CDI Rate-balancing with Per-base-station Constraints

David Neumann, Andreas Gründinger, Michael Joham, and Wolfgang Utschick
Associate Institute for Signal Processing, Technische Universität München, 80290 Munich, Germany
{d.neumann,gruendinger,joham,utschick}@tum.de

***Abstract*—We previously introduced a rate-balancing precoding technique based on channel distribution information, which significantly reduces the impact of pilot-contamination in the downlink of a massive MIMO system. We now extend the rate-balancing optimization problem for the precoders to multiple power constraints, to be able to apply the technique in a coordinated multi-point setting. For this problem, we derive an efficient and adaptive method to find a globally optimal solution based on transformation in the dual uplink. Finally, the results are compared to the optimization with a single sum-power constraint, where the beamforming vectors are simply scaled to satisfy the constraints at each base station.**

I. Introduction

Massive MIMO is currently drawing a lot of interest from both, academia and industry, due to the promise of full multiplexing gains with simple linear signal processing methods. This gain, however, is limited by the dimensionality bottleneck imposed by the fixed coherence interval of the channel [1]–[3].

Not all is lost, though, since the well-known results on the degrees of freedom achievable in a fixed coherence interval only consider independently and identically distributed channel coefficients [1]. In fact, structure of the channel vectors in form of second-order information can be exploited to break out of the dimensionality bottleneck [2], [4], [5].

In [4], we introduced a precoding method based on second-order information, which is a generalization of the pilot-contamination-precoding in [5]. In [6], a rate-balancing formulation based on the statistical precoding approach is optimized under a single sum-power constraint. Per base-station constraints, however, are more sensible in a coordinated multi-point setting. The additional constraints lead to challenges in the optimization, which we discuss herein.

The considered rate-balancing problem is similar to the one introduced in [7] for a cell-free system. However, our formulation is more general in that it exploits arbitrary correlations of the channel coefficients of a user. In contrast to [7], we solve the problem in the dual domain, which significantly reduces the complexity. The proposed iterative method is also applicable in an adaptive manner, i.e., in practice, we can perform a single iteration of the algorithm each time the covariance matrices are updated.

We demonstrate with numerical simulations, that significant gains are possible by explicitly taking the per-base-station constraints into consideration as compared to a simple scaling of the optimal beamformers with a single sum-power constraint.

II. System Model

We consider a cellular network of L cells where each base station is equipped with M antennas. The vector channel of user k to the base station in cell ℓ is given by

$$\boldsymbol{h}_{\ell k} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{0}, \boldsymbol{C}_{\ell k}) \in \mathbb{C}^M.$$

In the coordinated multi-point case, the user k is served from all base stations $\ell = 1, \ldots, L$ simultaneously. Thus, we simply consider the compound channel vectors

$$\boldsymbol{h}_k = \begin{bmatrix} \boldsymbol{h}_{1k} \\ \vdots \\ \boldsymbol{h}_{Lk} \end{bmatrix} \sim \mathcal{N}_{\mathbb{C}} \left(\boldsymbol{0}, \boldsymbol{C}_k = \begin{bmatrix} \boldsymbol{C}_{1k} & & \\ & \ddots & \\ & & \boldsymbol{C}_{Lk} \end{bmatrix} \right).$$

We assume that in each channel coherence interval all users in the network transmit one of T_{tr} orthonormal pilot sequences with the effective training SNR ρ_{tr}. Thus, for each user we can acquire a least-squares estimate

$$\hat{\boldsymbol{h}}_k = \boldsymbol{h}_k + \sum_{n \in \mathcal{I}_k} \boldsymbol{h}_n + \frac{1}{\sqrt{\rho_{\text{tr}}}} \boldsymbol{w}_k$$

where $\mathcal{I}_k$ is the set of users in the network, which transmit the same pilot sequence as user k and $\boldsymbol{w}_k$ is additive white Gaussian noise with i.i.d. zero-mean and unit-variance entries.

The idea behind channel-density-information (CDI) precoding is to apply a deterministic transformation to the matched filter estimate $\hat{\boldsymbol{h}}_k$, i.e., the precoding vectors are calculated as

$$\boldsymbol{t}_k = \boldsymbol{A}_k \hat{\boldsymbol{h}}_k.$$

Note again, that the transformation $\boldsymbol{A}_k$ is deterministic and only depends on the channel covariance matrices.

