Distributed Gain Matrix Optimization in Non-Regenerative MIMO Relay Networks

Raphael Rolny, Jörg Wagner, Celal Eşli, and Armin Wittneben
Communication Technology Laboratory, ETH Zurich
{rolny, wagner, cesli, wittneben}@nari.ee.ethz.ch

Abstract—**We consider a half-duplex multiple-input multiple-output relay network with multiple source and relay nodes and a set of collocated destination antennas. Each relay is equipped with multiple antennas and is constrained to map its receive signal linearly to its transmit signal. We devise a gradient based scheme that enables each relay to locally optimize its gain matrix with respect to the achievable sum-rate without requiring explicit knowledge of all fading coefficients characterizing the channel state. In particular, we demonstrate by numerical evidence that the overhead due to exchange of channel state information required by the scheme does either not scale at all or, if so, very slowly with the number of relay nodes or antennas therein.**

I. INTRODUCTION

The upcoming LTE Advanced standard [1] will be the first mobile telecommunication standard that allows for the use of relay stations (beyond simple repeaters) to assist communication between mobile devices and base stations. In this context, relays not only promise gains in terms of coverage, but due to more sophisticated signal processing might also turn out to be valuable means for enhancing efficiency of multiple-input multiple-output (MIMO) communication. For example, a mobile device in communication range of relay stations can exploit them in order to enforce well conditioned MIMO channel matrices even though the environment provides only little scattering. This potential of relay nodes to act as "active scatterers" has been identified in [2].

In this contribution, we consider a communication network that corresponds to a relay assisted MIMO uplink scenario (cf. Fig. 1). We study a non-regenerative orthogonal relaying protocol. More specifically, relay stations are assumed to determine their transmit signals as a linear map of their receive signals. The term "orthogonal" stems from the fact that the relay signals are separated from source signals either in time or frequency. This kind of relaying scheme has previously been used in [3], e.g., and is commonly termed as "generalized amplify-and-forward (AF) relaying." It is noteworthy in this context, that the modality of the operation of relay stations in LTE Advanced is fairly unspecified at the moment.

Coherent AF relaying is known to asymptotically (in the number of relay nodes) achieve the capacity of two-hop networks [4]. In networks with a finite number of relays, however, AF relaying is suboptimal in terms of the capacity. Nevertheless, it can have substantial implementation advantages. In particular, in a frequency division duplex implementation, AF relays have to perform an amplification and a frequency translation, which is easy to implement, power-efficient, and avoids coding delays. Moreover, AF relays are fully transparent to the modulation alphabet used by the terminal nodes.

An inherent drawback of non-regenerative relaying systems is the accumulation of relay and destination noise. In AF-based systems this accumulation arises through amplification and forwarding of the noise components of the receive signals at the relays. It is therefore important to efficiently exploit the knowledge of channel state information (CSI) to the extent it is available at the relay terminals in order to mitigate this effect. Likewise, such CSI can be exploited to further improve the condition of the MIMO channel matrix as seen between mobile devices and base station. A measure inherently capturing both of these effects is the sum-rate achieved by the mobile devices communicating to the base station through the set of relays. Accordingly, our work aims for optimizing relay gain matrices with respect to this quantity, focusing on generating only a moderate overhead in terms of CSI dissemination. Key enabler for our approach we believe to generate a CSI dissemination overhead that scales either not or very slowly with the number of relay stations/antennas, is the method of steepest descent. The underlying assumption that the number of required iterations of the optimization scheme scales not or, if so, very slowly with the number of relays is in accordance with numerical results. As iterative optimization methods in general, our approach proves particularly useful for channel tracking.

The use of gradient based methods is inspired by [5], which studies gain allocation in relay networks with distributed destination antennas and single-antenna relays. The gradient computation as well as the effects leading to the reduced CSI dissemination overhead, however, are fundamentally different in this contribution.

In the following, we identify mobile devices with source nodes, relay stations with relay nodes, and the base station with a destination node.
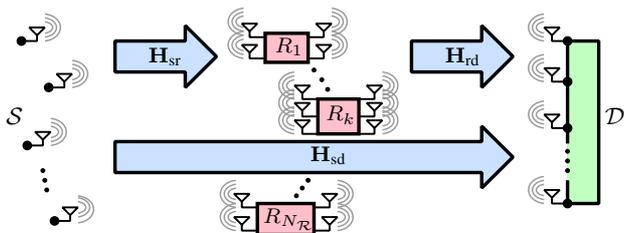

Fig. 1. Structure of the considered relay network.

II. I-O RELATION & ACHIEVABLE RATE

The half-duplex relay network under consideration consists of a destination terminal $\mathcal{D}$ with $n_\mathcal{D}$ fully cooperating antennas interested in messages sent by a set $\mathcal{S}$ of $n_\mathcal{S}$ distributed single-antenna source nodes. The direct link between $\mathcal{S}$ and $\mathcal{D}$ is assisted by an intermediate cluster $\mathcal{R}$ of $N_\mathcal{R}$ non-cooperating relays. The relay $R_k \in \mathcal{R}$ is equipped with n_k antennas, which corresponds to a total number of $n_\mathcal{R} = \sum_{k=1}^{N_\mathcal{R}} n_k$ relay antennas. Relay R_k performs a multiplication of its receive vector by a complex gain matrix $\mathbf{G}_k \in \mathbb{C}^{n_k \times n_k}$ depending on the channel state and retransmits it.

