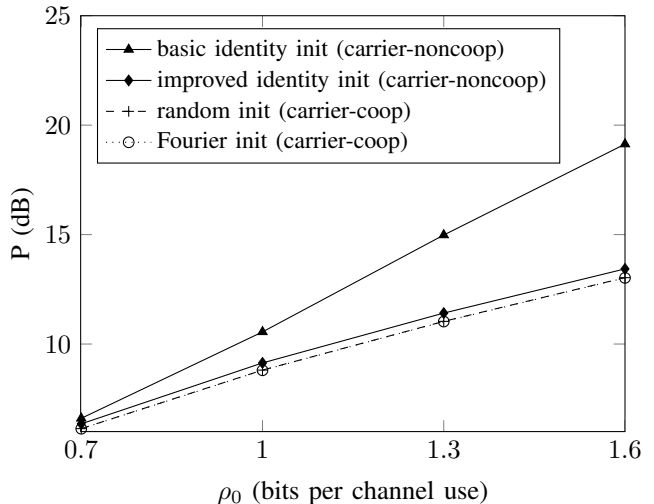
2 transmit antennas, 4 carriers, 8 users

Fig. 1. Average transmit power needed to achieve $r_k \geq 2\rho_0$ for half of the users and $r_k \geq \rho_0$ for the remaining users (1000 channel realization).

the problem of power minimization under per-user rate constraints in parallel multiple-input single-output (MISO) broadcast channels was considered in [13]. The iterative power minimization algorithm from [7] was initialized both with truncated identity matrices (carrier-noncooperative) and with matrices with random orthonormal columns (carrier-cooperative). For the initialization with truncated identity matrices, the algorithm from [7] additionally requires a feasible initialization of so-called per-stream rate targets. Loosely speaking, this can be understood as the initialization of $r_k^{(c)}$ (see Section II) and can be related to an initial scaling of the columns of the transmit filter matrices. Since a given combination of such per-user per-carrier rates is not necessarily feasible even if the requested per-user rates r_k are feasible [23], finding initial $r_k^{(c)}$ is a problem in itself. In [13], the method from [23] was used to find a basic initialization. As an improved version of the initialization with truncated identity matrices it was proposed to use the rates obtained with the greedy zero-forcing scheme from [24] as initial rate targets.[4]

In Fig. 1, we reproduce the results from [13], which show that the random initialization outperforms both versions of the initialization with truncated identity matrices. Furthermore, we have added a curve for the initialization with truncated Fourier matrices, as proposed in this paper. It turns out that this new initialization leads to the same average performance as the random initialization.

To see how significantly the outcome of an iterative algorithm for a nonconvex problem can depend on the initialization, it is insufficient to only study the average power over many channel realizations. Instead, we have to compute the power difference after convergence of the considered versions of the algorithm for each channel realization individually. This

---

[4]Note that only the per-stream rates achieved by the zero-forcing scheme are used as initialization, but not the zero-forcing filters.

2 transmit antennas, 4 carriers, 8 users

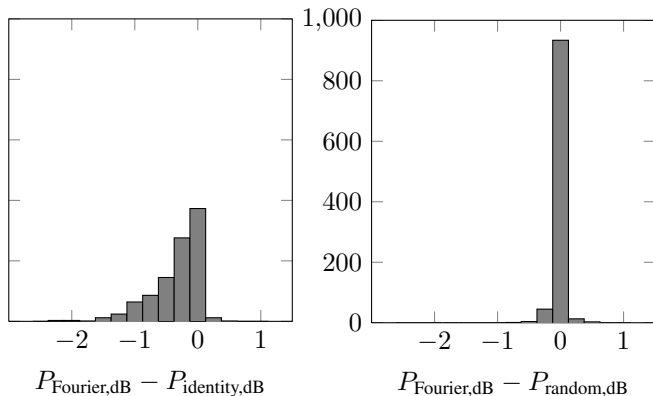$P_{\text{Fourier,dB}} - P_{\text{identity,dB}}$ $P_{\text{Fourier,dB}} - P_{\text{random,dB}}$

Fig. 2. Histograms of the power difference when using the Fourier initialization instead of the improved identity initialization or instead of the random initialization for the system in Fig. 1 with $\rho_0 = 1$.