For CDI precoding, a lower bound on the achievable rate is given by [8]

$$r_k = \log_2(1 + \gamma_k)$$

with the effective SINR

$$\gamma_k = \frac{\left|\mathrm{E}[\boldsymbol{h}_k^{\mathrm{H}} \boldsymbol{t}_k]\right|^2}{1 + \mathrm{var}[\left|\boldsymbol{h}_k^{\mathrm{H}} \boldsymbol{t}_k\right|] + \sum_n \mathrm{E}[\left|\boldsymbol{h}_k^{\mathrm{H}} \boldsymbol{t}_n\right|^2]}.$$

Since for zero-mean $\boldsymbol{x}$ and $\boldsymbol{y}$

$$\mathrm{E}[\left|\boldsymbol{x}^{\mathrm{H}} \boldsymbol{y}\right|^2] = \mathrm{tr}(\mathrm{E}[\boldsymbol{x}\boldsymbol{x}^{\mathrm{H}}]\,\mathrm{E}[\boldsymbol{y}\boldsymbol{y}^{\mathrm{H}}]) + \left|\mathrm{tr}(\mathrm{E}[\boldsymbol{y}\boldsymbol{x}^{\mathrm{H}}])\right|^2$$

we get [4]

$$\gamma_k = \frac{\left|\mathrm{tr}(\boldsymbol{A}_k \boldsymbol{C}_k)\right|^2}{1 + \sum_n \mathrm{tr}(\boldsymbol{C}_k \boldsymbol{A}_n \hat{\boldsymbol{C}}_n \boldsymbol{A}_n^{\mathrm{H}}) + \sum_{n \in \mathcal{I}_k} \left|\mathrm{tr}(\boldsymbol{A}_n \boldsymbol{C}_k)\right|^2}$$

where

$$\hat{\boldsymbol{C}}_k = \mathrm{E}[\hat{\boldsymbol{h}}_k \hat{\boldsymbol{h}}_k^{\mathrm{H}}].$$

Note that the last summand in the denominator is due to interference during the training phase, so-called pilot-contamination.

Since the covariance matrices $\boldsymbol{C}_k$ and $\hat{\boldsymbol{C}}_k$ are block-diagonal, the optimal transformations $\boldsymbol{A}_k$ have the same structure. That is, we have

$$\boldsymbol{A}_k = \begin{bmatrix} \boldsymbol{A}_{1k} & & \\ & \ddots & \\ & & \boldsymbol{A}_{Lk} \end{bmatrix}.$$

If we further assume, that for each base station, the covariance matrices are jointly diagonalized by some unitary transformation, e.g., by transforming in the angular domain of a uniform linear array [9], [10], we can work with the diagonal covariance matrices. Thus, also the optimal transformations $\boldsymbol{A}_k$ are diagonal (see Appendix) and the number of parameters reduces from M^2L to ML.

In general, by vectorization, the downlink SINRs can be written as [4]

$$\gamma_k(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_K) = \frac{\boldsymbol{a}_k^{\mathrm{H}} \boldsymbol{c}_k \boldsymbol{c}_k^{\mathrm{H}} \boldsymbol{a}_k}{1 + \sum_n \boldsymbol{a}_n^{\mathrm{H}} \boldsymbol{B}_{kn} \boldsymbol{a}_n} \tag{1}$$

where $\boldsymbol{a}$ has $(ML)^2$ elements for general covariance matrices $\boldsymbol{C}_k$, M^2L elements if we assume block-diagonal covariance matrices and ML elements if the blocks can be approximately jointly diagonalized. Correspondingly, the vectors $\boldsymbol{c}_k$ contain the stacked columns of $\boldsymbol{C}_k$, the stacked columns of all blocks $\boldsymbol{C}_{\ell k}$ or the elements of the diagonal covariance matrices, respectively. Due to the impractical complexity for the other cases, the simulation results focus on jointly diagonalizable channel covariance matrices.

III. Problem Formulation

In our previous work, we demonstrated how the deterministic transformations $\boldsymbol{A}_k$ can be chosen to suppress pilot-contamination [4] or provide a rate-balanced solution [6] in a single-cell scenario. In the following, we extend the rate-balancing formulation to the multi-cell case with per-base-station power constraints and provide an algorithm to find the optimal solution.

The average power constraint for base station ℓ is given by

$$\sum_k \mathrm{tr}(\boldsymbol{A}_{\ell k} \hat{\boldsymbol{C}}_{\ell k} \boldsymbol{A}_{\ell k}^{\mathrm{H}}) \leq 1. \tag{2}$$

Analogously to (1), we can rewrite the power constraint in (2) as

$$\sum_k \boldsymbol{a}_k^{\mathrm{H}} \boldsymbol{Q}_{\ell k} \boldsymbol{a}_k \leq 1$$

with a positive semi-definite $\boldsymbol{Q}_{\ell k}$. Consequently, the rate-balancing problem is stated as

$$\max_{\boldsymbol{a}_1,\ldots,\boldsymbol{a}_K} r_0 \quad \text{s.t. } \begin{array}{l} r_0 \tau_k \leq r_k(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_K) \;\forall k \\ \sum_k \boldsymbol{a}_k^{\mathrm{H}} \boldsymbol{Q}_{\ell k} \boldsymbol{a}_k \leq 1 \;\forall \ell \end{array} \tag{3}$$

where τ_k denotes the rate factor for user k.