The matrices describing the channels between $\mathcal{S}$ and $\mathcal{D}$ (direct link), between $\mathcal{S}$ and $\mathcal{R}$, and between $\mathcal{R}$ and $\mathcal{D}$ are denoted by $\mathbf{H}_{sd} \in \mathbb{C}^{n_\mathcal{D} \times n_\mathcal{S}}$, $\mathbf{H}_{sr} \in \mathbb{C}^{n_\mathcal{R} \times n_\mathcal{S}}$, and $\mathbf{H}_{rd} \in \mathbb{C}^{n_\mathcal{D} \times n_\mathcal{R}}$, respectively. The channel matrices are assumed to be drawn from a continuous, non-degenerate distribution, the fading is assumed to be slow and frequency flat.

Transmission is divided into two time slots of equal length. The source nodes transmit their signals in a first time slot, the relays forward them in a second time slot to the destination. The input-output (I-O) relation over both time slots is written in terms of the transmit vector $\mathbf{s} \in \mathbb{C}^{n_\mathcal{S}}$, the relay noise vector $\mathbf{v} = [\mathbf{v}_1^T, \ldots, \mathbf{v}_{N_\mathcal{R}}^T]^T \in \mathbb{C}^{n_\mathcal{R}}$ over all relays ($\mathbf{v}_k \in \mathbb{C}^{n_k}$ corresponds to the noise of R_k), and the destination noise vector (over two time slots) $\mathbf{w} \in \mathbb{C}^{2n_\mathcal{D}}$:

$$\mathbf{y} = \begin{bmatrix} \mathbf{H}_{sd} \\ \mathbf{H}_{rd}\mathbf{G}\mathbf{H}_{sr} \end{bmatrix} \mathbf{s} + \begin{bmatrix} \mathbf{O}_{n_\mathcal{D} n_\mathcal{R}} \\ \mathbf{H}_{rd}\mathbf{G} \end{bmatrix} \mathbf{v} + \mathbf{w},$$

with $\mathbf{O}_{n_\mathcal{D} n_\mathcal{R}}$ the $n_\mathcal{D} \times n_\mathcal{R}$ all zero matrix and $\mathbf{G}$ the block-diagonal matrix containing the gain matrices $\mathbf{G}_1, \ldots, \mathbf{G}_{N_\mathcal{R}}$ on its diagonal. The elements of $\mathbf{v}$ and $\mathbf{w}$ are assumed to be i.i.d. $\mathcal{CN}(0, \sigma_v^2)$ and $\mathcal{CN}(0, \sigma_w^2)$, respectively.

We assume spatially white input at the source nodes, i.e. $\mathbb{E}\{\mathbf{s}\mathbf{s}^H\} = P_\mathcal{S}/n_\mathcal{S} \cdot \mathbf{I}_{n_\mathcal{S}}$, where $P_\mathcal{S}$ denotes the respective sum transmit power. This kind of signaling is reasonable in absence of transmit CSI, which is assumed. We impose a power constraint on each relay node:

$$P_k \leq P_\mathcal{R}/N_\mathcal{R}, \quad \forall k \in \{1, \ldots, N_\mathcal{R}\}, \tag{1}$$

where P_k is the transmit power of relay R_k. An achievable sum-rate is then given in terms of the receive signal and noise covariance matrices $\mathbf{K}_s$ and $\mathbf{K}_n$ by [6]

$$r = \frac{1}{2}\log\det(\mathbf{K}_n + \mathbf{K}_s) - \frac{1}{2}\log\det(\mathbf{K}_n) \quad \text{[nats/channel use]}. \tag{2}$$

The covariance matrices are given by

$$\mathbf{K}_s = \frac{P_\mathcal{S}}{n_\mathcal{S}} \begin{bmatrix} \mathbf{H}_{sd}\mathbf{H}_{sd}^H & \mathbf{H}_{sd}\mathbf{H}_{sr}^H\mathbf{G}^H\mathbf{H}_{rd}^H \\ \mathbf{H}_{rd}\mathbf{G}\mathbf{H}_{sr}\mathbf{H}_{sd}^H & \mathbf{H}_{rd}\mathbf{G}\mathbf{H}_{sr}\mathbf{H}_{sr}^H\mathbf{G}^H\mathbf{H}_{rd}^H \end{bmatrix} \tag{3}$$

and

$$\mathbf{K}_n = \begin{bmatrix} \sigma_w^2\mathbf{I}_{n_\mathcal{D}} & \mathbf{O}_{n_\mathcal{D} n_\mathcal{D}} \\ \mathbf{O}_{n_\mathcal{D} n_\mathcal{D}} & \sigma_w^2\mathbf{I}_{n_\mathcal{D}} + \sigma_v^2\mathbf{H}_{rd}\mathbf{G}\mathbf{G}^H\mathbf{H}_{rd}^H \end{bmatrix}. \tag{4}$$

They depend on the relay gain coefficients, which can be chosen to maximize the achievable sum-rate (2). The pre-log factor $\frac{1}{2}$ is due to the use of two time slots.

III. GRADIENT BASED OPTIMIZATION

In order to find a local optimum of (2) for a given channel realization, we propose a gradient based algorithm. For the sake of simplicity, we turn in first instance the constrained optimization problem into an unconstrained one by fulfilling the relay power constraints with equality through the substitution

$$\mathbf{G}_k = \sqrt{\frac{P_\mathcal{R}}{N_\mathcal{R}}} \tilde{\mathbf{G}}_k \Big/ \sqrt{\tilde{P}_k}, \quad \forall k \in \{1, \ldots, N_\mathcal{R}\}, \tag{5}$$

whereas

$$\tilde{P}_k = \text{Tr}\left(\frac{P_\mathcal{S}}{n_\mathcal{S}}\tilde{\mathbf{G}}_k\mathbf{H}_{sr,k}\mathbf{H}_{sr,k}^H\tilde{\mathbf{G}}_k^H + \sigma_v^2\tilde{\mathbf{G}}_k\tilde{\mathbf{G}}_k^H\right), \tag{6}$$

with $\mathbf{H}_{sr,k}$ the elements of $\mathbf{H}_{sr}$ that correspond to R_k. Fulfilling the power constraint with equality is suboptimal in general. A method to adapt the algorithm for optimization without this additional restriction is outlined in Section V of this paper.