2 transmit antennas, 5 carriers, 5 users with 2 antennas
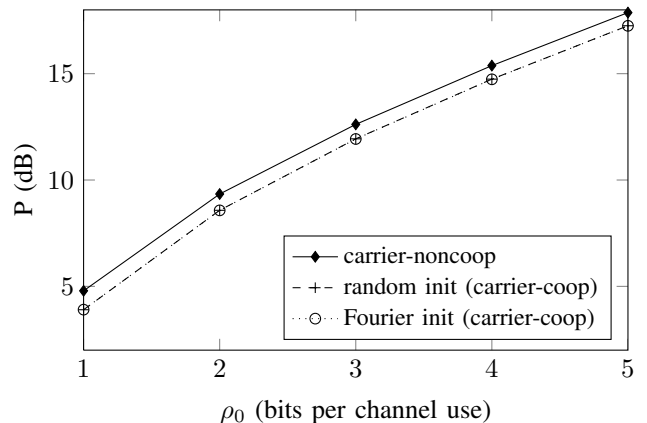
ρ_0 (bits per channel use)

Fig. 3. Average transmit power needed to achieve $r_k \geq \rho_0$ for all users in the case of spectrally similar channels [19] (1000 channel realization).

is done for the case $\rho_0 = 1$ in the histograms in Fig. 2. It can be seen that for about 40% of the channel realizations, the sum power needed after convergence is nearly the same no matter if the improved identity initialization or the proposed Fourier initialization is used. However, for the remaining 60% of the realizations, we observe a difference in power which can be as large as 2dB. Cases in which the carrier-noncooperative identity initialization outperforms the carrier-cooperative Fourier initialization are very rare.

In the second histogram in Fig. 2, we see the same comparison for the Fourier initialization and the random initialization, which are both carrier-cooperative. Apparently, these two initializations do not only lead to the same average performance as observed in Fig. 1, but also to the same individual performance for most channel realizations. Nevertheless, the histogram also reveals that these two initializations are not completely equivalent: there is a nonvanishing number of cases in which the algorithm converges to a different solution when initialized with the random initialization instead of with the Fourier initialization.

The basic identity initialization is not considered in the more detailed comparison in Fig. 2 since the large power gap between this initialization and all other considered ones already becomes very clear in Fig. 1.

B. Power Minimization in Parallel MIMO Broadcast Channels with Zero-Forcing

The second example considered in this paper is the carrier-cooperative zero-forcing algorithm proposed in [19]. The aim of this algorithm is to minimize the sum transmit power in multicarrier MIMO broadcast channels with per-user rate constraints as in (12) subject to additional zero-forcing constraints

$$\boldsymbol{v}_{k,s}^{\mathrm{H}} \boldsymbol{H}_k \boldsymbol{b}_{\ell,t} = 0 \quad \forall (k,s) \neq (\ell,t) \tag{13}$$

where $\boldsymbol{v}_{k,s}^{\mathrm{H}}$ is the sth row of $\boldsymbol{V}_k^{\mathrm{H}}$ and $\boldsymbol{b}_{\ell,t}$ is the tth column of $\boldsymbol{B}_\ell$.

Since the total number of data streams is limited by the degrees of freedom at the base station MC in the case of zero-forcing beamforming, a stream allocation is necessary. On the other hand, to obtain a carrier-cooperative solution, iterative methods with a suitable initialization are known to be a sensible approach (see Section III). Therefore, the method proposed in [19] combines a greedy stream allocation with a gradient-based optimization of the transmit and receive filters. In the original version of the algorithm, a random initialization of the filter vectors is used.

In Fig. 3, we reproduce simulation results from [19] and compare them to a modified version of the carrier-cooperative zero-forcing algorithm where the random initialization is replaced by the initialization with columns of Fourier matrices. The simulation is performed in a broadcast channel with $C = 5$ carriers, $M = 2$ transmit antennas, and $K = 5$ users with $N_k = 2$ receive antennas each. As channel model, we have used the model of spectrally similar channels defined in [19] with parameter $w^{(c)} = 0.1 + 0.9 \frac{c-1}{C-1}$ on carrier c. In this model, the channel quality depends on the carrier index. This enables us to model, e.g., interference from neighboring cells whose power varies as a function of the carrier index c. For details on this channel model, the reader is referred to [19].

Just like in the first example, we can again observe that the initialization based on Fourier matrices leads to the same average performance as the random initialization. Both versions of the algorithm outperform the carrier-noncooperative reference algorithm from [20].

V. INTERPRETATION

Having observed that the Fourier-based initialization leads to the same average performance as the random initialization in numerical simulations, we have to ask why this is the case. In the following, we try to answer this question by discussing two lines of interpretation.

First, we relate the proposed initialization to the concept of equal gain transmission [25] and then, we discuss it in the

context of mutual incoherence of orthonormal bases [26].