Instead of solving (3) directly, we solve a dual uplink problem (cf. [6], [11], [12]). To derive the uplink problem, we first consider the downlink power-minimization problem, which is very similar in structure to the rate-balancing problem, but has the advantage of being convex. The power minimization problem is given by

$$\min_{\boldsymbol{a}_1,\ldots,\boldsymbol{a}_K} \beta \quad \text{s.t. } \begin{array}{l} \bar{r}_k \leq r_k(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_K) \;\forall k \\ \sum_k \boldsymbol{a}_k^{\mathrm{H}} \boldsymbol{Q}_{\ell k} \boldsymbol{a}_k \leq \beta \;\forall \ell. \end{array} \tag{4}$$

That is, we minimize the power factor β under fixed rate targets $\bar{r}_k$. We rewrite the rate constraints as (cf. (II) and (1))

$$\begin{aligned} \bar{r}_k &\leq r_k(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_K) \\ \Leftrightarrow 2^{\bar{r}_k} - 1 &\leq \gamma_k(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_K) \\ \Leftrightarrow 1 + \sum_n \boldsymbol{a}_n^{\mathrm{H}} \boldsymbol{B}_{kn} \boldsymbol{a}_n &\leq \frac{1}{2^{\bar{r}_k} - 1} \boldsymbol{a}_k^{\mathrm{H}} \boldsymbol{c}_k \boldsymbol{c}_k^{\mathrm{H}} \boldsymbol{a}_k \end{aligned}$$

which allows us to apply Lagrange duality to get the dual problem

$$\max_{\boldsymbol{\mu},\boldsymbol{\lambda}} \boldsymbol{\lambda}^{\mathrm{T}} \mathbf{1} \quad \text{s.t. } \begin{array}{l} \gamma_k^{\mathrm{ul}}(\boldsymbol{\lambda}, \boldsymbol{\mu}) \leq 2^{\bar{r}_k} - 1 \;\forall k \\ \boldsymbol{\mu}^{\mathrm{T}} \mathbf{1} = 1 \\ \boldsymbol{\mu} \geq 0, \quad \boldsymbol{\lambda} \geq 0 \end{array} \tag{5}$$

where

$$\gamma_k^{\mathrm{ul}}(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \lambda_k \boldsymbol{c}_k^{\mathrm{H}} \Big(\sum_n \lambda_n \boldsymbol{B}_{nk} + \sum_\ell \mu_\ell \boldsymbol{Q}_{\ell k} \Big)^{-1} \boldsymbol{c}_k \tag{6}$$

and $\mathbf{1}$ is an all-ones vector. The uplink SINRs in (6) can also be interpreted as the optimal SINRs

$$\gamma_k^{\mathrm{ul}}(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \max_{\boldsymbol{g}_1,\ldots,\boldsymbol{g}_K} \frac{\lambda_k \boldsymbol{g}_k^{\mathrm{H}} \boldsymbol{c}_k \boldsymbol{c}_k^{\mathrm{H}} \boldsymbol{g}_k}{\boldsymbol{g}_k^{\mathrm{H}} \big(\sum_n \lambda_n \boldsymbol{B}_{nk} + \sum_\ell \mu_\ell \boldsymbol{Q}_{\ell k} \big) \boldsymbol{g}_k}$$

with the optimal uplink filters

$$\boldsymbol{g}_k^{\star} = \Big(\sum_n \lambda_n \boldsymbol{B}_{nk} + \sum_\ell \mu_\ell \boldsymbol{Q}_{\ell k} \Big)^{-1} \boldsymbol{c}_k. \tag{7}$$

Due to the monotonicity properties of the uplink SINRs, the SINR constraints are all binding in the optimum and there is a unique solution for $\boldsymbol{\lambda} \geq \mathbf{0}$. The maximization over $\boldsymbol{\mu}$ in (5) leads to a worst-case "noise-covariance" in the uplink domain.

Note that the optimal value of the dual problem (5) is equal to the optimal power factor of the primal problem (4). For the rate-balancing problem we want $\beta = 1$ and optimal rate targets for given balancing factors τ_k. This leads to the dual rate-balancing problem (cf. [13])

$$\min_{\boldsymbol{\mu}} \min_{\boldsymbol{\lambda}, r_0} r_0 \quad \text{s.t. } \begin{array}{l} \bar{\gamma}_k(r_0) \geq \gamma_k^{\mathrm{ul}}(\boldsymbol{\lambda}, \boldsymbol{\mu}) \;\forall k \\ \boldsymbol{\lambda}^{\mathrm{T}} \mathbf{1} = 1, \quad \boldsymbol{\mu}^{\mathrm{T}} \mathbf{1} = 1 \\ \boldsymbol{\mu} \geq \mathbf{0}, \quad \boldsymbol{\lambda} \geq \mathbf{0}. \end{array}$$

where

$$\bar{\gamma}_k(r) = 2^{\tau_k r} - 1.$$

Since the SINR constraints are all binding, the optimal solution for the inner problem satisfies

$$\begin{aligned} &\bar{\boldsymbol{\gamma}}(r_0^{\star}) = \boldsymbol{\gamma}^{\mathrm{ul}}(\boldsymbol{\lambda}^{\star}, \boldsymbol{\mu}) \\ &\mathbf{1}^{\mathrm{T}} \boldsymbol{\lambda}^{\star} = 1. \end{aligned} \tag{8}$$

where we stacked the SINRs into the vectors

$$\bar{\gamma}(r_0) = \begin{bmatrix} \bar{\gamma}_1(r_0) \\ \vdots \\ \bar{\gamma}_K(r_0) \end{bmatrix} \quad \text{and} \quad \gamma^{\text{ul}}(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \begin{bmatrix} \gamma_1^{\text{ul}}(\boldsymbol{\lambda}, \boldsymbol{\mu}) \\ \vdots \\ \gamma_K^{\text{ul}}(\boldsymbol{\lambda}, \boldsymbol{\mu}) \end{bmatrix}.$$