A. Gradient Computation

The complex gradient of (2) with respect to $\tilde{\mathbf{G}}^*$, is defined as $\nabla_{\tilde{\mathbf{G}}^*} r \triangleq [\partial r/\partial \tilde{g}_{11}^* \ \ldots \ \partial r/\partial \tilde{g}_{n_\mathcal{R} n_\mathcal{R}}^*]^T$, with $(\cdot)^*$ denoting the complex conjugate. The gradient is obtained by the chain rule of differentiation as

$$\nabla_{\tilde{\mathbf{G}}^*} r = 2\left(\nabla_{\mathbf{G}^*} r\right)^T \cdot \hat{\mathbf{J}}_{\tilde{\mathbf{G}}^*} + 2\left(\nabla_{\mathbf{G}^*} r\right)^H \cdot \mathbf{J}_{\tilde{\mathbf{G}}^*}, \tag{7}$$

with $\mathbf{J}_{\tilde{\mathbf{G}}^*}$ and $\hat{\mathbf{J}}_{\tilde{\mathbf{G}}^*}$ the Jacobians corresponding to the change of variables (5) and its conjugate complex, respectively. The elements of $\nabla_{\mathbf{G}^*} r$, i.e. the elements of the gradient of r with respect to $\mathbf{G}^*$, are given by

$$\frac{\partial r}{\partial g_{ij}^*} = \frac{1}{2}\text{Tr}\left\{(\mathbf{K}_n + \mathbf{K}_s)^{-1}\left(\frac{\partial \mathbf{K}_n}{\partial g_{ij}^*} + \frac{\partial \mathbf{K}_s}{\partial g_{ij}^*}\right) - \mathbf{K}_n^{-1}\frac{\partial \mathbf{K}_n}{\partial g_{ij}^*}\right\}, \tag{8}$$

with

$$\frac{\partial \mathbf{K}_s}{\partial g_{ij}^*} = \frac{P_\mathcal{S}}{n_\mathcal{S}} \begin{bmatrix} \mathbf{O}_{n_\mathcal{D} n_\mathcal{D}} & \mathbf{H}_{sd}\mathbf{H}_{sr}^H\mathbf{E}_{(i,j)}^H\mathbf{H}_{rd}^H \\ \mathbf{O}_{n_\mathcal{D} n_\mathcal{D}} & \mathbf{H}_{rd}\mathbf{G}\mathbf{H}_{sr}\mathbf{H}_{sr}^H\mathbf{E}_{(i,j)}^H\mathbf{H}_{rd}^H \end{bmatrix} \tag{9}$$

and

$$\frac{\partial \mathbf{K}_n}{\partial g_{ij}^*} = \begin{bmatrix} \mathbf{O}_{n_\mathcal{D} n_\mathcal{D}} & \mathbf{O}_{n_\mathcal{D} n_\mathcal{D}} \\ \mathbf{O}_{n_\mathcal{D} n_\mathcal{D}} & \sigma_v^2\mathbf{H}_{rd}\mathbf{G}\,\mathbf{E}_{(i,j)}^H\mathbf{H}_{rd}^H \end{bmatrix}, \tag{10}$$

the partial derivatives of the signal and noise covariance matrices, respectively, with respect to g_{ij}^*. The matrix $\mathbf{E}_{(i,j)}$ is the single-entry matrix with all entries equal to zero except for the entry $\mathbf{E}_{(i,j)}[i,j]$, which is one.

The Jacobians $\mathbf{J}_{\tilde{\mathbf{G}}^*}$ and $\hat{\mathbf{J}}_{\tilde{\mathbf{G}}^*}$ contain the partial derivatives of (5) and of $\mathbf{G}_k^* = \sqrt{\frac{P_\mathcal{R}}{N_\mathcal{R}}}\tilde{\mathbf{G}}_k^* \big/ \sqrt{\tilde{P}_k}$, respectively, with respect to $\tilde{\mathbf{G}}^*$. The entries of $\mathbf{J}_{\tilde{\mathbf{G}}^*}$ are given by

$$\frac{\partial g_{i'j'}}{\partial \tilde{g}_{ij}^*} = -\frac{1}{2}\sqrt{\frac{P_\mathcal{R}}{N_\mathcal{R}}}\tilde{g}_{i'j'} \cdot \left(\tilde{P}_k\right)^{-\frac{1}{2}} \cdot \frac{\partial \tilde{P}_k}{\partial \tilde{g}_{ij}^*} \Big/ \tilde{P}_k$$

and those of $\hat{\mathbf{J}}_{\tilde{\mathbf{G}}^*}$ by

$$\frac{\partial g_{i'j'}^*}{\partial \tilde{g}_{ij}^*} = \xi_{(ij)}^{(i'j')} - \frac{1}{2}\sqrt{\frac{P_\mathcal{R}}{N_\mathcal{R}}}\tilde{g}_{i'j'}^* \cdot \left(\tilde{P}_k\right)^{-\frac{1}{2}} \cdot \frac{\partial \tilde{P}_k}{\partial \tilde{g}_{ij}^*} \Big/ \tilde{P}_k,$$

with

$$\xi_{(ij)}^{(i'j')} = \begin{cases} \sqrt{\frac{P_\mathcal{R}}{N_\mathcal{R}} \tilde{P}_k}, & \text{if} \quad i = i' \text{ and } j = j' \\ 0, & \text{otherwise.} \end{cases}$$