A. Equal Gain Transmission

The concept of equal gain transmission (EGT) was discussed, e.g., in [25], [27]. The original intention of the concept is to benefit from the potential of multiantenna systems while keeping the requirements on the transmit amplifiers modest. By using a constant amplitude of the transmit signal on all antennas and only modifying the phases, inexpensive amplifiers can be used [25]. Our interest in equal gain transmission is, however, a different one. Instead of studying implementation-related questions, we adopt the concept of EGT only as a tool to get an interpretation of filter vectors with a particular structure.

As mentioned in Section III, the columns of the Fourier matrix have the notable property that all their entries have the same magnitude. Therefore, these column vectors are suitable as transmit filters for EGT, as was already pointed out in [25].

In the equivalent single-carrier representation of a multicarrier MIMO communication system, each channel input represents a space-frequency dimension of the original system. Therefore, applying EGT to the equivalent single-carrier model means that the power is not only distributed equally among all antennas, but also among all carriers.

On the other hand, using columns of the identity matrix (i.e., canonical unit vectors) as filter vectors corresponds to the concept of antenna selection in a MIMO system [25], [28]. Applied to the equivalent single-carrier formulation of a multicarrier MIMO system, we again have to replace the term antenna by space-frequency dimension.

This means that the initialization with truncated identity matrices maps a stream to exactly one space-frequency dimension while the initialization with truncated Fourier matrices does just the opposite: it distributes the signal corresponding to a data stream equally on all space-frequency dimensions.

Note that this does not mean that the solution obtained after convergence of the iterative procedures is constrained to consist of filter vectors that are feasible for EGT. The iterative methods based on gradient steps and on alternating filter updates are able to redistribute power among the space-frequency dimensions and can potentially converge to any carrier-cooperative or carrier-noncooperative solution when initialized with EGT filters. This is in contrast to the initialization with truncated identity matrices where convergence to carrier-cooperative solutions is impossible as shown in [13].

In comparison to the initialization with random orthonormal vectors used in [13], [19], we note the following. The random initialization from [13], [19] treats all space-frequency dimensions equally from a statistical point of view, i.e., on average, the signal corresponding to a data stream is distributed equally among all antennas and all carriers. The initialization based on the Fourier matrix does the same in a deterministic manner, i.e., the signal is distributed equally on all space-frequency dimensions not only on average, but every time the algorithm is started.

Apparently, both initialization strategies (random orthonormal and Fourier) lead to the same good average performance in the power minimization simulations in Section IV. This reveals that not the randomness is the important factor when choosing the initial filters of iterative algorithms to optimize carrier-cooperative transmit strategies. Instead, the equal distribution of power to space-frequency dimensions seems to be a factor of success.

B. Mutual Incoherence of Orthonormal Bases

For a second line of interpretation, we make use of the notion of mutual incoherence of orthonormal bases. This concept was introduced in [26] with the aim of studying sparse representations of signals and became one of the foundations of compressed sensing [29]. Here, however, we apply the very same concept in a completely different context.

Consider two orthonormal bases consisting of the columns of the unitary matrices $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_N] \in \mathbb{C}^{N \times N}$ and $\boldsymbol{\Psi} = [\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_N] \in \mathbb{C}^{N \times N}$, respectively. Then, the mutual coherence of these two bases is given by [26], [30]

$$\mathrm{M}(\boldsymbol{\Phi}, \boldsymbol{\Psi}) = \sup \left\{ |\boldsymbol{\phi}_i^{\mathrm{H}} \boldsymbol{\psi}_j| : i, j \in \{1, \ldots, N\} \right\}. \qquad (14)$$

It is easy to show that $\frac{1}{\sqrt{N}} \leq \mathrm{M}(\boldsymbol{\Phi}, \boldsymbol{\Psi}) \leq 1$ [26].

As pointed out in [26], the basis defined by the identity matrix and the Fourier basis are a so-called *most mutually incoherent pair*, since their coherence achieves the lower bound, i.e., $\mathrm{M}(\boldsymbol{I}_N, \boldsymbol{F}_N) = \mathrm{M}_{\min}(N) = \frac{1}{\sqrt{N}}$.

Based on this concept, the Fourier initialization can be understood as the opposite of the initialization with truncated identity matrices. The rows of the channel matrices are sparse in the identity basis since they have nonzero elements only in the components corresponding to one of the carriers. Using the vectors of the identity basis as initialization, each initial filter vector has a common support with some rows of the channel matrices (those that correspond to the same carrier) and is orthogonal to all others. This property of the filters has previously been described by the formulation that the filters match the block-structure of the channels [13], and it has been shown that this perfect match inevitably leads to convergence to carrier-noncooperative solutions [13]. However, we are interested in an initialization that can potentially lead to both carrier-cooperative and carrier-noncooperative solutions.