Different approaches have been suggested in literature to solve this non-linear system of equations. For example, the equations can be reformulated into a fixed-point form [12]. For the power-minimization problem, the Newton-Raphson method is known to converge globally with a feasible initialization [11]. In our proposed optimization method for the rate-balancing problem, we also adopt the Newton-Raphson method and ensure that the dual variables $\boldsymbol{\lambda}$ are always non-negative. In our experiments, the algorithm always converged to the unique solution of (9).

For a single power constraint, we simply have $\mu = 1$. However, since we consider one constraint per base station, we find the optimal $\boldsymbol{\mu}$ with a projected gradient method.

To this end, we obtain the gradient $\partial r_0^\star / \partial \boldsymbol{\mu}$ by applying the implicit function theorem to the optimality conditions in (8). Defining

$$\boldsymbol{q}_\ell = \frac{\partial \boldsymbol{\gamma}^{\text{ul}}}{\partial \mu_\ell}, \ \boldsymbol{\varXi} = \frac{\partial \boldsymbol{\gamma}^{\text{ul}}}{\partial \boldsymbol{\lambda}^{\text{T}}}, \ \text{and} \ \boldsymbol{d} = \frac{\partial \bar{\boldsymbol{\gamma}}}{\partial r_0}$$

the resulting linear system of equations for the derivative with respect to μ_ℓ is given by

$$\begin{bmatrix} \boldsymbol{\varXi} & -\boldsymbol{d} \\ \boldsymbol{1}^{\text{T}} & 0 \end{bmatrix} \begin{bmatrix} \frac{\partial \boldsymbol{\lambda}^\star}{\partial \mu_\ell} \\ \frac{\partial r_0^\star}{\partial \mu_\ell} \end{bmatrix} = \begin{bmatrix} \boldsymbol{q}_\ell \\ 0 \end{bmatrix}. \tag{9}$$

With

$$\boldsymbol{x} = (\boldsymbol{\varXi}^{\text{T}})^{-1} \boldsymbol{1} \tag{10}$$

the desired derivative is thus given by

$$\frac{\partial r_0^\star}{\partial \mu_\ell} = -\frac{\boldsymbol{x}^{\text{T}} \boldsymbol{q}_\ell}{\boldsymbol{x}^{\text{T}} \boldsymbol{d}}.$$

Note that the matrix on the left-hand side of (9) and thus the vector $\boldsymbol{x}$ is the same for all μ_ℓ. Therefore,

$$\frac{\partial r_0^\star}{\partial \boldsymbol{\mu}} = \frac{1}{\boldsymbol{x}^{\text{T}} \boldsymbol{d}} \boldsymbol{Q}^{\text{T}} \boldsymbol{x} \ \text{ with } \ \boldsymbol{Q} = [\boldsymbol{q}_1, \ldots, \boldsymbol{q}_\ell].$$

The matrix $\boldsymbol{\varXi}$ also needs to be inverted for the Newton steps on (8) which allows us to save some complexity in the proposed optimization algorithm.

Due to the structure of the dual problem, we know that the optimal uplink SINRs $\gamma_k^{\text{ul}}(\boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star) = \bar{\gamma}_k(r_0^\star)$ are also achievable in the downlink [13] with the beamformers

$$\boldsymbol{a}_k^\star = p_k \boldsymbol{g}_k^\star.$$

The downlink power allocation is found by simply solving the system of equations

$$\gamma_k(p_1 \boldsymbol{g}_1^\star, \ldots, p_K \boldsymbol{g}_K^\star) = \bar{\gamma}_k(r_0^\star) \ \forall k$$

which is linear in the p_k. In fact, the linear system is given by

$$\boldsymbol{\varXi}^{\text{T}} \operatorname{diag}(\boldsymbol{\lambda})^{-1} \boldsymbol{p} = \frac{1}{\rho_{\text{ul}}} \boldsymbol{1}.$$

Consequently,

$$\boldsymbol{p} = \frac{1}{\rho_{\text{ul}}} \operatorname{diag}(\boldsymbol{\lambda}) \boldsymbol{x}$$

with $\boldsymbol{x}$ from (10).

For a practical implementation, an adaptive algorithm is preferred, such that we are able to update the transformations $\boldsymbol{A}_k$ for small changes in the covariance matrices. The proposed algorithm is described in Alg. 1. Since the optimal $\boldsymbol{\mu}$ changes only slightly from one coherence interval to the next, we can use a fixed, small step size $\alpha_{\boldsymbol{\mu}}$ for the gradient step.