The inner derivative of $\tilde{P}_k$, given by (6), with respect to $\tilde{g}_{ij}^*$ is computed as

$$\frac{\partial \tilde{P}_k}{\partial \tilde{g}_{ij}^*} = \frac{P_\mathcal{S}}{n_\mathcal{S}} \tilde{\mathbf{G}}_k[i,:] \mathbf{H}_{\mathrm{sr},k} \mathbf{H}_{\mathrm{sr},k}^H \cdot \mathbf{e}_{(j)} + \sigma_v^2 \tilde{\mathbf{G}}_k[i,:] \cdot \mathbf{e}_{(j)},$$

where $\tilde{\mathbf{G}}_k[i,:]$ is the i-th row of $\tilde{\mathbf{G}}_k$ and $\mathbf{e}_{(j)}$ is the single entry vector which is zero everywhere except for the j-th entry, where it is one. Note that the Jacobians are block diagonal matrices, since the change of variables is applied to each node separately.

B. Optimization Algorithm

The steepest descent algorithm in its original form approaches a local optimum iteratively. The update equation of the gain coefficients is given by

$$\tilde{g}_{ij}^{(\ell+1)} = \tilde{g}_{ij}^{(\ell)} + \mu^{(\ell)} \cdot \boldsymbol{\Delta}_{ij}^{(\ell)}, \quad \forall i, j, \text{ and } \ell = 0, 1, 2, \dots, \quad (11)$$

where the search direction

$$\boldsymbol{\Delta}_{ij}^{(\ell)} = \left(\nabla_{\tilde{\mathbf{G}}^*} r(\tilde{\mathbf{G}}^{(\ell)}) \right)_{ij}$$

corresponds to the gradient of r evaluated at $\tilde{\mathbf{G}}^{(\ell)}$, and $\mu^{(\ell)}$ is the step-size of the actual iteration cycle. We optimize $\mu^{(\ell)}$ in each iteration cycle by a heuristic line search method that aims to achieve low complexity. Starting with some small initial step-size $\mu_0 = \mu_{\min}$, the line search increases the step-size according to $\mu_{m+1} = \alpha \cdot \mu_m$, for some $\alpha > 1$, as long as

$$r\left(\tilde{\mathbf{G}}^{(\ell)} + \mu_{m+1} \cdot \boldsymbol{\Delta}^{(\ell)} \right) > r\left(\tilde{\mathbf{G}}^{(\ell)} + \mu_m \cdot \boldsymbol{\Delta}^{(\ell)} \right).$$

As soon as μ_{m+1} leads to an achievable sum-rate smaller than the sum-rate for the previous step-size, the line search terminates and chooses $\mu^{(\ell)} = \mu_m$ for the ℓ-th iteration cycle of the optimization in (11). With this heuristic approach, we achieve a fast increase of the step-sizes in the beginning of the optimization and small step-sizes when the gain coefficients are close to the (local) optimum.

We use a refined version of the optimization algorithm by utilizing conjugate gradients (CG) which yield a faster convergence behavior as compared to the conventional steepest descent method. To this end, the search directions are modified according to

$$\boldsymbol{\Delta}^{(\ell)} = \nabla_{\tilde{\mathbf{G}}^*} r(\tilde{\mathbf{G}}^{(\ell)}) + \beta_{\mathrm{cg}}^{(\ell)} \cdot \boldsymbol{\Delta}^{(\ell-1)}, \quad \ell = 0, 1, 2, \dots.$$

The parameter $\beta_{\mathrm{cg}}^{(\ell)}$ is a design parameter. We choose

$$\beta_{\mathrm{cg}}^{(\ell)} = \frac{\nabla_{\tilde{\mathbf{G}}^*} r(\tilde{\mathbf{G}}^{(\ell)})^H \cdot \left(\nabla_{\tilde{\mathbf{G}}^*} r(\tilde{\mathbf{G}}^{(\ell)}) - \nabla_{\tilde{\mathbf{G}}^*} r(\tilde{\mathbf{G}}^{(\ell-1)}) \right)}{\nabla_{\tilde{\mathbf{G}}^*} r(\tilde{\mathbf{G}}^{(\ell-1)})^H \cdot \nabla_{\tilde{\mathbf{G}}^*} r(\tilde{\mathbf{G}}^{(\ell-1)})} \quad (12)$$

according to Polak-Ribière [7].

IV. DISTRIBUTED OPTIMIZATION

The optimization scheme discussed in Section III can be distributed among the relays without disseminating global CSI. In particular, we show that each relay can compute its gain matrix based on CSI that is locally available and limited feedback from the destination.

A. Gradient Computation with Limited CSI

In the proposed optimization scheme, each relay R_k needs to compute its own gain coefficients only. Therefore, R_k is only required to compute its own components of the search direction. In the following, we refer to local CSI at R_k as the channel coefficients that can be estimated locally at R_k, i.e. the channel coefficients from all sources to R_k and the channel coefficients form R_k to all destination antennas.