According to [26], a vector that is sparse in one basis cannot be sparse in a second basis if the two bases are mutually incoherent. Therefore, the rows of the channel matrices cannot be sparse in the Fourier basis. From this point of view, using the columns of the Fourier matrix as initial filter vectors can be understood as one of the worst possible matches with the block-structure of the channel vectors. As we want to avoid a perfect match, the worst possible match is an intuitively sensible choice.

Moreover, by initializing the algorithm with one of the worst possible matches, we also avoid picking an initialization with a small amount of carrier cooperation. Recall that carrier-noncooperative transmission is stable under the considered algorithms, i.e., if we have a carrier-noncooperative strategy

in one step of the iteration, all strategies obtained in the following iterations are carrier-noncooperative as well [13]. This suggests that when starting at a point that is not far from being carrier-noncooperative, we might be within a region of attraction of the set of carrier-noncooperative strategies and quickly converge towards this set. Since the aim was to find a general initialization that does not necessarily lead to carrier-noncooperative solutions, it makes sense to choose initial filters that are far from being carrier-noncooperative, i.e., far from being a good match. As explained above, the Fourier initialization has this property.

To obtain a complete picture, we have to interpret the random orthonormal initialization within the same framework. Since we consider complex vectors, the results about the coherence of the identity basis and a real-valued random orthonormal basis from [26] do not apply. However, we can make use of the extension to complex random orthonormal (i.e., random unitary) bases given in [31]. In fact, the difference between the complex case and the real-valued case is only of quantitative, but not of qualitative nature.

Let $\boldsymbol{U}_N \in \mathcal{C}^{N \times N}$ be a random matrix that is uniformly distributed on the unitary group, i.e., a complex Haar matrix [31], [32]. This can be considered as the most natural choice for a random unitary matrix. Then, with high probability, the mutual coherence $\mathrm{M}(\boldsymbol{I}_N, \boldsymbol{U}_N)$ is not larger than $\mathrm{M}_{\mathrm{rand}}(N) = \sqrt{\frac{2 \log N}{N}}$ [31], and for $N \to \infty$, $\frac{\mathrm{M}(\boldsymbol{I}_N, \boldsymbol{U}_N)}{\mathrm{M}_{\mathrm{rand}}(N)}$ converges to 1 in probability [31].

Note that $\mathrm{M}_{\mathrm{rand}}(N) > \mathrm{M}_{\mathrm{min}}(N) = \frac{1}{\sqrt{N}}$. However, especially for large N, we have that $\mathrm{M}_{\mathrm{rand}}(N) \ll 1 = \mathrm{M}_{\mathrm{max}}$. For the scenario studied in this paper, the total dimension N is the product of the number of transmit antennas M and the number of carriers C, i.e., we have a large N if either M or C or both are large.

Consequently, as was already pointed out in [26], a random orthonormal basis is typically quite incoherent with the identity basis. For the initialization of an iterative algorithm to optimize carrier-cooperative transmit strategies, this has the following implications. Just like the Fourier initialization, the random orthonormal initialization is typically a rather bad match with the block-structure of the channel matrices.

It is obvious that a random initialization drawn from a general continuous distribution is carrier-cooperative almost surely. However, we now have a stronger statement: we now know that by using the columns of a Haar matrix $\boldsymbol{U}_N$ as initial filter vectors, we also avoid picking initializations that are, despite being carrier-cooperative, very close to carrier-noncooperative filters. In other words: it is clear that with random matrices, we do not obtain a *perfect* match of the block-structure of the channels, but the columns of Haar matrices as random orthonormal initialization, we also avoid any kind of a *good* match. The study based on the concept of mutual incoherence tells us that the columns of a Haar matrix are typically not far from being one of the *worst possible* matches. Recall that we have obtained the same statement for the initialization based on the Fourier matrix.

This gives us an additional intuitive explanation of why the average performances of the Fourier initialization and the random initialization are very similar in our numerical experiments.

VI. Summary and Outlook

By proposing an initialization based on the Fourier matrix and by demonstrating its favorable behavior in numerical simulations, we have shown that randomness of the initialization is not a requirement for the optimization of carrier-cooperative transmit strategies using iterative algorithms. The numerical experiments have been performed for the example of power minimization under per-user rate constraints in parallel MIMO broadcast channels, but we conjecture that the Fourier initialization also leads to good results in other system models and for other optimization problems. A study of such other scenarios is left open for future research.