Algorithm 1 Adaptive rate-balancing algorithm

1: **for** each coherence interval **do**
2: Calculate $\boldsymbol{\gamma}^{\text{ul}}$, $\boldsymbol{g}_1, \ldots, \boldsymbol{g}_K$, $\boldsymbol{\varXi}$, $\boldsymbol{d}$, $\boldsymbol{Q}$ using $\boldsymbol{\lambda}$, $\boldsymbol{\mu}$ and r_0
3: Apply LU-factorization to $\boldsymbol{\varXi}$
4: Use backsubstitutions to get

$$\boldsymbol{x} \leftarrow (\boldsymbol{\varXi}^{\text{T}})^{-1} \boldsymbol{1}$$

5: Calculate Newton steps

$$\Delta r_0 \leftarrow -\frac{\boldsymbol{x}^{\text{T}}(\bar{\gamma} - \gamma^{\text{ul}})}{\boldsymbol{x}^{\text{T}} \boldsymbol{d}}$$
$$\Delta \boldsymbol{\lambda} \leftarrow \boldsymbol{\varXi}^{-1}(\bar{\gamma} - \gamma^{\text{ul}}) + \Delta r_0 \boldsymbol{\varXi}^{-1} \boldsymbol{d}$$

6: Update r_0 and $\boldsymbol{\lambda}$

$$r_0 \leftarrow r_0 + \Delta r_0$$
$$\alpha_{\boldsymbol{\lambda}} \leftarrow \max_{\alpha \in [0,1]} \alpha \text{ s.t. } \boldsymbol{\lambda} + \alpha \Delta \boldsymbol{\lambda} \geq 0$$
$$\boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + \alpha_{\boldsymbol{\lambda}} \Delta \boldsymbol{\lambda}$$

7: Perform projected gradient step for $\boldsymbol{\mu}$ with projection $\mathcal{P}(\cdot)$ onto the unit-simplex

$$\Delta \boldsymbol{\mu} \leftarrow \frac{1}{\boldsymbol{x}^{\text{T}} \boldsymbol{d}} \boldsymbol{Q}^{\text{T}} \boldsymbol{x}$$
$$\boldsymbol{\mu} \leftarrow \mathcal{P}(\boldsymbol{\mu} + \alpha_{\boldsymbol{\mu}} \Delta \boldsymbol{\mu})$$

8: Calculate downlink power allocation

$$\boldsymbol{p} \leftarrow \frac{1}{\rho_{\text{dl}}} \operatorname{diag}(\boldsymbol{\lambda}) \boldsymbol{x}$$

9: **end for**

IV. Computational Complexity

As noted in [10], it is advantageous in terms of complexity but also in terms of estimation accuracy to approximate the channel covariance matrices by

$$\boldsymbol{C}_{\ell k} \approx \boldsymbol{F}^{\text{H}} \operatorname{diag}(\boldsymbol{c}_{\ell k}) \boldsymbol{F} \tag{11}$$

where $\boldsymbol{F}$ is some unitary transformation which approximately diagonalizes the covariance matrices. Important examples are uniform linear arrays and uniform rectangular arrays. For typical physical channel models, the respective covariance matrices are approximately diagonalized by the discrete Fourier transform (DFT) matrix or a Kronecker product of DFT matrices [2], [9], [10].

These approximations significantly reduce the complexity of the proposed method, since for covariance matrices with the structure in (11) the optimal $\boldsymbol{A}_{\ell k}$ under per-base-station constraints have the same structure (see Appendix), i.e.,

$$\boldsymbol{A}_{\ell k} = \boldsymbol{F}^{\mathrm{H}} \operatorname{diag}(\boldsymbol{a}_{\ell k}) \boldsymbol{F}. \tag{12}$$

Thus, if the approximation in (11) is applicable, we have

$$\boldsymbol{c}_k = \begin{bmatrix} \boldsymbol{c}_{1k} \\ \vdots \\ \boldsymbol{c}_{Lk} \end{bmatrix} \in \mathbb{C}^{LM}$$

and

$$\boldsymbol{B}_{kn} = \begin{cases} \operatorname{diag}(\boldsymbol{c}_k \odot \hat{\boldsymbol{c}}_n) + \boldsymbol{c}_k \boldsymbol{c}_k^{\mathrm{H}} & \text{for } k \in \mathcal{I}_n \\ \operatorname{diag}(\boldsymbol{c}_k \odot \hat{\boldsymbol{c}}_n) & \text{otherwise} \end{cases}$$
$$\boldsymbol{Q}_{\ell n} = \operatorname{diag}((\boldsymbol{e}_\ell \otimes \mathbf{1}_M) \odot \hat{\boldsymbol{c}}_n)$$

with the element-wise multiplication $\odot$ and the Kronecker product $\otimes$. The optimal filters are consequently given by

$$\boldsymbol{g}_k^\star = \Big(\operatorname{diag}(\hat{\boldsymbol{c}}_k \odot (\boldsymbol{\mu} \otimes \mathbf{1} + \sum_n \lambda_n \boldsymbol{c}_n)) + \sum_{n \in \mathcal{I}_k} \lambda_n \boldsymbol{c}_n \boldsymbol{c}_n^{\mathrm{H}} \Big)^{-1} \boldsymbol{c}_k.$$

The computational complexity of the calculation is $O(M^3 L^3)$ per user, however, by applying the matrix inversion lemma (cf. Appendix), we reduce the complexity to $O(|\mathcal{I}_k|^3 + ML\,|\mathcal{I}_k|^2)$. Note that the weighted sum over all covariance matrices, which needs $O(MLK)$ operations, is the same for all users.