Proposition: *Each relay $R_k \in \mathcal{R}$ can locally compute its corresponding components $(\nabla_{\tilde{\mathbf{G}}^*} r)_k$ of the gradient of the achievable sum-rate based on limited CSI, which consists of*

- *the a-priori known quantities $P_\mathcal{S}$, $P_\mathcal{R}$, σ_v^2,*
- *its local CSI,*
- *the matrices $\mathbf{H}_{\mathrm{sd}}$, $\mathbf{H}_{\mathrm{rd}} \mathbf{G} \mathbf{H}_{\mathrm{sr}}$, and $\mathbf{K}_n$ (fed back from $\mathcal{D}$).*

Proof: The overall gradient (7) consists of the "outer" gradient $\nabla_{\mathbf{G}^*} r$ as well as the Jacobi matrices $\mathbf{J}_{\tilde{\mathbf{G}}^*}$ and $\hat{\mathbf{J}}_{\tilde{\mathbf{G}}^*}$. We focus on the "outer" gradient and the Jacobians separately, before we complete the proof by combining these components to the overall gradient.

First, we show that R_k can locally compute its corresponding elements of $\nabla_{\mathbf{G}^*} r$. These elements are given by (8) for all g_{ij} belonging to $\mathbf{G}_k$. The signal covariance matrix $\mathbf{K}_s$ can be computed locally, since it contains only products of $\mathbf{H}_{\mathrm{sd}}$ and $\mathbf{H}_{\mathrm{rd}} \mathbf{G} \mathbf{H}_{\mathrm{sr}}$ or their Hermitians as well as the a-priori known $P_\mathcal{S}$. The value $n_\mathcal{S}$ can be deduced from the number of columns of $\mathbf{H}_{\mathrm{sd}}$. The noise covariance matrix $\mathbf{K}_n$ is explicitly available in the limited CSI. The derivatives of the covariance matrices are given in (9) and (10). In first instance, we study the following components $\mathbf{H}_{\mathrm{sr}}^H \mathbf{E}_{(i,j)}^H \mathbf{H}_{\mathrm{rd}}^H$ and $\mathbf{H}_{\mathrm{rd}} \mathbf{G} \mathbf{E}_{(i,j)}^H \mathbf{H}_{\mathrm{rd}}^H$. In order to compute them, local CSI is sufficient:

$$\mathbf{H}_{\mathrm{sr}}^H \mathbf{E}_{(i,j)}^H \mathbf{H}_{\mathrm{rd}}^H = \left(\mathbf{H}_{\mathrm{rd}} \mathbf{E}_{(i,j)} \mathbf{H}_{\mathrm{sr}} \right)^H$$
$$= \left(\mathbf{H}_{\mathrm{rd}}[\mathcal{K}_\mathcal{D}, \{i\}] \cdot \mathbf{H}_{\mathrm{sr}}[\{j\}, \mathcal{K}_\mathcal{S}] \right)^H,$$

with $\mathcal{K}_\mathcal{D} = \{1, \dots, n_D\}$ and $\mathcal{K}_\mathcal{S} = \{1, \dots, n_\mathcal{S}\}$, since $\mathbf{E}_{(i,j)}$ contains only a single one, while all other entries are zero. Both $\mathbf{H}_{\mathrm{rd}}[\mathcal{K}_\mathcal{D}, \{i\}]$ and $\mathbf{H}_{\mathrm{sr}}[\{j\}, \mathcal{K}_\mathcal{S}]$ contain only elements that can be estimated locally at R_k. Due to the block-diagonal structure of $\mathbf{G}$, also the computation of $\mathbf{H}_{\mathrm{rd}} \mathbf{G} \mathbf{E}_{(i,j)}^H \mathbf{H}_{\mathrm{rd}}^H$ involves only channel coefficients that are locally available:

$$\mathbf{H}_{\mathrm{rd}} \mathbf{G} \mathbf{E}_{(i,j)}^H \mathbf{H}_{\mathrm{rd}}^H = \left(\mathbf{H}_{\mathrm{rd}} \mathbf{E}_{(i,j)} \mathbf{G}^H \mathbf{H}_{\mathrm{rd}}^H \right)^H$$
$$= \left(\mathbf{H}_{\mathrm{rd}} \mathbf{E}_{(i,j)} \underbrace{\begin{bmatrix} \mathbf{G}_1^H \left(\mathbf{H}_{\mathrm{rd}}[\mathcal{K}_\mathcal{D}, \mathcal{K}_1] \right)^H \\ \vdots \\ \mathbf{G}_k^H \left(\mathbf{H}_{\mathrm{rd}}[\mathcal{K}_\mathcal{D}, \mathcal{K}_k] \right)^H \\ \vdots \\ \mathbf{G}_{N_\mathcal{R}}^H \left(\mathbf{H}_{\mathrm{rd}}[\mathcal{K}_\mathcal{D}, \mathcal{K}_{N_\mathcal{R}}] \right)^H \end{bmatrix}}_{\triangleq \mathbf{X}} \right)^H$$
$$= \left(\mathbf{H}_{\mathrm{rd}}[\mathcal{K}_\mathcal{D}, \{i\}] \cdot \mathbf{X}[\{j\}, \mathcal{K}_\mathcal{D}] \right)^H,$$

with $\mathcal{K}_k$ the index set corresponding to the antennas of R_k. The single-entry matrix $\mathbf{E}_{(i,j)}$ selects a single row of $\mathbf{X}$ that

contains only channel coefficients that are locally available at R_k. Therefore, we can conclude that limited CSI available at R_k is sufficient to compute $\nabla_{\mathbf{G}^*} r$.