Our conjecture is based on the fact that the Fourier initialization has a clear rationale, which we have discussed in detail in this paper. The good performance and the similarity to the random orthonormal initialization can be interpreted in the context of equal gain transmission as well as based on the notion of mutual incoherence of orthonormal bases. We think that a further study, in particular of the latter aspect, might help to gain more insights on the problem of optimizing carrier-cooperative strategies.

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] 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.

[5] 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.

[6] 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.

[7] 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.

[8] M. Mohseni, R. Zhang, and J. Cioffi, "Optimized transmission for fading multiple-access and broadcast channels with multiple antennas," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1627–1639, Aug. 2006.

[9] C. Hellings and W. Utschick, "On the inseparability of parallel MIMO broadcast channels with linear transceivers," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 6273–6278, Dec. 2011.

[10] ——, "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," Oct. 2013, unpublished, preprint available at http://arxiv.org/abs/1303.2766.

[12] D. P. Palomar, M. A. Lagunas, and J. M. Cioffi, "Optimum linear joint transmit-receive processing for MIMO channels with QoS constraints," *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1179–1197, May 2004.

[13] 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.

[14] R. T. L. Rolny, M. Kuhn, A. Amah, and A. Wittneben, "Multi-cell cooperation using subcarrier-cooperative two-way amplify-and-forward relaying," in *Proc. 10th Int. Symp. Wireless Commun. Syst. (ISWCS)*, Aug. 2013, pp. 259–263.

[15] 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.

[16] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 528–541, Mar. 2006.

[17] F. Boccardi and H. Huang, "A Near-Optimum Technique using Linear Precoding for the MIMO Broadcast Channel," in *Proc. Int. Conf. Acoust., Speech, and Signal Process. (ICASSP) 2007*, vol. 3, Apr. 2007, pp. III–17–III–20.

[18] C. Guthy, W. Utschick, G. Dietl, and P. Tejera, "Efficient linear successive allocation for the MIMO broadcast channel," in *Conf. Record 42nd Asilomar Conf. Signals, Syst., and Comput. (ACSSC)*, Oct. 2008, pp. 336–340.

[19] S. Herrmann, C. Hellings, and W. Utschick, "Carrier-cooperative zero-forcing for power minimization in parallel MIMO broadcast channels," presented at the 46th Asilomar Conf. Signals, Syst., and Comput. (ACSSC), Pacific Grove, CA, USA, Nov. 4–7, 2012.

[20] C. Guthy, W. Utschick, and G. Dietl, "Spatial resource allocation for the multiuser multicarrier MIMO broadcast channel – a QoS optimization perspective," in *Proc. Int. Conf. Acoust., Speech, and Signal Process. (ICASSP) 2010*, Mar. 2010, pp. 3166–3169.

[21] Z. Ho, E. Jorswieck, and S. Engelmann, "Information leakage neutralization for the multi-antenna non-regenerative relay-assisted multi-carrier interference channel," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 9, pp. 1672–1686, Sep. 2013.

[22] J. O. Smith III, *Mathematics of the Discrete Fourier Transform: With Audio Applications*. W3K Publishing, 2006.

[23] M. Joham, C. Hellings, and R. Hunger, "QoS feasibility for the MIMO broadcast channel: Robust formulation and multi-carrier systems," in *Proc. 8th Int. Symp. Modeling and Optim. Mobile, Ad Hoc, and Wireless Netw. (WiOpt)*, May 2010, pp. 610–614.

[24] C. Hellings, M. Joham, and W. Utschick, "Power minimization in parallel vector broadcast channels with zero-forcing beamforming," presented at the IEEE GLOBECOM 2010, Miami, FL, USA, Dec. 6–10, 2010.

[25] D. J. Love and R. W. Heath, "Equal gain transmission in multiple-input multiple-output wireless systems," *IEEE Trans. Commun.*, vol. 51, no. 7, pp. 1102–1110, Jul. 2003.

[26] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845–2862, Nov. 2001.

[27] S.-H. L. Tsai, "An Equal Gain Transmission in MIMO Wireless Communications," presented at the IEEE GLOBECOM 2010, Miami, FL, USA, Dec. 6–10, 2010.

[28] S. Sanayei and A. Nosratinia, "Antenna selection in MIMO systems," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 68–73, Oct. 2004.

[29] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[30] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via 1 minimization," *Proc. National Academy Sciences*, vol. 100, no. 5, pp. 2197–2202, Mar. 2003.

[31] J. Tiefeng, "Maxima of entries of haar distributed matrices." *Probability Theory & Related Fields*, vol. 131, no. 1, pp. 121–144, Jan. 2005.

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