The main complexity for the remaining steps is $O(MLK^2)$ to calculate $\boldsymbol{\Xi}$ and $O(K^3)$ for the factorization. All other steps need $O(MLK)$ operations. The total complexity of one iteration of Alg. 1 is thus $O(MLK^2 + K^3 + \sum_k |\mathcal{I}_k|^3 + ML\,|\mathcal{I}_k|^2)$. Note that the complexity is linear in the total number of transmit antennas ML and typically $ML \gg |\mathcal{I}_k|$.

V. Per-Antenna Constraints

The proposed algorithm can be extended to per-antenna constraints in a straight-forward manner. As a consequence, the number of dual variables is increased to ML, which leads to a quadratic complexity in ML for the calculation of the gradient $\partial r_0^\star / \partial \boldsymbol{\mu}$.

The per-base-station power constraints for the approximated covariance matrices in (11) contain terms of the form

$$\begin{aligned} &\operatorname{tr}(\boldsymbol{F}^{\mathrm{H}} \operatorname{diag}(\boldsymbol{a}_{\ell k}) \boldsymbol{F} \boldsymbol{F}^{\mathrm{H}} \operatorname{diag}(\hat{\boldsymbol{c}}_{\ell k}) \boldsymbol{F} \boldsymbol{F}^{\mathrm{H}} \operatorname{diag}(\boldsymbol{a}_{\ell k})^{\mathrm{H}} \boldsymbol{F}) \\ =& \operatorname{tr}(\boldsymbol{F}^{\mathrm{H}} \operatorname{diag}(\boldsymbol{a}_{\ell k} \odot \hat{\boldsymbol{c}}_{\ell k} \odot \boldsymbol{a}_{\ell k}^*) \boldsymbol{F}) \\ =& \boldsymbol{a}_{\ell k}^{\mathrm{H}} \operatorname{diag}(\hat{\boldsymbol{c}}_{\ell k}) \boldsymbol{a}_{\ell k}. \end{aligned}$$

The constraint for antenna m on the other hand sums over

$$\begin{aligned} &\boldsymbol{e}_m^{\mathrm{T}} \boldsymbol{F}^{\mathrm{H}} \operatorname{diag}(\boldsymbol{a}_{\ell k} \odot \hat{\boldsymbol{c}}_{\ell k} \odot \boldsymbol{a}_{\ell k}^*) \boldsymbol{F} \boldsymbol{e}_m \\ =& \boldsymbol{f}_m^{\mathrm{H}} \operatorname{diag}(\boldsymbol{a}_{\ell k} \odot \hat{\boldsymbol{c}}_{\ell k} \odot \boldsymbol{a}_{\ell k}^*) \boldsymbol{f}_m \end{aligned}$$

which due to the constant modulus of the entries in the DFT basis vectors $\boldsymbol{f}_m$ is equal to

$$\frac{1}{M} \boldsymbol{a}_{\ell k}^{\mathrm{H}} \operatorname{diag}(\hat{\boldsymbol{c}}_{\ell k}) \boldsymbol{a}_{\ell k}.$$

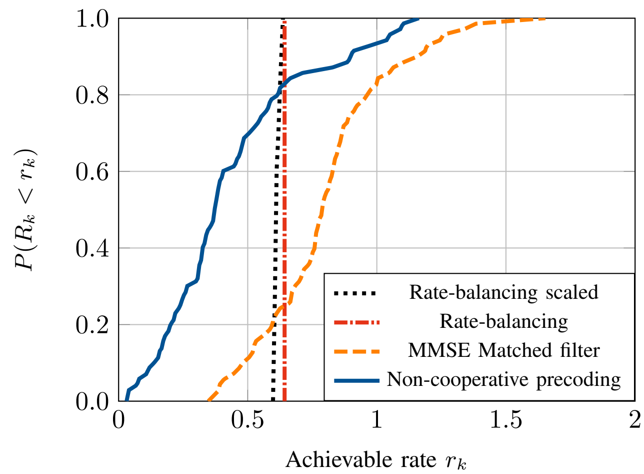

Fig. 1. Cumulative distribution of the achievable rates for one snapshot of $K = 70$ uniformly distributed users in a network with $L = 7$ cells. Each base station is equipped with $M = 20$ antennas and $T_{\mathrm{tr}} = 10$ channel accesses are used for training. The results are for per-base-station constraints. The dotted graph shows the result for rate-balancing with a single sum-power constraint scaled to meet the per-base-station requirements.

In other words, the per-antenna and per base-station constraints are equivalent for the approximation in (11) and thus the complexity does not increase when considering per-antenna constraints.