The Jacobians are block-diagonal, since the change of variables is applied to each relay separately. Each R_k can locally compute its corresponding block $(\mathbf{J}_{\tilde{\mathbf{G}}^*})_k$ and $(\hat{\mathbf{J}}_{\tilde{\mathbf{G}}^*})_k$, because the change of variables depends only on the gain coefficients $\mathbf{G}_k$, the a-priori known quantities $P_{\mathcal{S}}$, σ_v^2, and the channel coefficients $\mathbf{H}_{\text{sr},k}$ which can be estimated at R_k.

Due to the block-diagonal structure of the Jacobians, we conclude that the components of the gradient that belong to R_k can be written as

$$(\nabla_{\tilde{\mathbf{G}}^*} r)_k = \left((\nabla_{\mathbf{G}^*} r)^T \cdot \hat{\mathbf{J}}_{\tilde{\mathbf{G}}^*} + (\nabla_{\mathbf{G}^*} r)^H \cdot \mathbf{J}_{\tilde{\mathbf{G}}^*} \right)_k^T$$
$$= \left[(\nabla_{\mathbf{G}^*} r)_k^T \cdot \left(\hat{\mathbf{J}}_{\tilde{\mathbf{G}}^*} \right)_k + (\nabla_{\mathbf{G}^*} r)_k^H \cdot (\mathbf{J}_{\tilde{\mathbf{G}}})_k \right]^T,$$

where $(\nabla_{\mathbf{G}^*} r)_k$, $(\hat{\mathbf{J}}_{\tilde{\mathbf{G}}^*})_k$, and $(\mathbf{J}_{\tilde{\mathbf{G}}^*})_k$ can be computed with limited CSI. Therefore, the assumed limited CSI available at R_k is sufficient to compute the k-th block of the overall gradient (7). ∎

B. Distributed Gain Allocation

In the first optimization cycle, each relay $R_k \in \mathcal{R}$ initializes its gain matrix $\mathbf{G}_k$ with random values fulfilling the power constraint and computes the corresponding components of the update vector. In the first iteration cycle, R_k updates $\mathbf{G}_k$ with a step in the direction of $(\nabla_{\tilde{\mathbf{G}}^*} r)_k$. The direction can be computed locally based on limited CSI as shown in the proposition.

If the conjugate gradient (CG) algorithm is used, the following iteration cycles require the computation of β_{cg} according to (12). This involves the computation of inner products that can be obtained in a distributed manner by using *over-the-air addition* (cf. [5]). Thereby, each R_k computes a value

$$\gamma_k^{(\ell)} = \frac{\nabla_{\tilde{\mathbf{G}}^*} r(\tilde{\mathbf{G}}^{(\ell)})_k^H \cdot \left(\nabla_{\tilde{\mathbf{G}}^*} r(\tilde{\mathbf{G}}^{(\ell)})_k - \nabla_{\tilde{\mathbf{G}}^*} r(\tilde{\mathbf{G}}^{(\ell-1)})_k \right)}{\mathbf{H}_{\text{rd}}[p_k, q]}$$

with $\mathbf{H}_{\text{rd}}[p_k, q]$ denoting the channel coefficient from an arbitrary antenna p_k belonging to R_k to an arbitrary destination antenna q. When all relays simultaneously transmit analog values corresponding to $\gamma_k^{(\ell)}$ in equivalent baseband, the receive signal at destination antenna q becomes

$$d_q^{(\ell)} = \sum_{k=1}^{N_{\mathcal{R}}} \mathbf{H}_{\text{rd}}[p_k, q] \cdot \gamma_k^{(\ell)} =$$
$$= \sum_{k=1}^{N_{\mathcal{R}}} \nabla_{\tilde{\mathbf{G}}^*} r(\tilde{\mathbf{G}}^{(\ell)})_k^H \cdot \left(\nabla_{\tilde{\mathbf{G}}^*} r(\tilde{\mathbf{G}}^{(\ell)})_k - \nabla_{\tilde{\mathbf{G}}^*} r(\tilde{\mathbf{G}}^{(\ell-1)})_k \right)$$
$$= \nabla_{\tilde{\mathbf{G}}^*} r(\tilde{\mathbf{G}}^{(\ell)})^H \cdot \left(\nabla_{\tilde{\mathbf{G}}^*} r(\tilde{\mathbf{G}}^{(\ell)}) - \nabla_{\tilde{\mathbf{G}}^*} r(\tilde{\mathbf{G}}^{(\ell-1)}) \right),$$

which is (by neglecting the noise) equivalent to the numerator of β_{cg}. In order to perform an over-the-air addition to compute the denominator of β_{cg}, each relay R_k transmits another value in equivalent baseband:

$$\delta_k^{(\ell)} = \frac{\nabla_{\tilde{\mathbf{G}}^*} r(\tilde{\mathbf{G}}^{(\ell-1)})_k^H \cdot \nabla_{\tilde{\mathbf{G}}^*} r(\tilde{\mathbf{G}}^{(\ell-1)})_k}{\mathbf{H}_{\text{rd}}[p_k, q]}.$$

This leads to a noisy version of the denominator of β_{cg} at destination antenna q. The destination estimates the receive signal of both transmissions, computes the fraction (12), and broadcasts the resulting β_{cg} back to the relays. Note that in such a *measurement cycle*, all relays transmit simultaneously with a single antenna to a single destination antenna.