VI. Results

We compare the rate-balancing performance with matched-filter precoding based on the MMSE channel estimation. That is, for diagonal covariance matrices, we have

$$\boldsymbol{a}_k^{\mathrm{MMSE}} = \frac{\operatorname{diag}(\hat{\boldsymbol{c}}_k)^{-1} \boldsymbol{c}_k}{\boldsymbol{c}_k^{\mathrm{T}} \operatorname{diag}(\hat{\boldsymbol{c}}_k)^{-1} \boldsymbol{c}_k}.$$

We use a uniform power allocation for the MMSE approach with joint scaling for all users to meet the power constraints. Additionally, we show results for rate-balancing with a sum-power constraint, where the precoders are also jointly scaled to meet the per-base-station constraints. As a baseline we also depict results for non-cooperative transmission, i.e., each user is served only by the closest base-station with a matched filter based on the local MMSE estimation of the channel.

Results are presented in Fig. 1 for one snapshot of $K = 70$ uniformly distributed users in a network with $L = 7$ cells. Due to the even distribution of users in the network, the per-base-station constraints do not significantly affect the performance. The performance of the MMSE matched filter scaled to meet the sum-power constraint is in fact almost identical to the case with per-base-station constraints.

These results are expected, since even with the proposed rate-balancing approach, there is a significant amount of residual interference and even more so for the matched filter approach. The joint scaling of the precoders has a notable effect only if the noise power is comparable to the interference. To illustrate this behavior, we depict the minimal achievable

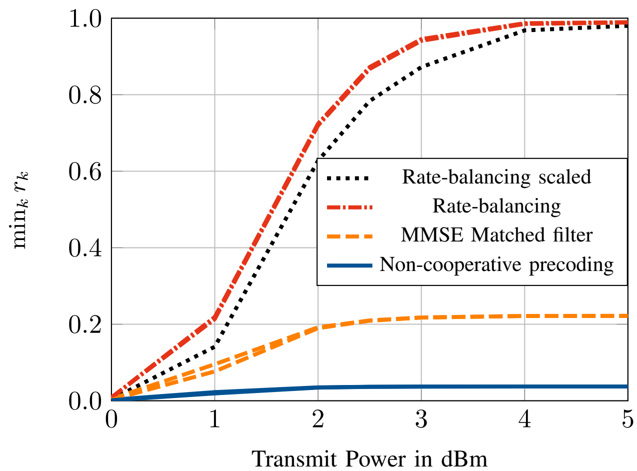

Fig. 2. Minimal achievable rates averaged over several snapshot of $K = 70$ uniformly distributed users in a network with $L = 7$ cells. Each base station is equipped with $M = 20$ antennas and $T_{\text{tr}} = 10$ channel accesses are used for training. The results for both, sum-power and per-base-station constraints are depicted in the same style. The sum-power curves are strictly higher than the per base-station ones, but in most cases, the difference is barely notable. The dotted graph shows the result for rate-balancing with a single sum-power constraint scaled to meet the per-base-station requirements.

rate averaged over several user placements with respect to the transmit power in Fig. 2. We note that the relative gain of explicitly taking all constraints into account is larger for low transmit power and vanishes for high transmit power.

VII. Conclusion

We demonstrated in numerical simulations the performance gains of explicitly taking per-base station power constraints into account. The gains depend on the distribution of the user terminals and the transmit power. Under favorable conditions, simpler heuristics yield almost the same performance. Nevertheless, the proposed adaptive optimization algorithm allows us to deal with multiple power constraints with a reasonable amount of complexity. The proposed approach is particularly interesting in a scenario with groups of distributed antennas.

Appendix

To show that the optimal transformation matrices $\boldsymbol{A}_k$ and thus the optimal filter $\boldsymbol{g}_k$ have the same structure as the covariance matrices, we consider the optimal filters for general covariance matrices. For general channel covariance matrices $\boldsymbol{C}_k$ with $\boldsymbol{c}_k = \text{vec}(\boldsymbol{C}_k)$, we have

$$\boldsymbol{B}_{kn} = \begin{cases} \hat{\boldsymbol{C}}_n^* \otimes \boldsymbol{C}_n + \boldsymbol{c}_k \boldsymbol{c}_k^{\text{H}} & \text{for } k \in \mathcal{I}_n \\ \hat{\boldsymbol{C}}_n^* \otimes \boldsymbol{C}_n & \text{otherwise} \end{cases}$$

$$\boldsymbol{Q}_{\ell n} = \hat{\boldsymbol{C}}_n^* \otimes \text{diag}(\boldsymbol{e}_\ell \otimes \mathbf{1}_M).$$