Once the search direction of the ℓ-th iteration cycle is known, the relays have to choose an appropriate step size $\mu^{(\ell)}$. To this end, the line search procedure is applied in a distributed manner. Each relay updates its gain coefficients with an initial step size $\mu_{\min}$, which is a-priori known to all relay nodes. Then, the relays scale their gain coefficients according to the power constraint and transmit pilot sequences such that $\mathcal{D}$ can estimate its achievable sum-rate r. This procedure is repeated for increasing step sizes μ_m, $m = 0, 1, 2, \ldots$, as explained in Section III. As soon as $\mathcal{D}$ detects a sum-rate that is smaller than in the previous line search iteration, a stop bit is broadcasted to all relays. The relays then terminate the line search and use $\mu^{(\ell)} = \mu_m$ for the update of their gain coefficients.

C. Overhead

The knowledge of $\mathbf{H}_{\text{sd}}$, $\mathbf{H}_{\text{rd}}\mathbf{G}\mathbf{H}_{\text{sr}}$, and $\mathbf{K}_n$ is required by all relays for the computation of the search directions. Estimates of these matrices can be obtained at $\mathcal{D}$ and fed back to the relays. The channel matrices $\mathbf{H}_{\text{sd}}$ and $\mathbf{H}_{\text{rd}}\mathbf{G}\mathbf{H}_{\text{sr}}$ can be estimated by the use of training sequences transmitted by the source antennas. These pilot sequences traverse the network through the direct channel and the two-hop channel via the relays. The pilot sequences from $\mathcal{S}$ and additional pilot sequences transmitted simultaneously by $\mathcal{D}$ can be used for estimates of the local CSI in the relays. In order to estimate $\mathbf{K}_n$, all relays can simultaneously transmit pseudo random sample sequences that can be averaged to an estimate of this matrix in $\mathcal{D}$. The direct link $\mathbf{H}_{\text{sd}}$ is independent of $\mathbf{G}$ and has to be broadcasted once at the beginning of the optimization only. The matrices $\mathbf{H}_{\text{rd}}\mathbf{G}\mathbf{H}_{\text{sr}}$ and $\mathbf{K}_n$, however, depend on the relay gain matrices and thus need to be estimated and fed back in every iteration cycle. The overhead caused by the CSI dissemination in each iteration cycle is independent of the number of relay antennas $n_{\mathcal{R}}$, since all matrices $\mathbf{H}_{\text{sd}}$, $\mathbf{H}_{\text{rd}}\mathbf{G}\mathbf{H}_{\text{sr}}$, and $\mathbf{K}_n$ are either $n_{\mathcal{D}} \times n_{\mathcal{S}}$ or $n_{\mathcal{D}} \times n_{\mathcal{D}}$ matrices. For a fixed number of source and destination antennas, the dimensions of these matrices remain constant as $n_{\mathcal{R}}$ grows large. Numerical results show that also the number of iterations required to achieve close to optimal sum-rates does not scale at all or, if so, very slowly with the number of relays. Accordingly, our gain allocation scheme proves to be particularly useful for large $n_{\mathcal{R}}$.

V. INEQUALITY CONSTRAINT

In the previous sections, the transmit power constraint imposed on the relays has been fulfilled with equality by applying the change of variables (5). This is suboptimal in general. A relay node that receives a weak signal due to small channel coefficients from the sources, retransmits mainly noise and little desired signal. Forcing a relay to transmit at a fixed

power can, in this case, possibly reduce the achievable sum-rate. Therefore, we modify the change of variables to

$$\mathbf{G}_k = \begin{cases} \sqrt{\frac{P_{\mathcal{R}}}{N_{\mathcal{R}}}} \cdot \tilde{\mathbf{G}}_k \big/ \tilde{P}_k, & \text{if } \tilde{\mathbf{G}}_k \notin \Omega_k \\ \tilde{\mathbf{G}}_k, & \text{if } \tilde{\mathbf{G}}_k \in \Omega_k \end{cases}, \quad \forall k \in \{1, \dots, N_{\mathcal{R}}\},$$

where $\Omega_k \triangleq \{\tilde{\mathbf{G}}_k : \tilde{P}_k \leq P_{\mathcal{R}}/N_{\mathcal{R}}\}$ is defined for each relay as the region of $\tilde{\mathbf{G}}_k$ in which the power constraint is fulfilled. After each update operation, the optimization scheme has to distinguish different cases:

- If the matrix $\tilde{\mathbf{G}}_k^{(\ell)}$ in the ℓ-th iteration cycle is in Ω_k, then the gain coefficients are not scaled. The gradient of the next iteration cycle is then accordingly computed with the Jacobians $(\hat{\mathbf{J}}_{\tilde{\mathbf{G}}^*})_k = \mathbf{I}_{n_k}$ and $(\mathbf{J}_{\tilde{\mathbf{G}}^*})_k = \mathbf{O}_{n_k n_k}$. If the power constraint is violated during the line search, the resulting gain coefficients are scaled to fulfill the power constraint with equality.
- If $\tilde{\mathbf{G}}_k^{(\ell)}$ is not in Ω_k, the gain coefficients are scaled to fulfill the power constraint with equality. For the next iteration cycle, two different search directions need to be considered:
  - If $\left(\nabla_{\mathbf{G}^*} r(\mathbf{G}^{(\ell)})\right)_k$ points *into* Ω_k, this search direction is used.
  - If $\left(\nabla_{\mathbf{G}^*} r(\mathbf{G}^{(\ell)})\right)_k$ points *outwards* Ω_k, then the gradient $\left(\nabla_{\tilde{\mathbf{G}}^*} r(\tilde{\mathbf{G}}^{(\ell)})\right)_k$ with the Jacobians according to Section III is used.

Each relay chooses its search direction individually.

VI. Simulation Results

We assess the achievable sum-rate of the proposed gain allocation scheme by means of computer simulations. We choose networks with $n_{\mathcal{S}} = n_{\mathcal{D}} = 4$ and varying relay configurations. We fix the source power to $P_{\mathcal{S}} = 10^2$ and the noise variances to $\sigma_v^2 = \sigma_w^2 = 1$ and optimize for the inequality constraint $P_k \leq \frac{P_{\mathcal{R}}}{N_{\mathcal{R}}}$ for $P_{\mathcal{R}} = 10^2$. The elements of $\mathbf{H}_{\text{sr}}$, and $\mathbf{H}_{\text{rd}}$ are assumed to be i.i.d $\mathcal{CN}(0, 1)$, and those of $\mathbf{H}_{\text{sd}}$ are assumed to be i.i.d. $\mathcal{CN}(0, \sigma_{\text{sd}}^2)$ with $\sigma_{\text{sd}}^2 \in \{0, 1\}$.

Fig. 2 compares the convergence behavior for different cooperation restrictions at the relays with $n_{\mathcal{R}} = 8$ relay antennas in total, once with a strong direct link ($\sigma_{\text{sd}}^2 = 1$) and once with a blocked direct link ($\sigma_{\text{sd}}^2 = 0$). We observe that the sum-rate performance improves when the number of antennas per relay, and thus the cooperation, increases (as expected). Room for optimization is reduced in the case $\sigma_{\text{sd}}^2 = 1$, since $\mathbf{H}_{\text{sd}}$ is independent of $\mathbf{G}$.

Fig. 3 shows the empirical cumulative distribution function (CDF) of r for networks with $N_{\mathcal{R}} = 4, 8, 16, 32$ with two antennas per relay, $\sigma_{\text{sd}}^2 = 0$, and varying number of iterations. We observe that a moderate number of iterations achieves close to optimum rates. The number of iterations required to achieve close to optimal solutions scales not or very slowly with the number of relays.

VII. Conclusion

In this work, we have identified a gradient based algorithm as an efficient means to optimize the achievable sum-rate in
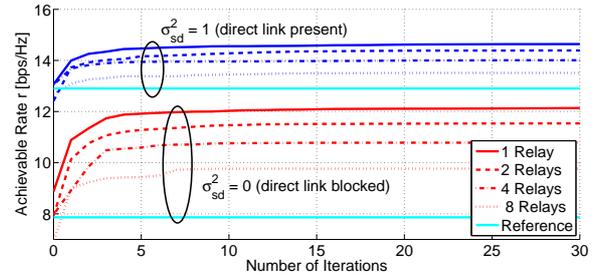

Fig. 2. Typical convergence behavior of the sum-rate in networks with four source/destination antennas and eight relay antennas. References are obtained by averaging over random relay phase shifts.
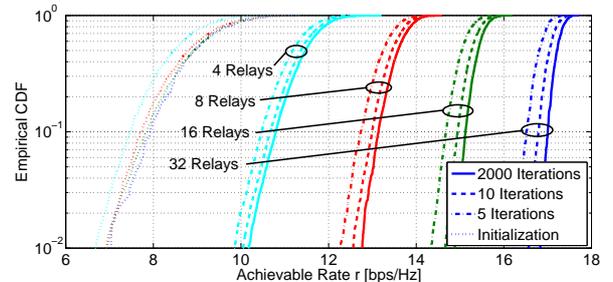

Fig. 3. Empirical CDF of the achievable sum-rate with initial gain coefficients and after 5, 10, and 2000 iterations with 4, 8, 16, and 32 relays.

MIMO relay networks. Even though we do not claim to end up in a global optimum, we observe that the optimization with different initializations converges to the same solution for fixed channel coefficients. In small networks with few nodes, a centralized scheme could find optimal gain coefficients with less overhead, even if global CSI needs to be disseminated. Therefore, the proposed scheme is more efficient in large networks with many relays involved. The scheme is also useful for channel tracking or dynamic environments with relays that arbitrarily enter or leave the network. In such a case, the distributed scheme can continue without any precautions, since each relay considers its own contribution to the search direction only and can compute its update vector independently of other relays. This is not possible in a centralized scheme, since global CSI needs to be disseminated whenever the network topology changes.

References

[1] 3GPP, Technical Specification Group Radio Access Network, "Further advancements for E-UTRA: Physical layer aspects (release 9)," Tech. Rep. 3GPP TR 36.814 V0.4.1 (2009-02), Feb. 2009.

[2] B. Rankov and A. Wittneben, "On the capacity of relay-assisted wireless MIMO channels," in *Signal Processing Advances in Wireless Communications, SPAWC*, July 2004.

[3] O. Munoz, J. Vidal, and A. Agustin, "Non-regenerative MIMO relaying with channel state information," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, March 2005, pp. 361–364.

[4] H. Bölcskei, R. U. Nabar, O. Oyman, and A. J. Paulraj, "Capacity scaling laws in MIMO relay networks," *IEEE Trans. Wireless Communications*, vol. 5, no. 6, pp. 1433–1444, June 2006.

[5] C. Eşli, J. Wagner, and A. Wittneben, "Distributed gradient based gain allocation for coherent multiuser multihop AF relaying networks," in *IEEE International Conference on Communications*, Jun. 2009.

[6] E. Telatar, "Capacity of multi-antenna gaussian channels," *European Transactions on Telecommunications*, vol. 10, pp. 585–595, 1999.

[7] J. R. Shewchuk, "An introduction to the conjugate gradient method without the agonizing pain," Carnegie Mellon University, Pittsburgh, PA, USA, Tech. Rep., 1994.