The optimal filter is thus given by

$$\boldsymbol{g}_k^\star = \Big(\sum_n \lambda_n \hat{\boldsymbol{C}}_k^* \otimes \boldsymbol{C}_n + \sum_{n \in \mathcal{I}_k} \lambda_n \boldsymbol{c}_n \boldsymbol{c}_n^{\text{H}} + \hat{\boldsymbol{C}}_k^* \otimes \boldsymbol{\Gamma}(\boldsymbol{\mu})\Big)^{-1} \boldsymbol{c}_k.$$

where $\boldsymbol{\Gamma}(\boldsymbol{\mu}) = \text{blkdiag}(\mu_1 \mathbf{I}, \ldots, \mu_L \mathbf{I})$. For the interfering users $\mathcal{I}_k = \{a, b, c, \ldots\}$ define

$$\boldsymbol{Z}_k = [\boldsymbol{c}_a, \boldsymbol{c}_b, \boldsymbol{c}_c, \ldots], \quad \bar{\boldsymbol{\lambda}}_k = [\lambda_a, \lambda_b, \lambda_c, \ldots]^{\text{T}}$$

and

$$\boldsymbol{S}_k = \hat{\boldsymbol{C}}_k^* \otimes \big(\boldsymbol{\Gamma}(\boldsymbol{\mu}) + \sum_n \lambda_n \boldsymbol{C}_n\big).$$

With the matrix inversion lemma, we have

$$\boldsymbol{g}_k^\star = \big(\boldsymbol{S}_k^{-1} - \boldsymbol{S}_k^{-1} \boldsymbol{Z}_k (\text{diag}(\bar{\boldsymbol{\lambda}}_k)^{-1} + \boldsymbol{Z}_k^{\text{H}} \boldsymbol{S}_k^{-1} \boldsymbol{Z}_k)^{-1} \boldsymbol{Z}_k^{\text{H}} \boldsymbol{S}_k^{-1}\big) \boldsymbol{c}_k.$$

Note that the optimal filter $\boldsymbol{g}_k^\star$ is a linear combination of terms of the form

$$\boldsymbol{S}_k^{-1} \boldsymbol{c}_n = \text{vec}\Big(\big(\boldsymbol{\Gamma}(\boldsymbol{\mu}) + \sum_i \lambda_i \boldsymbol{C}_i\big) \boldsymbol{C}_n \hat{\boldsymbol{C}}_k\Big). \tag{13}$$

Thus, for block-diagonal covariance matrices $\boldsymbol{C}_k$, the optimal $\boldsymbol{A}_k$ are also block-diagonal. Additionally, if each block $\boldsymbol{C}_{\ell k}$ satisfies (11) the optimal $\boldsymbol{A}_{\ell k}$ are of the form (12), since $\boldsymbol{\Gamma}(\boldsymbol{\mu})$ is block-diagonal with scaled identities for the different blocks.

References

[1] L. Zheng and D. Tse, "Communication on the Grassmann manifold: a geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, pp. 359–383, Feb. 2002.

[2] H. Huh, G. Caire, H. Papadopoulos, and S. Ramprashad, "Achieving "Massive MIMO" Spectral Efficiency with a Not-so-Large Number of Antennas," *IEEE Trans. Wireless Commun.*, vol. 11, pp. 3226–3239, Sep. 2012.

[3] T. Marzetta, "Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas," *IEEE Trans. Wireless Commun.*, vol. 9, pp. 3590 –3600, Nov. 2010.

[4] D. Neumann, M. Joham, and W. Utschick, "CDI Precoding for Massive MIMO," in *Conf. Syst. Commun. and Coding*, Feb. 2015, pp. 1–6.

[5] A. Ashikhmin and T. Marzetta, "Pilot contamination precoding in multi-cell large scale antenna systems," in *Int. Symp. Inf. Theory*, Jul. 2012, pp. 1137 –1141.

[6] D. Neumann, A. Gründinger, M. Joham, and W. Utschick, "Rate-balancing in massive MIMO using statistical precoding," in *Int. Workshop Signal Process. Adv. Wireless Commun.*, Jun. 2015, pp. 226–230.

[7] H. Q. Ngo, A. Ashikhmin, H. Yang, E. Larsson, and T. Marzetta, "Cell-Free Massive MIMO: Uniformly great service for everyone," in *Int. Workshop Signal Process. Adv. Wireless Commun.*, Jun. 2015, pp. 201–205.

[8] M. Medard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inf. Theory*, vol. 46, pp. 933–946, May 2000.

[9] F.-W. Sun, Y. Jiang, and J. Baras, "On the convergence of the inverses of Toeplitz matrices and its applications," *IEEE Trans. Inf. Theory*, vol. 49, pp. 180–190, Jan. 2003.

[10] D. Neumann, M. Joham, L. Weiland, and W. Utschick, "Low-Complexity Computation of LMMSE Channel Estimates in Massive MIMO," in *Workshop on Smart Antennas*, Mar. 2015, pp. 1–6.

[11] M. Schubert and H. Boche, "A Generic Approach to QoS-Based Transceiver Optimization," *IEEE Trans. Commun.*, vol. 55, pp. 1557–1566, Aug. 2007.

[12] M. Joham, A. Gründinger, A. Pastore, J. Fonollosa, and W. Utschick, "Rate balancing in the vector BC with erroneous CSI at the receivers," in *Conf. Inf. Sci. and Syst.*, Mar. 2013, pp. 1–6.

[13] A. Fehske, F. Richter, and G. Fettweis, "SINR Balancing for the Multi-User Downlink under General Power Constraints," in *Glob. Telecom. Conf.*, Nov. 2008, pp. 1–6.